ZFX acts as a transcriptional activator in multiple types of human tumors by binding downstream of transcription start sites at the majority of CpG island promoters

Suhn Kyong Rhie, Lijun Yao, Zhifei Luo, et al.

Genome Res. published online February 2, 2018
Access the most recent version at doi:10.1101/gr.228809.117

ZFX acts as a transcriptional activator in multiple types of human tumors by binding downstream of transcription start sites at the majority of CpG island promoters

Suhn Kyong Rhie, Lijun Yao, Zhifei Luo, Heather Witt, Shannon Schreiner,
Yu Guo, Andrew A. Perez, and Peggy J. Farnham*

Department of Biochemistry and Molecular Medicine and the Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles 90089, CA, USA.

* To whom correspondence should be addressed. Tel: +1 323 442 8015; Fax: +1 323 865 1509; Email: peggy.farnham@med.usc.edu

Running title: ZFX activates transcription at CGI promoters

Keywords: ZFX, zinc finger proteins, CpG islands, promoters, transcriptional regulation, chromatin structure

ABSTRACT

High expression of the transcription factor ZFX is correlated with proliferation, tumorigenesis, and patient survival in multiple types of human cancers. However, the mechanism by which ZFX influences transcriptional regulation has not been determined. We performed ChIP-seq in four cancer cell lines (representing kidney, colon, prostate, and breast cancers) to identify ZFX binding sites throughout the human genome. We identified ~9,000 ZFX binding sites and found that the majority of the sites are in CpG island promoters. Moreover, genes with promoters bound by ZFX are expressed at higher levels than genes with promoters not bound by ZFX. To determine if ZFX contributes to regulation of the promoters to which it is bound, we performed RNA-seq analysis after knockdown of ZFX by siRNA in prostate and breast cancer cells. Many genes with promoters bound by ZFX were downregulated upon ZFX knockdown, supporting the hypothesis that ZFX acts as a transcriptional activator. Surprisingly, ZFX binds at +240bp downstream of the TSS of the responsive promoters. Using Nucleosome Occupancy and Methylome Sequencing (NOMe-seq), we show that ZFX binds between the open chromatin region at the TSS and the first downstream nucleosome, suggesting that ZFX may play a critical role in promoter architecture. We have also shown that a closely related zinc finger protein ZNF711 has a similar binding pattern at CpG island promoters, but ZNF711 may play a subordinate role to ZFX. This functional characterization of ZFX provides important new insights into transcription, chromatin structure, and the regulation of the cancer transcriptome.

INTRODUCTION

Altered transcriptomes are a general characteristic of human cancers. In many cases, the transcriptional dysregulation is driven by altered expression levels or activity of transcription factors (TFs) (Yao et al. 2015; Rhie et al. 2016). There are ~2,000 DNA-binding TFs in human genome, but little is known about most of these regulators (Vaquerizas et al. 2009; Wingender et al. 2015). We have previously identified distinct sets of TFs having increased expression associated with different cancers (Yao et al. 2015; Rhie et al. 2016). In contrast, ZFX, a zinc finger protein (ZNF) that contains a DNA binding domain, has been implicated in the initiation or progression of many different types of human cancers, including prostate cancer, breast cancer, colorectal cancer, renal cell carcinoma, glioma, gastric cancer, gallbladder adenocarcinoma, non-small cell lung carcinoma, and laryngeal squamous cell carcinoma (Zhou et al. 2011; Fang et al. 2012; Jiang et al. 2012b; Jiang et al. 2012c; Nikpour et al. 2012; Li et al. 2013; Fang et al. 2014a; Fang et al. 2014b; Yang et al. 2014; Weng et al. 2015). In these previous studies, it was shown that high expression of ZFX is linked to tumorigenesis and that knocking down ZFX can suppress cellular proliferation and increase the proportion of apoptotic cells (Fang et al. 2014a; Jiang and Liu 2015; Yang et al. 2015; Yan et al. 2016). In addition, high ZFX expression correlates with poor survival of cancer patients (Jiang and Liu 2015; Li et al. 2015; Yang et al. 2015; Yan et al. 2016). For example, ZFX expression is significantly related to histological grade (p-value <0.001) in gallbladder adenocarcinoma and patients that survived less than 1 year were found to have significantly higher ZFX expression than patients that survived more than 1 year (Weng et al., 2015). Taken together, these studies suggest that ZFX may function as an oncogene. However, the mechanism by which ZFX may influence transcriptional regulation in such a diverse set of human tumors has not been determined.

ZFX is encoded on the X Chromosome and highly conserved in vertebrates. Among the ~2,000 site-specific DNA-binding TFs, the C2H2 ZNFs are the largest class encoded in the human genome. Although the biological functions of the majority of ZNFs are unknown, the molecular functions of ZNFs include not only sequence-specific binding to DNA but also protein-

protein interactions and RNA binding (Stubbs et al. 2011; Najafabadi et al. 2015). DNA-binding ZNFs generally have multiple, adjacent, properly spaced zinc fingers in their DNA binding domain; ZNFs with fewer than 3 properly spaced fingers are more likely to be involved in protein-protein or protein-RNA interactions (Brown 2005; Brayer and Segal 2008). ZFX has 13 C2H2-type zinc fingers in its putative DNA binding domain, the last 9 of which are properly spaced, supporting the hypothesis that ZFX is a DNA binding factor. ZFX contains a large acidic activation domain in addition to the C2H2-type zinc finger-containing DNA binding domain, suggesting that, in contrast to the hundreds of ZNFs that contain a KRAB repression domain, ZFX may be a transcriptional activator.

Although the structure of ZFX suggests that it is a DNA binding transcriptional activator that is expressed at high levels in many different types of cancers, ZFX binding sites have not yet been mapped in cancer cells. To understand the mechanism by which ZFX may regulate the cancer transcriptome, we performed ChIP-seq, NOMe-seq, and RNA-seq assays with knockdown experiments in HEK293T kidney, HCT116 colon, C42B prostate, and MCF-7 breast cancer cells, identifying ZFX-binding sites and ZFX-regulated genes throughout the human genome.

RESULTS

ZFX binds to CpG island promoter regions.

To profile the genome-wide binding sites of ZFX, we performed two biological ZFX ChIP-seq replicates using chromatin from HEK293T kidney, HCT116 colon, C42B prostate, and MCF-7 breast cancer cells (Fig. 1A, see Supplemental Fig. S1 for ZFX antibody validation and Supplemental Table S1 for access information for all genomic datasets). We chose to use these cancer models because there is a strong link in these 4 cancer types between ZFX expression and cell proliferation, tumor development, or patient survival. For example, prostate cancer tissues exhibit significantly higher ZFX expression than benign prostatic hyperplasia and adjacent tissues and siRNA-mediated knockdown of ZFX suppresses the proliferation of

prostate cancer cells and reduces the number of colonies in colony forming assays (Jiang et al. 2012a). Similarly, expression of ZFX is high in advanced invasive breast cancers and knockdown of ZFX reduces the proliferation rate of breast cancer cells (Yang et al. 2014). High expression of ZFX promotes tumor growth of colon cancer cells and colorectal cancer patients with high ZFX expression have poorer overall and disease-free survival (Jiang and Liu 2015). Moreover, knockdown of ZFX suppresses proliferation and invasion of colon cancer cell lines (Jiang and Liu 2015). Finally, ZFX is significantly upregulated in renal cell carcinomas (RCC) and has been suggested to be a strong predictor for prognosis of RCC patients (Li et al. 2015). Also, knockdown of ZFX expression in renal carcinoma cells results in significantly inhibited proliferation and cell cycle progression (Fang et al. 2014a).

We identified ~9,000 reproducible ZFX binding sites in each cancer cell line (HEK293T: 9,955, HCT116: 9,039, C42B: 8,708, MCF-7: 9,382). Annotation of the ZFX binding sites with respect to different genomic regions showed that ~80% of the sites are located in a promoter region (±2kb of a TSS). In each of the cell types examined, only ~1,000 sites are located in distal elements (i.e. distal sites having H3K27ac or CTCF peaks or other distal sites that are not marked with H3K27ac or CTCF) (Fig. 1B). To further classify the promoter binding sites, we determined if ZFX preferentially binds to housekeeping, CpG island promoters or to more cell type-specific, non-CpG island promoters. We found that the majority of the ZFX peaks are in CpG island promoters (Fig. 1C). In fact, we identified more than 13,000 CpG island promoters that are bound by ZFX in the union of the 4 cell lines, with ~60% of all active CpG island promoters in a given cell type being bound by ZFX, including a strong peak at the promoter of the ZFX gene (Fig. 1A right panel, Supplemental Table S2, Supplemental Table S3A-D).

The ZFX binding pattern at promoter regions is very similar in different cancer types.

Many oncogenic TFs bind to distinct cell type-specific distal regulatory elements in different types of tumors (Rhie et al. 2016). However, the majority of ZFX peaks are in promoter regions, suggesting that ZFX may bind to regulatory elements that are common to all cell types, rather

than to cell type-specific regulatory elements. To further analyze the ZFX binding patterns, we compared the ZFX binding sites in the 4 cancer cell lines; the binding patterns are similar to each other in general (correlation coefficient >0.5) (Fig. 1D). We then separated the peaks into TSS proximal and non-TSS (>2kb from a TSS). We found that the ZFX binding sites in promoter regions are largely shared among the 4 cell lines, with ZFX commonly binding to more than 5,000 promoter peaks in all 4 cell lines (Fig. 1E, Supplemental Table S3E). In contrast, the distal sites bound by ZFX are not always the same in the different cell types. We note that both the common and the cell type-specific ZFX binding sites are robust and reproducible (Fig. 1F).

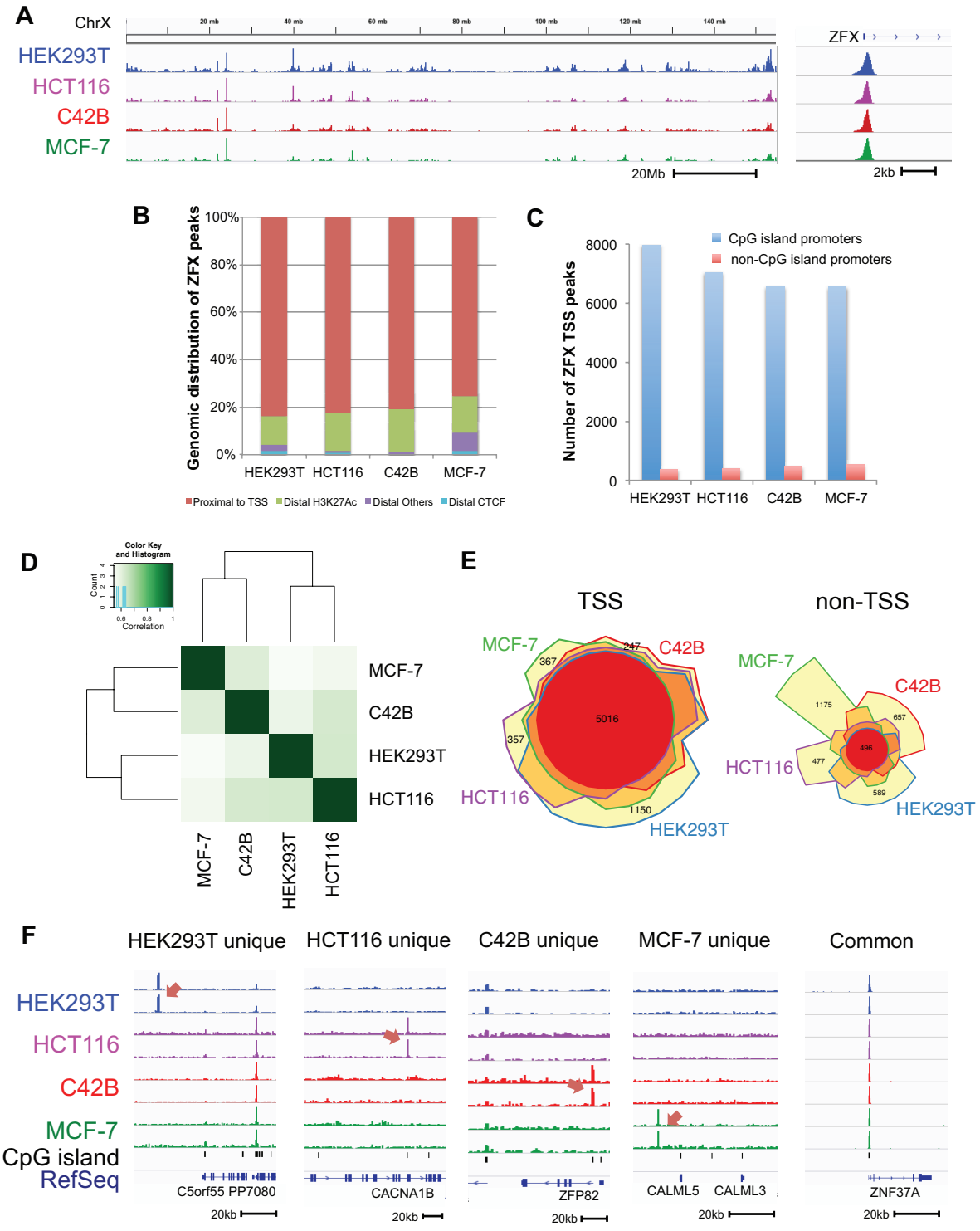


Figure 1. Genome-wide ZFX binding profiles in multiple types of human tumors. (A) Shown is ZFX ChIP-seq data for a ~150 Mb region of Chromosome X for HEK293T kidney, HCT116 colon, C42B prostate, and MCF-7 breast cancer cells (left panel) and for a ~6 kb region near the ZFX promoter (right panel). (B) The percentage of ZFX binding sites in

promoters (±2kb from the TSS), distal enhancers (H3K27ac), distal insulators (CTCF not in enhancers), and at other locations is shown for ZFX ChIP-seq data for 4 cell types. (C) The number of ZFX binding sites located in CpG island promoters vs non-CpG island promoters is shown for the 4 cell types. (D) Shown is a heatmap of the ChIP-seq signal correlation for ZFX binding sites in the 4 tumor types. (E) Shown are Venn diagrams of ZFX binding sites near promoters (left) and distal regions (right) for HEK293T kidney, HCT116 colon, C42B prostate, and MCF-7 breast cancer cells. (F) Examples of cell-type-specific and common ZFX binding sites.

ZFX motifs are enriched at 240bp up and downstream of the TSS in CpG island promoters, but ZFX prefers to bind downstream of the TSS.

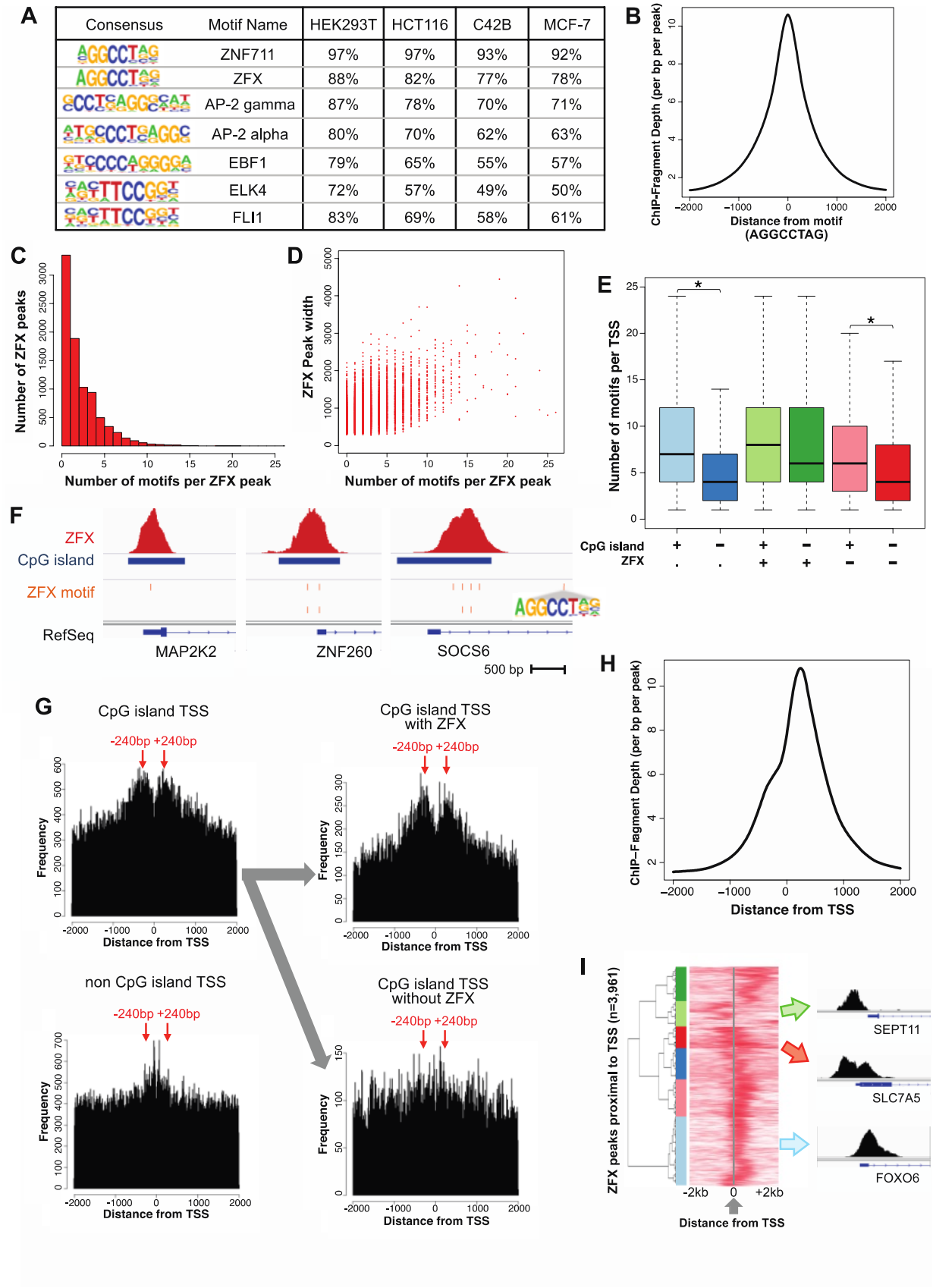
To determine the preferred binding motif for ZFX, we performed motif analyses using 20bp windows from ZFX peak summits. ZFX has 9 properly spaced zinc fingers; because a zinc finger can bind to 3 nucleotides of DNA (Desjarlais and Berg 1992), one would expect a 27 nucleotide motif if all of these fingers are involved in DNA binding. However, the motif we identified in the majority of the ZFX peaks was only 8 nucleotides (AGGCCTAG) with a strong 6 nucleotide consensus (AGGCCT) (Fig. 2A). This motif was originally identified from ZNF711 ChIP-seq data from the brain tumor cell line SH-SY5Y (Kleine-Kohlbrecher et al. 2010) and from ZFX ChIP-seq data from mouse embryonic stem cells (Chen et al. 2008). We also identified several other CT-containing motifs, such as the motif for AP-2. It is possible that multiple sets of zinc fingers in ZFX may recognize a repeating unit of a short motif. Alternatively, it is possible that ZFX utilizes only a subset of its fingers to bind DNA; we have previously described this situation for two artificial 6-finger ZNFs (Grimmer et al. 2014). To further analyze the ZFX binding preferences, we focused on the AGGCCTAG motif, which is found in >90% of the ZFX peaks in each cell line. We showed that the motif is centered in the ZFX peaks, suggesting that this motif directly recruits ZFX (Fig. 2B). However, we also found that some ZFX peaks have many copies of the motif (Fig. 2C), with larger peaks having, in general, more copies (Fig. 2D).

Next, we asked whether this motif is found in all promoter regions or only in promoters bound by ZFX. Surprisingly, we found that 48,373 out of 57,820 promoters of all known genes in the human genome have at least one copy of this motif ±2kb from the TSS. In fact, there are multiple copies of this motif in most promoter regions, with more motifs in CpG island promoters (Fig. 2E). We note that promoters bound by ZFX have, on average, slightly more motifs than promoters not bound by ZFX. However, the number of motifs does not necessarily correspond to the number of ZFX peaks in a promoter. Some promoters, like *MAP2K2*, have one motif and one binding site. Other promoters, such as *ZNF260* and *SOCS6*, have multiple copies of the motif, but not all of the motifs are bound by ZFX (Fig. 2F)

We further investigated the distribution of the AGGCCTAG motif relative to the TSS. We found that this motif is symmetrically enriched ±240bp from the TSS, with CpG island promoters showing greater enrichment than non-CpG island promoters and CpG island promoters bound by ZFX showing greater enrichment than those not bound by ZFX (Fig. 2G). The symmetrical distribution of the motif ±240bp relative to the TSS is unusual and specific for the ZFX motif and not scrambled variants (Supplemental Fig. S2). Therefore, we asked if ZFX binding has a similar distribution or, as for most TFs, if ZFX binds mainly upstream of the TSS. Surprisingly, we found that ZFX has a stronger preference for binding at +240bp downstream of the TSS (Fig. 2H). The reason that there are some peaks at -240bp could be due to the inclusion of bidirectional promoters. To test this hypothesis, we selected the ZFX binding sites that have only one known TSS within a ±2kb window (n=3,961) and plotted heatmaps centered to each TSS (Fig. 2I, Supplemental Table S3F). Most of the time, ZFX is bound at +240bp of the TSS (e.g. *FBXO6*). However, there are a small number of promoters (<10%) that have a ZFX peak at -240bp, (e.g. *SEPT11*) and some promoters that have two ZFX peaks symmetrically located on either side of the TSS (e.g. *SLC7A5*). We conclude that the presence of motif, which is symmetrically enriched ±240bp from the TSS may be necessary, but is not sufficient, to recruit ZFX because ZFX has a higher binding frequency at +240bp than at -240bp.

A

Consensus	Motif Name	HEK293T	HCT116	C42B	MCF-7
	ZNF711	97%	97%	93%	92%
	ZFX	88%	82%	77%	78%
	AP-2 gamma	87%	78%	70%	71%
	AP-2 alpha	80%	70%	62%	63%
	EBF1	79%	65%	55%	57%
	ELK4	72%	57%	49%	50%
	FLI1	83%	69%	58%	61%



1

Figure 2. Characterization of ZFX motifs and binding sites. (A) Shown are motifs enriched at summits of the ZFX binding sites and the percentage of ZFX peaks having each motif for HEK293T kidney, HCT116 colon, C42B prostate, and MCF-7 breast cancer cells; the same motifs were identified in the set of ZFX peaks located near a TSS and the set of distal ZFX peaks. (B) Shown is the average ZFX ChIP-seq signal in C42B relative to ±2kb from the motif AGGCCTAG. (C) Shown is a histogram of the number of motifs per ZFX peak. (D) Shown is a scatterplot of the relationship between the number of motifs per ZFX peak and ZFX peak width. (E) Shown is the number of AGGCCTAG motifs per TSS region (±2kb from the TSS) for different groups of promoters (light blue: CpG island promoters; blue: non-CpG island promoters; light green: CpG island promoters bound by ZFX; green: non-CpG island promoters bound by ZFX; light pink: CpG island promoters not bound by ZFX; red: non-CpG island promoters not bound by ZFX). Comparisons of datasets that show a significant difference (adjusted p-value <0.05) are marked with an asterisk. (F) Shown is an example of ZFX binding site and motif position at CpG island promoters having one (*MAP2K2*) or more (*ZNF260*, *SOCS6*) motifs. (G) Shown is the frequency of AGGCCTAG motifs located ±2kb from the TSS of CpG island promoters, of non-CpG island promoters, of CpG island promoters bound by ZFX, and of CpG island promoters not bound by ZFX; a comparison to results obtained using a scrambled motif can be found in Supplemental Fig. S2. (H) Shown is the average ZFX ChIP-seq signal ±2kb from the TSS of promoters bound by ZFX in MCF-7. (I) Shown is a heatmap representing unsupervised clustering of ZFX ChIP-seq signals in MCF-7 cells at promoters bound by ZFX that only have one TSS in a ±2kb region (n=3,961). Also shown are examples of ZFX binding sites from the light green cluster (promoters having a ZFX peak only upstream of the TSS), the red cluster (promoters having a ZFX peak both up and downstream of the TSS), and the light blue cluster (promoters having a ZFX peak only downstream of the TSS). Genes from each cluster are listed in Supplemental Table S3F.

ZFX has properties of a transcriptional activator.

To determine if ZFX functions as a transcriptional activator or repressor, we separately analyzed expression levels of genes regulated by promoters that are bound by ZFX vs. those promoters not bound by ZFX. We found that the median expression level of genes with promoters bound by ZFX is much higher than the median expression level of genes with promoters not bound by ZFX (Fig. 3A), suggesting that ZFX may be a transcriptional activator. To gain insight into possible mechanisms by which ZFX might activate transcription, we used transient transfection with siRNA to knockdown the levels of ZFX in C42B prostate cancer cells and identified 1,271 genes whose expression decreased and 1,249 genes whose expression increased upon reduction of ZFX levels in C42B cells (FDR <0.05, fold change >1.5) (Fig. 3B and Supplemental Table S4A). Genes identified as responsive to changes in the level of a TF include both direct target genes and genes that are in downstream signaling pathways regulated by the direct target genes (i.e. indirect targets). One approach to identify direct ZFX target genes is to determine which of the deregulated genes have ZFX binding sites in their promoter regions. We found that the promoters of 744 of the 1,271 downregulated genes (58.5%) but only 143 promoters of the 1,249 upregulated genes (11.4%) of the upregulated genes are bound by ZFX in C42B cells (Fig. 3C).

We repeated the siRNA experiments in MCF-7 breast cancer cells. Again, we found that genes having a bound ZFX in their promoter had a median higher expression levels than genes without a bound ZFX (Fig. 3D). However, when we knocked down ZFX, we identified only 418 genes whose expression decreased and 183 genes whose expression increased (Fig. 3E and Supplemental Table S4B). Although the number of deregulated genes in the MCF-7 knockdown experiments is less than in the C42B knockdown experiments, the downregulated genes again have a higher percentage of promoter-bound ZFX (59.1%) than do the upregulated genes (17.5%) (Fig. 3F).

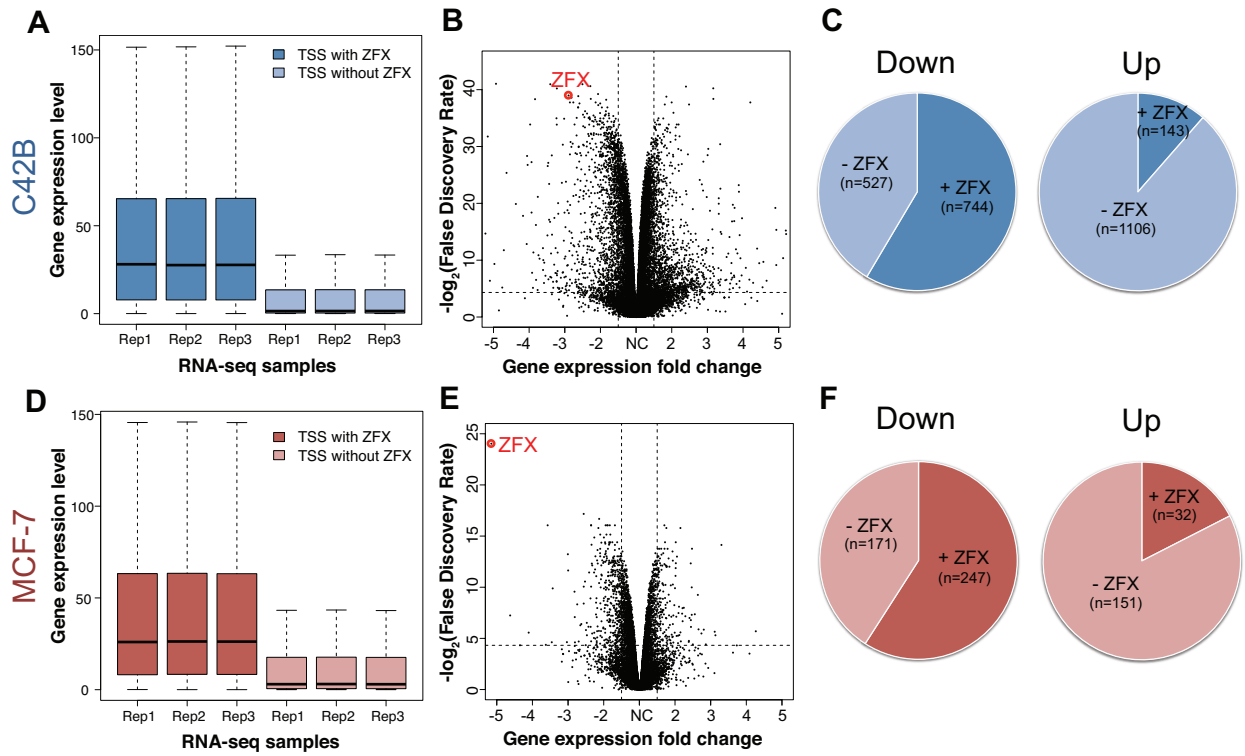


Figure 3. The role of ZFX in transcription regulation. Shown are expression levels of genes with active promoters bound by ZFX and genes with active promoters not bound by ZFX for C42B (A) and MCF-7 (D); active promoters are defined by detectable expression of a transcript from that promoter in that particular cell line. Shown are volcano plots demonstrating differential gene expression after knockdown of ZFX in C42B (B) and MCF-7 (E). Shown are comparisons of the percentage of downregulated vs. upregulated genes that have ZFX bound at their promoters for C42B (C) and MCF-7 (F) cells.

One explanation for the smaller effect on the transcriptome of MCF-7 cells could be inefficient knockdown of ZFX. However, the reduction in ZFX was similar in the siRNA-treated C42B and MCF-7 cells (Supplemental Fig. S3A). An alternative explanation could be that another TF is functionally redundant with ZFX in MCF-7. C2H2 ZNFs comprise the largest class of site-specific DNA-binding proteins encoded in the human genome and have arisen through gene duplication followed by mutation. Specifically, ZFX is very similar to ZFY and ZNF711 (Supplemental Fig. S3B); ZFX and ZNF711 are both encoded on the X Chromosome, whereas

ZFY is located on the Y Chromosome. Overall protein homology is 92% between ZFX and ZFY, with the zinc finger domains having 97% homology, suggesting that these two proteins may have fully redundant activities (Schreiber et al. 2014). However, MCF-7 are female breast cancer cells and thus do not express ZFY (we note that although C42B are male, they also do not express ZFY). There is 55% identity between the entire ZFX and ZNF711 proteins, with the zinc finger domains having 87% identity, also suggesting that these two TFs may have similar functions. ZNF711 is not expressed in C42B, but it is expressed in MCF-7 with the expression slightly increasing upon knockdown of ZFX (Supplemental Fig. S3C). To investigate the possible functional redundancy of these two TFs, in an independent set of siRNA experiments than shown in Fig. 3, we knocked down ZFX, ZNF711, or both TFs simultaneously in MCF-7 (Supplemental Table S4C). We again observed an increase in ZNF711 expression when ZFX was knocked down (Fig. 4A). Several thousand genes changed upon knockdown of ZFX but very few genes changed upon knockdown of ZNF711. However, we detected more differentially expressed genes (n=1,847, FDR <0.05, fold change >1.5) upon knockdown of both ZFX and ZNF711 in MCF-7 cells than in the combined single knockdown experiments (Fig. 4B). In the double knockdown, we identified 371 additional downregulated genes that have ZFX bound to their promoters in MCF-7 cells (Fig. 4C). These findings suggest that ZNF711 may substitute for ZFX when ZFX expression is reduced by knockdown in MCF-7 cells; similar results were found when ZNF711 and ZFX were knocked down in HEK293T cells (Supplemental Fig. S4).

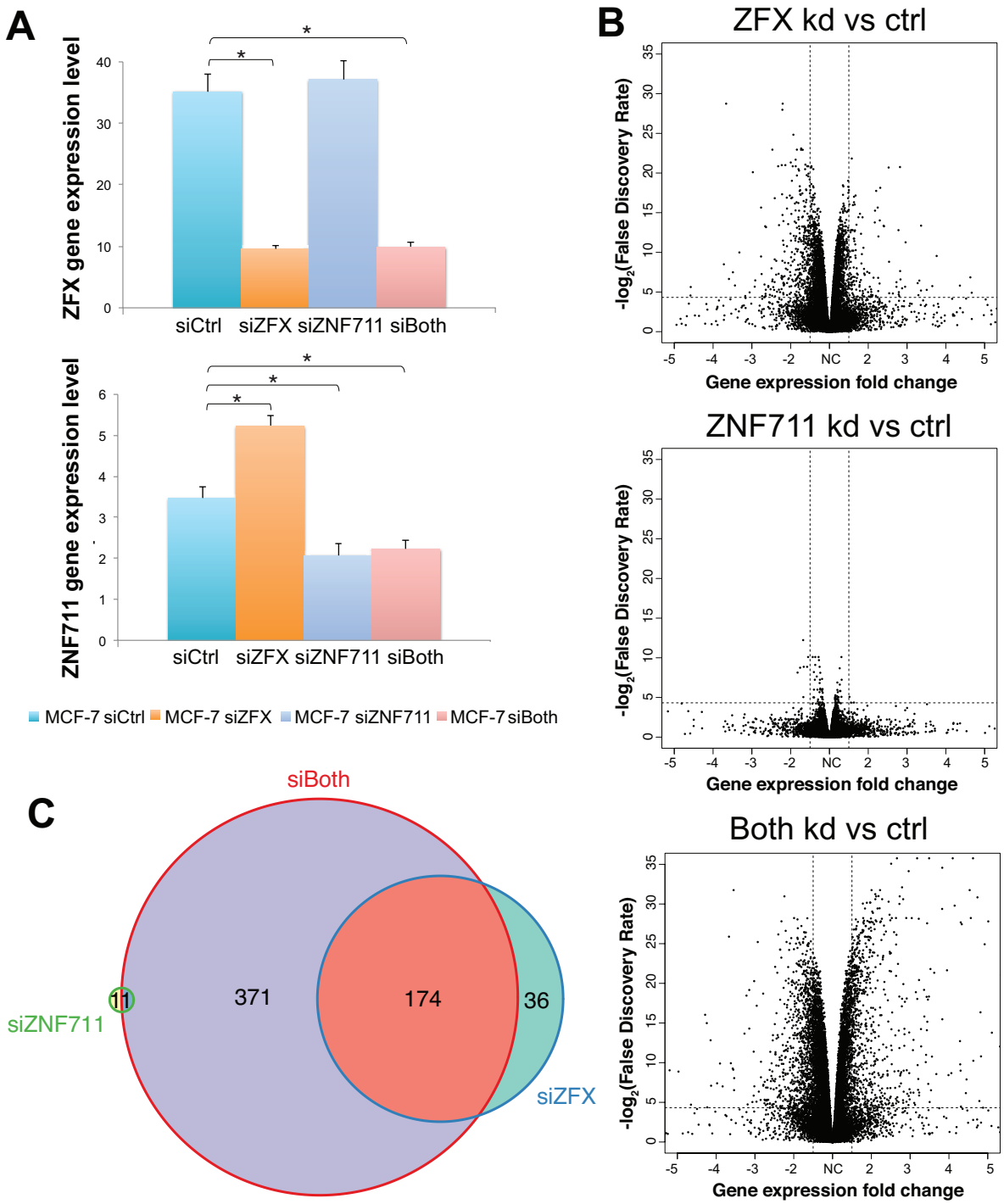


Figure 4. Combinatorial knockdown of ZFX and ZNF711 in MCF-7 cells. (A) ZFX and ZNF711 expression levels upon knockdown of ZFX, ZNF711 or both TFs in MCF-7. Comparisons of datasets that show a significant difference are marked with an asterisk (FDR <0.05). (B) Shown are volcano plots demonstrating differential gene expression after

knockdown of ZFX, ZNF711, or both TFs. (C) Shown is a comparison of downregulated genes having ZFX bound at their promoters after knockdown of ZFX, ZNF711, or both TFs.

To gain further support for the hypothesis that ZFX and ZNF711 are functionally redundant, we performed ChIP-seq for ZNF711 in MCF-7 cells. We identified 2,708 ZNF711 binding sites genome-wide (Supplemental Table S5), 98.6% of which overlapped with ZFX binding sites in MCF-7 cells. As expected, ZNF711 binding sites were also enriched at CpG island promoter regions, both in MCF-7 cells and in SH-SY5Y cells (Supplemental Fig. S5A-B). Unlike the ZFX binding pattern in the 4 cancer cell types, ZNF711 appears to have more cell type-specific binding sites (Supplemental Fig. S5). A comparison of ChIP-seq tags shows enrichment of ZNF711 at the promoter regions bound by ZFX, but the ZNF711 ChIP-seq signals are weaker than the ZFX ChIP-seq signals in MCF-7 cells (Fig. 5A), perhaps due to differences in expression levels of the two TFs (see Fig. 4A). Importantly, the ZNF711 binding sites are also enriched at +240bp downstream of the TSS (Fig. 5B). A comparison of the set of ZNF711-bound active promoters to the set of ZFX-bound active promoters revealed that 99% of promoters bound by ZNF711 are also bound by ZFX (Fig. 5C). This binding site redundancy supports the hypothesis that ZNF711 can substitute for ZFX when ZFX is knocked down. For example, *FAAH* and *LIPT2* show statistically significantly reduced expression in the double knockdown cells (FDR <0.05) but not in the single knockdown cells and the promoters of the *FAAH* and *LIPT2* genes are bound by both ZFX and ZNF711 in MCF-7 cells (Fig. 5D).

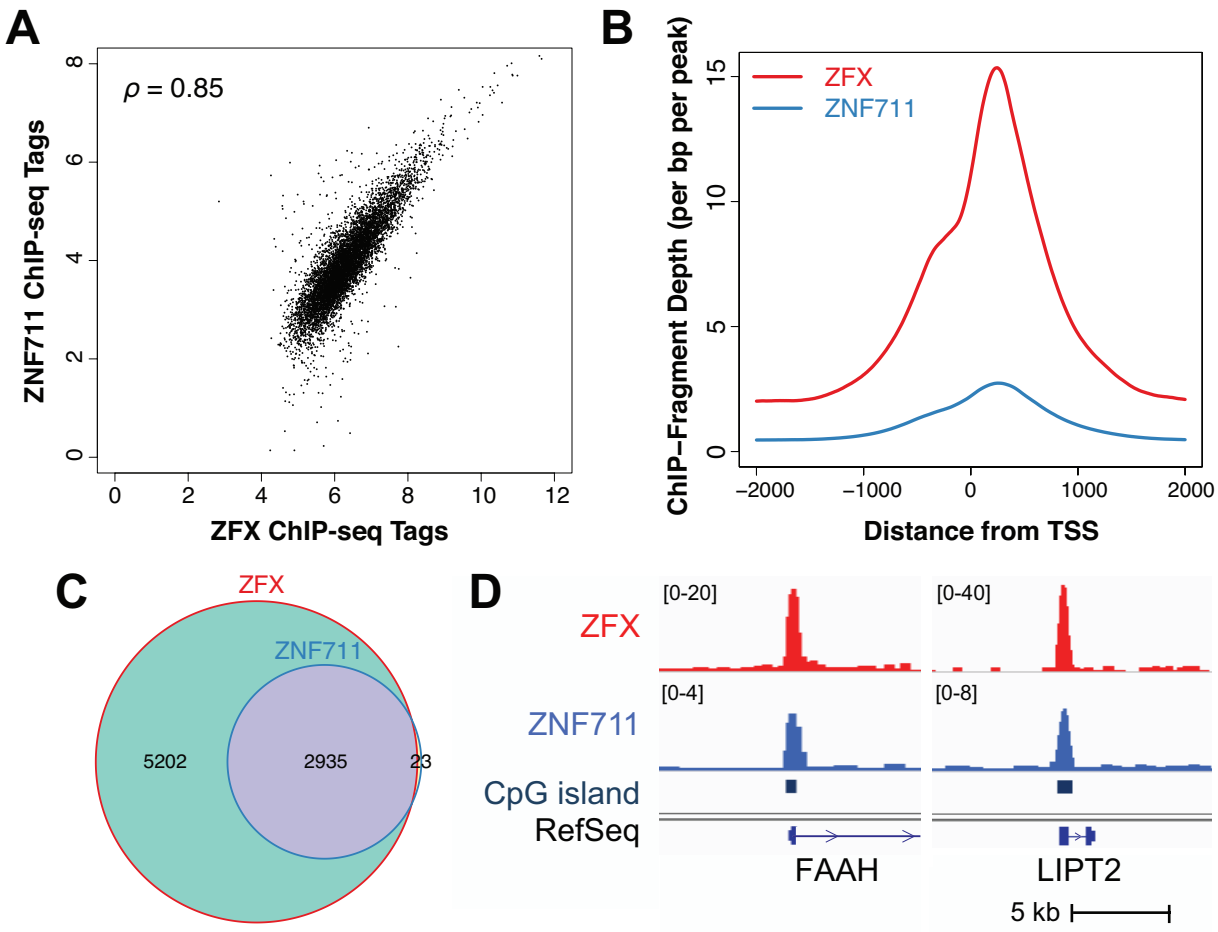


Figure 5. Comparison of ZFX and ZNF711 binding at promoters in MCF-7 cells. (A) Shown is a scatter plot of the normalized ZFX vs. ZNF711 ChIP-seq tags for the union set of ZFX and ZNF711 binding sites found in promoters (ρ = 0.85, Spearman's rank correlation coefficient). (B) Shown is the average ZFX (red) and ZNF711 (blue) ChIP-seq signal ±2kb from the TSS of promoters bound by ZNF711 in MCF-7. (C) Shown is a comparison of expressed genes having ZFX or ZNF711 bound at their promoters in MCF-7. (D) Shown are examples of ZFX and ZNF711 binding at CpG island promoters for two genes downregulated upon knockdown of both ZFX and ZNF711 in MCF-7.

ZFX binds adjacent to the first phased nucleosome downstream of the TSS.

As shown above, ZFX motifs are enriched at 240bp both upstream and downstream of the TSS but the majority of ZFX peaks are located at +240bp. However, it was possible that the

promoters activated by ZFX have a distinct binding pattern as compared to all ZFX peaks. Therefore, we compared ZFX binding patterns at promoters of all expressed genes vs. genes downregulated or upregulated upon ZFX knockdown (Fig. 6A). Interestingly, the downregulated genes (genes that may be directly activated by ZFX) show a nicely positioned ZFX bound at +240bp. In contrast, the upregulated genes (genes that may be repressed or indirectly regulated by ZFX) are not as highly enriched for ZFX at the +240bp position; rather there were more peaks at the -240bp position for upregulated genes (especially in MCF-7 cells). Taken together, these results suggest that ZFX may be a positive activator only when bound at +240bp of the TSS.

The preferred location of ZFX at +240bp is a unique position for a DNA binding TF. To further characterize the relationship between the bound ZFX and open chromatin surrounding the TSS, we used Nucleosome Occupancy and Methylome Sequencing (NOMe-seq). This genome-wide method identifies nucleosome-depleted regions (NDRs) and provides single molecule resolution for both accessibility and DNA methylation, which can very precisely identify specific TF binding sites (Kelly et al. 2012). When we used NOMe-seq to profile accessibility and DNA methylation in MCF-7 and C42B cells, we found that promoters bound by ZFX have a more accessible region with lower levels of DNA methylation near the TSS and more highly phased nucleosomes downstream of the TSS, as compared to promoters that are active in those cells but not bound by ZFX (Fig. 6B). Although ZFX ChIP-seq does not allow precise positioning of the bound ZFX, it appears that the summit of the ZFX peak is located in the NDR downstream of both the TSS and a bound RNA Polymerase II (RNAPII), just upstream of the first phased nucleosome (Fig. 6C-D). Indeed, >70% of ZFX peaks that have a summit near +240bp of the TSS overlapped with NDRs called by NOMe-seq in MCF-7 and C42B cells (Supplemental Table S6). We note that although Fig. 6D shows the pattern for all ZFX-bound promoters in MCF-7 cells, a similar pattern is also seen if the small subset of promoters bound by ZFX only in MCF-7 cells is analyzed (Supplemental Fig. S6).

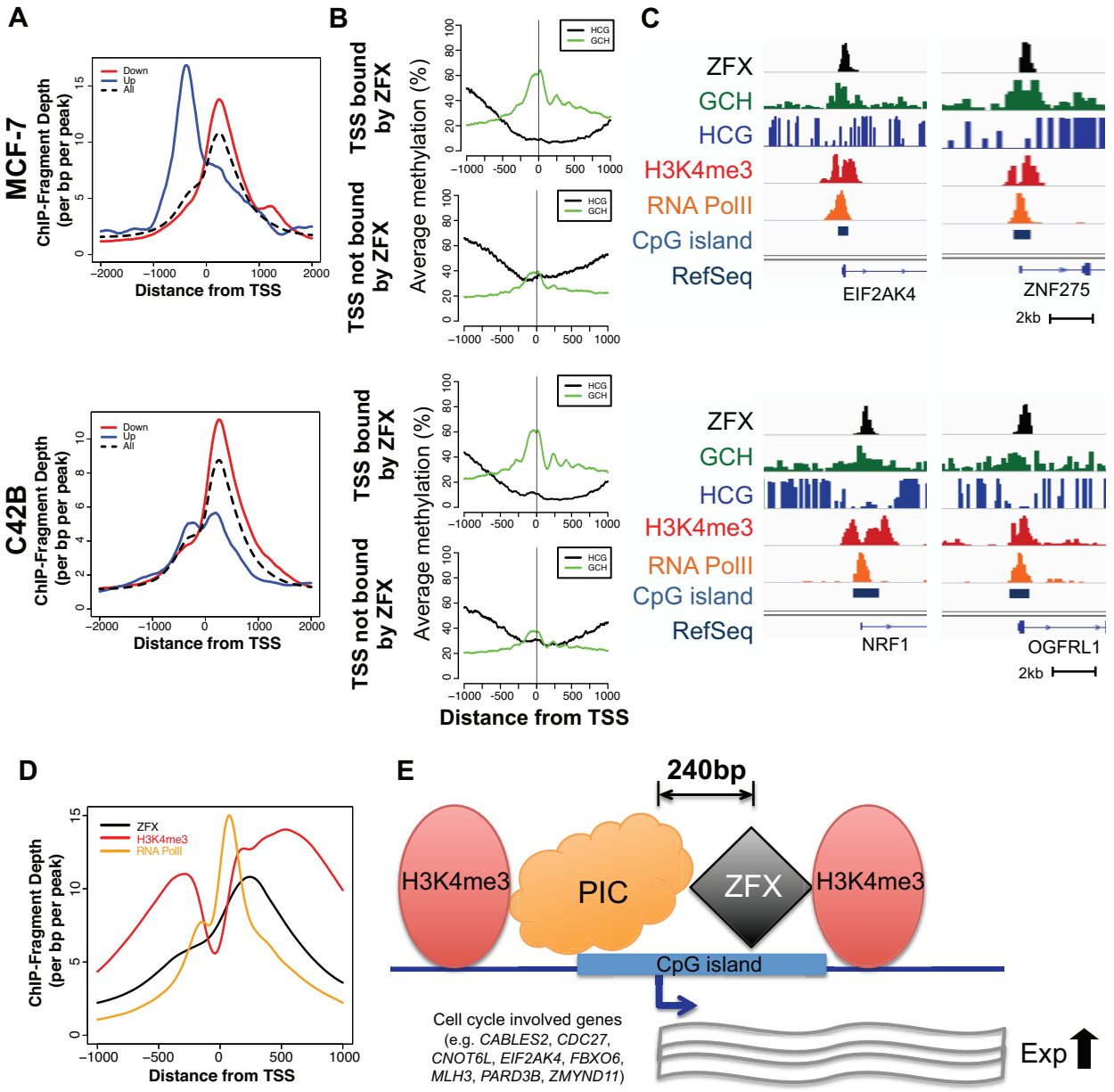


Figure 6. The relationship between ZFX binding and chromatin structure at promoters. (A) Shown are average ZFX ChIP-seq signals ±2kb from the TSS of downregulated (red), upregulated (blue), and all (black) genes bound by ZFX in MCF-7 (top) and C42B (bottom). (B) Shown is the endogenous DNA methylation (HCG) (black) and the accessibility (GCH) (green) levels from NOMe-seq data ±1kb from the TSS of active promoters bound by ZFX and from the TSS of active promoters not bound by ZFX in MCF-7 (top) and C42B (bottom). (C) Shown are examples of ZFX binding sites with ZFX ChIP-seq, NOMe-seq, H3K4me3 ChIP-seq, and RNA Polymerase II ChIP-seq signals in MCF-7 (top) and C42B (bottom). (D) Shown are the average

ZFX (black), H3K4me3 (red), and RNA Polymerase II (orange) ChIP-seq signals ±1kb from the TSS of genes bound by ZFX in MCF-7. (E) A model demonstrating the relationship of ZFX to other components of CpG island promoter structure. ZFX binds at +240bp in the nucleosome-depleted region of CpG island promoters, between the general transcription preinitiation complex (PIC) and the first nucleosome in the transcribed region. When ZFX is bound to this downstream site, it increases the expression levels of genes involved in cell proliferation; the wavy lines represent RNA levels.

DISCUSSION

We have profiled ZFX binding sites genome-wide in kidney, colon, prostate, and breast cancer cell lines. Unlike many oncogenic TFs that bind to distal elements, ZFX binds to the majority of CpG island promoters that are active in cancer cells, and many genes with promoters bound by ZFX were downregulated upon knockdown. Surprisingly, ZFX binds at +240bp downstream of the TSS of ZFX-regulated promoters, in the open chromatin region between the TSS and the first downstream nucleosome. Genome-wide analyses of open chromatin and DNA methylation demonstrate that promoters bound by ZFX have a more accessible region with lower levels of DNA methylation near the TSS and more highly phased nucleosomes downstream of the TSS, as compared to promoters that are active but not bound by ZFX. Taken together, these findings support the hypothesis that ZFX may act as a transcription activator and play an important role in maintaining a nucleosome-free promoter region and/or in positioning nucleosomes at many CpG island promoters in the human genome.

In accordance with findings from previous studies (Fang et al. 2012; Jiang et al. 2012a; Fang et al. 2014a; Yang et al. 2014; Yang et al. 2015), we found that the top categories of genes affected by ZFX knockdown are related to the cell cycle, to the DREAM complex (which contains E2F family members), and/or to genes regulated by E2F family members (Supplemental Fig. S7). For example, cell division cycle 27 homolog (*CDC27*), a component of the anaphase promoting complex/cyclosome that ubiquitinates Cyclin B (Lee and Langhans

2012), and MutL homolog 3 (*MLH3*), which is implicated in maintaining DNA replication and mismatch repair (Lipkin et al. 2000), both have a ZFX binding site downstream of the TSS and their expression levels are decreased upon ZFX knockdown in both MCF-7 and C42B. Thus, our results support the previous studies that ZFX expression is linked to cell proliferation. We also mapped ZFX binding sites in human normal prostate epithelial cells (PrEC) (Supplemental Fig. S8, Supplemental Table S3G). Although the ZFX binding pattern in normal prostate cells is very similar to the ZFX binding pattern in the prostate cancer cell line, the ChIP-seq peaks in PrEC were considerably smaller, suggesting that high ZFX expression in cancer cells may result in stronger binding and higher expression of genes involved in cell proliferation.

What distinguishes a "functional" bound ZFX from a "non-functional" bound ZFX?

Although ZFX binds to ~8,000-9,000 promoters in a given cell type, siRNA-mediated knockdown of ZFX resulted in altered activity of only a subset of these promoters. Although the ZFX motif is enriched symmetrically ±240bp from the TSS, our results suggest that ZFX acts as a transcriptional activator only when bound at +240bp. However, not all promoters with a ZFX bound at +240bp responded in the knockdown experiments. There are several possibilities that can explain why reduction of ZFX levels only affected a small percentage of promoters to which it is bound. First, it is possible that the incomplete knockdown of ZFX by siRNA treatment may have prevented the identification of all ZFX-regulated genes. In the future, knockout of ZFX by CRISPR/Cas9 could be performed to determine if a larger set of ZFX-regulated genes is identified upon complete removal of ZFX from the cell. It is also possible that co-binding of ZFX with other TFs is required for ZFX to regulate transcription. Finally, there is the possibility that other TFs are functionally redundant with ZFX. We tested the possibility that ZNF711, a TF that shares high homology and a similar DNA binding motif to ZFX, can substitute for ZFX. Indeed, ZNF711 binding sites are shared by ZFX binding sites and we identified several hundred additional ZFX-bound target genes that are downregulated upon knockdown of both ZFX and

ZNF711; perhaps complete loss of both proteins is required to observe the full effect of ZFX on the transcriptome.

How does ZFX regulate transcription of CpG island promoters from a downstream position?

There are two main types of transcriptional regulatory elements, promoters and enhancers. Unlike enhancers, which are located far from a TSS, are cell-type specific, and are closely linked to cellular identity (Rhie et al. 2014; Rhie et al. 2016), promoter elements are crucial for basal transcription of genes. The majority of human promoters are classified as CpG island promoters; these promoters are generally active in most cell types (Deaton and Bird 2011). Interestingly, ZFX is bound to most of the active CpG island promoters in a given cell. Other TFs have been shown to preferentially bind to CpG island promoters (Rozenberg et al. 2008; Jaeger et al. 2010; Landolin et al. 2010; Blattler et al. 2013). However, these CpG island-binding TFs tend to bind upstream of the TSS (Cao et al. 2011), whereas ZFX binds 240bp downstream of +1.

Comparison of the binding patterns of ZFX with RNAPII and H3K4me3 revealed that the bound ZFX is slightly downstream of the RNAPII signal and slightly upstream of the downstream peak of H3K4me3 signal (Fig. 6D). Although it is possible that ZFX regulates release of a paused RNAPII, factors implicated in this process are usually bound at +30 to +40bp relative to the TSS (Krumm et al. 1995; Shao and Zeitlinger 2017). We also note that it is unlikely that ZFX is involved in splicing, since the binding site can be in the first exon or at various places within the first intron, depending on the size of the first exon. Moreover, RNAPII and H3K4me3 signals are more enriched at ZFX-bound promoters than at promoters not bound by ZFX (Supplemental Fig. S9); these findings are consistent with a role for ZFX in transcriptional (not post-transcriptional) regulation. We note that ZFX does appear to be uniquely placed in relation to the phased nucleosomes located downstream of the TSS and ZFX-bound promoters have a more open region near the start site than do promoters that are active but not bound by ZFX.

Therefore, we postulate that perhaps ZFX is involved in creating a nucleosome-depleted region in CpG island promoters by recruiting the transcription pre-initiation complex and/or in positioning the downstream nucleosomes (Fig. 6E).

In conclusion, we have profiled ZFX binding sites genome-wide in kidney, colon, prostate and breast cancer cells and found that ZFX may function as a transcriptional activator, regulating as many as 60% of active CpG island promoters. Because tumor cells require abnormally high levels of transcription for their inappropriate proliferation and survival, increased overall transcription mediated by ZFX may explain why this TF has been correlated with poor prognosis for a variety of human cancers (Jiang and Liu 2015; Li et al. 2015; Yang et al. 2015; Yan et al. 2016). Future studies will focus on testing the hypothesis that ZFX contributes to overall high levels of transcription via a role in maintaining the large NDR found at the ZFX-bound promoters. Our demonstration that ZNF711, a TF highly related to ZFX, has a similar binding pattern suggests that we may have identified a new class of regulatory TFs. Further characterization of these TFs and their role in gene regulation will provide important new insights into transcription, chromatin structure, and the regulation of the cancer transcriptome.

METHODS

Cell culture. The human kidney HEK293T (ATCC# CRL-3216), colon HCT116 (ATCC# CCL-247), and breast cancer MCF-7 (ATCC# HTB-22) cells were obtained from ATCC (https://www.atcc.org/). The human prostate cancer C42B cells were obtained from ViroMed Laboratories (Minneapolis, MN, USA). The human normal prostate epithelial cells (PrEC) were obtained from Lonza (Cat# CC-2555, Lonza, Walkersville, MD, USA) The corresponding medium of each cell lines (DMEM for HEK293T, McCoy's 5A for HCT116, RPMI 1640 for C42B, DMEM for MCF-7) was supplemented 10% fetal bovine serum (Gibco by Thermo Fisher Scientific, Waltham, MA, USA) and 1% penicillin and streptomycin at 37ºC with 5% CO_2. PrEC cells were grown with PrEGM Bullet Kit (Prostate Epithelial Cell Growth Medium with

supplements), which were obtained from Lonza (Cat# CC-3166, Walkersville, MD, USA). All cell stocks except primary cells, PrEC were authenticated at the USC Norris Cancer Center cell culture facility by comparison to the ATCC and/or published genomic criteria for that specific cell line; all cells are documented to be free of mycoplasma. Pre-authentication was performed at Lonza (Walkersville, MD, USA) for PrEC and the first passage from the cultured cells were used for ChIP assay.

ChIP-seq. ZFX ChIP assays were performed in HEK293T, HCT116, C42B, MCF-7, and PrEC cells using a ZFX antibody (Cat# L28B6 Lot# 1, Cell Signaling Technology, Boston, MA, USA), according to ENCODE standards (Blattler et al. 2014). The ZFX antibody was validated using siRNAs, followed by western blots to demonstrate loss of the detected protein band (Supplemental Fig. S1). ZNF711 ChIP-seq experiments in MCF-7 cells were performed using antibodies from two different rabbits that were generated against ZNF711 amino acids 1-358; these antibodies have been previously used in ChIP-seq and were provided by Dr. Kristian Helin (Kleine-Kohlbrecher et al. 2010). H3K4me3 and RNAPII ChIP-seq experiments in C42B cells were performed using antibodies from Cell Signaling Technology (Cat# 9751S, Boston, MA, USA) for H3K4me3 and BioLegend (Cat# 664906, San Diego, CA, USA) for RNAPII. Each ZFX/ZNF711 ChIP-seq experiment in cancer cells was performed using two biological replicates and ChIP-seq libraries were sequenced on an Illumina HiSeq. All ChIP-seq data were mapped to hg19 and peaks were called using MACS2 (Zhang et al. 2008) with the IDR tool (https://github.com/nboley/idr) after preprocessing data with the ENCODE3 ChIP-seq pipeline (https://www.encodeproject.org/chip-seq/). ZFX and ZNF711 binding sites are listed in Supplemental Tables S3 and S5. A detailed description of ChIP-seq analyses can be found in Supplemental Methods.

Motif analyses. To discover de novo motifs enriched in the ChIP-seq peaks, we collected sequences of 20bp windows of the ZFX peak summits and used MEME version 4.10.1 (Bailey

and Elkan 1994) with a minimum motif width of 6 and a maximum motif width of 12, scanning both strands of the DNA sequences. The discovered motifs were very similar to the known motifs for ZNF711 and ZFX; AGGCCTAG motif found from HOMER (http://homer.ucsd.edu/homer/) (Heinz et al. 2010) was originally identified from ZNF711 ChIP-seq in SH-SY5Y (Kleine-Kohlbrecher et al. 2010) and ZFX ChIP-seq in mouse embryonic stem cells (Chen et al. 2008). Therefore, we used known motifs to scan ZFX and ZNF711 binding sites in 4 cell types using findMotifsGenome.pl script from HOMER to identify the enriched motifs and calculate the percentage of regions with the motifs (Fig. 2A). The motifs reported in Fig. 2A are the enriched motifs (FDR <0.05) found in more than 50% of ZFX peaks (sequences of 20bp windows of the ZFX peak summits) in each cell types. To further examine motif distribution in promoters, we compared the ZFX motif (AGGCCTAG), 10 randomly scrambled motifs having the same nucleotide composition as the ZFX motif, and the ETS motif (Supplemental Fig. S2).

siRNA knockdown, RT-qPCR, and RNA-seq. For transient knockdown, cells were transfected in triplicate with 100nM of siRNA oligonucleotides targeting human ZFX (Cat# L006572000005), ZNF711 (Cat# L008444020005) or control oligonucleotides (Cat# D0018101005) using SMART pool Dharmafect transfection reagent 3 (Cat# T200301) for C42B and reagent 1 (Cat# T200101) for MCF-7 (Dharmacon, Lafayette, CO, USA). Cells were incubated for 24 hours and transfected again with the same concentration of siRNAs and the incubation was continued for an additional 24 hours. RNA was extracted using TRIzol reagent (Cat# 15596-018, Thermo Fisher Scientific, NY, USA) following the manufacturer-suggested protocol. cDNA was synthesized using the SuperScript® VILO™ cDNA Synthesis Kit (Cat# 11754-050, Life technologies, Carlsbad, CA, USA). RNA-seq libraries were made using KAPA Stranded mRNA-Seq Kit with KAPA mRNA Capture Beads (Cat# KK8421, Kapa Biosystems, Woburn, MA, USA) and sequenced on an Illumina HiSeq. To remove batch effects, matched controls and knockdown samples were prepared and sequenced at the same time. Differentially expressed

genes were selected using the Gene Specific Algorithm from Partek® Flow® software with the upper quartile normalization method (Partek Inc., St. Louis, MO, USA). An FDR cut-off of 0.05 was used to select statistically significantly differently expressed genes. Differentially expressed genes with absolute fold change >1.5 are listed in Supplemental Table S4.

NOMe-seq. The NOMe-seq method is a combination of the genome-wide identification of open chromatin regions plus whole genome bisulfite sequencing (to identify methylated DNA). The first step of the method is based on the treatment of chromatin with the M.CviPI methyltransferase. This enzyme methylates Cs in the context of GpC dinucleotides. GpC^{m} does not occur in the human genome (the vast majority of DNA methylation in the human genome is at CpG dinucleotides, not GpC dinucleotides) and therefore there is no endogenous background of GpC^{m}. The enzyme can only methylate GpC dinucleotides that are accessible in the context of chromatin, i.e. not protected by nucleosomes or other proteins that are tightly bound to the chromatin. The second step of the method involves bisulfite treatment of the M.CviPI-methylated chromatin, which converts unmethylated Cs to Ts. This allows the distinction of GpC from GpC^{m} and CpG from C^{m}pG. Using this method, NDRs are defined as regions having increased GpC^{m} methylation over background (i.e. they are in open regions and thus were methylated by the M.CviPI enzyme) that are at least 140 bp in length. Because the bisulfite treatment also allows the distinction of CpG from C^{m}pG, the endogenous methylation status of the genome can also be obtained in the same sequencing reaction. It is important to note that in contrast to the induced GpC^{m,} which represents nucleosome-free, open chromatin that is available for TF binding, the endogenous C^{m}pG represents nucleosome-bound chromatin that is not available for TF binding. Open chromatin is expected to have high levels of GpC^{m} but low levels of C^{m}pG; thus, each of the two separate methylation analyses serve as independent (but opposite) measures which should provide matching chromatin designations (open vs. closed). C42B NOMe-seq data were generated as previously described (Rhie et al. 2018) and sequenced using an Illumina HiSeq 2000 PE 100bp to produce FASTQ files. FASTQ files of

MCF-7 NOMe-seq data were obtained from GSE57498 (Taberlay et al. 2014). Each FASTQ file was aligned to a bisulfite-converted genome (hg19), using BSMAP (Xi and Li 2009) and processed as previously described (Rhie et al. 2018). To identify the methylation status of CpG sites (in all HCG trinucleotides) and GpC sites (in all GCH trinucleotides) from the BAM file, the Bis-SNP (Liu et al. 2012) program was used and the Bis-tools was used to visualize DNA methylation and accessibility signals (https://github.com/dnaase/Bis-tools) (Lay et al. 2015). For identification of NDRs (Supplemental Table S6), the findNDRs function in the aaRon R package was used (https://github.com/astatham/aaRon).

GSEA and Gene ontology (GO) analysis. Differentially expressed genes upon ZFX knockdown are selected using FDR cutoff 0.05 and fold change cutoff 1.5 in C42B cells (Supplemental Table S4) and genes bound by ZFX were selected for GSEA and GO analysis. The differentially expressed genes were used to identify enriched gene sets using the GSEA (Gene Set Enrichment Analysis) tool (Subramanian et al. 2005). Hypergeometric test was used to calculate p-value, and false discovery rate (q-value) <0.05 was used to select significantly enriched gene sets. The same differentially expressed genes were analyzed for enrichment in particular GO categories using the TopGO program (https://bioconductor.org/packages/release/bioc/html/topGO.html). A Fisher's exact test was performed, and an adjusted p-value cutoff 0.05 was used to select statistically significantly enriched GO categories (Supplemental Fig. S7 and Supplemental Fig. S8E).

DATA ACCESS

All ChIP-seq, RNA-seq, and NOMe-seq data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE102616). Access to other publicly available datasets from GEO or ENCODE (Consortium 2012; Sloan et al. 2016) used in this study is detailed in Supplemental Table S1.

ACKNOWLEDGMENTS

We thank the ENCODE Consortium for data access, the USC/Norris Cancer Center Molecular Genomics core, the Stanford Sequencing Center, the UCLA Technology Center for Genomics & Bioinformatics, USC's Norris Medical Library bioinformatics service, and the USC Center for High-performance Computing (hpc.usc.edu). We also thank Kristian Helin for the ZNF711 antibody, Charlie Nicolet for assistance with the ZNF711 ChIP-seq experiments, and members of the Farnham lab for helpful discussions. This work was supported by the following grants from the National Institutes of Health: R01CA136924, U54HG006996 and P30CA014089.

AUTHOR CONTRIBUTIONS*:* Conceived and designed the experiments: SKR and PJF. Designed and performed experiments in cell lines: LY, ZL, HW, SS, YG, and AP. Performed data analysis: SKR and LY. Wrote and edited the manuscript: SKR and PJF.

REFERENCES

Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**: 28-36.

Blattler A, Yao L, Wang Y, Ye Z, Jin VX, Farnham PJ. 2013. ZBTB33 binds unmethylated regions of the genome associated with actively expressed genes. *Epigenetics & chromatin* **6**: 13.

Blattler A, Yao L, Witt H, Guo Y, Nicolet CM, Berman BP, Farnham PJ. 2014. Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. *Genome biology* **15**: 469.

Brayer KJ, Segal DJ. 2008. Keep your fingers off my DNA: protein-protein interactions mediated by C2H2 zinc finger domains. *Cell Biochem Biophys* **50**: 111-131.

Brown RS. 2005. Zinc finger proteins: getting a grip on RNA. *Curr Opin Struct Biol* **15**: 94-98.

Cao AR, Rabinovich R, Xu M, Xu X, Jin VX, Farnham PJ. 2011. Genome-wide analysis of transcription factor E2F1 mutant proteins reveals that N- and C-terminal protein interaction domains do not participate in targeting E2F1 to the human genome. *The Journal of biological chemistry* **286**: 11985-11996.

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J et al. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**: 1106-1117.

Consortium EP. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57-74.

Deaton AM, Bird A. 2011. CpG islands and the regulation of transcription. *Genes & development* **25**: 1010-1022.

Desjarlais JR, Berg JM. 1992. Toward rules relating zinc finger protein sequences and DNA binding site preferences. *Proceedings of the National Academy of Sciences of the United States of America* **89**: 7345-7349.

Fang J, Yu Z, Lian M, Ma H, Tai J, Zhang L, Han D. 2012. Knockdown of zinc finger protein, X-linked (ZFX) inhibits cell proliferation and induces apoptosis in human laryngeal squamous cell carcinoma. *Molecular and cellular biochemistry* **360**: 301-307.

Fang Q, Fu WH, Yang J, Li X, Zhou ZS, Chen ZW, Pan JH. 2014a. Knockdown of ZFX suppresses renal carcinoma cell growth and induces apoptosis. *Cancer Genet* **207**: 461-466.

Fang X, Huang Z, Zhou W, Wu Q, Sloan AE, Ouyang G, McLendon RE, Yu JS, Rich JN, Bao S. 2014b. The zinc finger transcription factor ZFX is required for maintaining the tumorigenic potential of glioblastoma stem cells. *Stem Cells* **32**: 2033-2047.

Grimmer MR, Stolzenburg S, Ford E, Lister R, Blancafort P, Farnham PJ. 2014. Analysis of an artificial zinc finger epigenetic modulator: widespread binding but limited regulation. *Nucleic acids research* **42**: 10856-10868.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**: 576-589.

Jaeger SA, Chan ET, Berger MF, Stottmann R, Hughes TR, Bulyk ML. 2010. Conservation and regulatory associations of a wide affinity range of mouse transcription factor binding sites. *Genomics* **95**: 185-195.

Jiang H, Zhang L, Liu J, Chen Z, Na R, Ding G, Zhang H, Ding Q. 2012a. Knockdown of zinc finger protein X-linked inhibits prostate cancer cell proliferation and induces apoptosis by activating caspase-3 and caspase-9. *Cancer Gene Ther* **19**: 684-689.

Jiang J, Liu LY. 2015. Zinc finger protein X-linked is overexpressed in colorectal cancer and is associated with poor prognosis. *Oncol Lett* **10**: 810-814.

Jiang M, Xu S, Yue W, Zhao X, Zhang L, Zhang C, Wang Y. 2012b. The role of ZFX in non-small cell lung cancer development. *Oncol Res* **20**: 171-178.

Jiang R, Wang JC, Sun M, Zhang XY, Wu H. 2012c. Zinc finger X-chromosomal protein (ZFX) promotes solid agar colony growth of osteosarcoma cells. *Oncol Res* **20**: 565-570.

Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, Jones PA. 2012. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome research* **22**: 2497-2506.

Kleine-Kohlbrecher D, Christensen J, Vandamme J, Abarrategui I, Bak M, Tommerup N, Shi X, Gozani O, Rappsilber J, Salcini AE et al. 2010. A functional link between the histone demethylase PHF8 and the transcription factor ZNF711 in X-linked mental retardation. *Molecular cell* **38**: 165-178.

Krumm A, Hickey LB, Groudine M. 1995. Promoter-proximal pausing of RNA polymerase II defines a general rate-limiting step after transcription initiation. *Genes & development* **9**: 559-572.

Landolin JM, Johnson DS, Trinklein ND, Aldred SF, Medina C, Shulha H, Weng Z, Myers RM. 2010. Sequence features that drive human promoter function and tissue specificity. *Genome research* **20**: 890-898.

Lay FD, Liu Y, Kelly TK, Witt H, Farnham PJ, Jones PA, Berman BP. 2015. The role of DNA methylation in directing the functional organization of the cancer epigenome. *Genome research* **25**: 467-477.

Lee SJ, Langhans SA. 2012. Anaphase-promoting complex/cyclosome protein Cdc27 is a target for curcumin-induced cell cycle arrest and apoptosis. *BMC Cancer* **12**: 44.

Li C, Li H, Zhang T, Li J, Ma F, Li M, Sui Z, Chang J. 2015. ZFX is a Strong Predictor of Poor Prognosis in Renal Cell Carcinoma. *Med Sci Monit* **21**: 3380-3385.

Li K, Zhu ZC, Liu YJ, Liu JW, Wang HT, Xiong ZQ, Shen X, Hu ZL, Zheng J. 2013. ZFX knockdown inhibits growth and migration of non-small cell lung carcinoma cell line H1299. *Int J Clin Exp Pathol* **6**: 2460-2467.

Lipkin SM, Wang V, Jacoby R, Banerjee-Basu S, Baxevanis AD, Lynch HT, Elliott RM, Collins FS. 2000. MLH3: a DNA mismatch repair gene associated with mammalian microsatellite instability. *Nature genetics* **24**: 27-35.

Liu Y, Siegmund KD, Laird PW, Berman BP. 2012. Bis-SNP: combined DNA methylation and SNP calling for Bisulfite-seq data. *Genome biology* **13**: R61.

Najafabadi HS, Mnaimneh S, Schmitges FW, Garton M, Lam KN, Yang A, Albu M, Weirauch MT, Radovani E, Kim PM et al. 2015. C2H2 zinc finger proteins greatly expand the human regulatory lexicon. *Nat Biotechnol* **33**: 555-562.

Nikpour P, Emadi-Baygi M, Mohammad-Hashem F, Maracy MR, Haghjooy-Javanmard S. 2012. Differential expression of ZFX gene in gastric cancer. *J Biosci* **37**: 85-90.

Rhie SK, Guo Y, Tak YG, Yao L, Shen H, Coetzee GA, Laird PW, Farnham PJ. 2016. Identification of activated enhancers and linked transcription factors in breast, prostate, and kidney tumors by tracing enhancer networks using epigenetic traits. *Epigenetics & chromatin* **9**: 50.

Rhie SK, Hazelett DJ, Coetzee SG, Yan C, Noushmehr H, Coetzee GA. 2014. Nucleosome positioning and histone modifications define relationships between regulatory elements and nearby gene expression in breast epithelial cells. *BMC genomics* **15**: 331.

Rhie SK, Schreiner S, Farnham PJ. 2018. Defining regulatory elements in the human genome using Nucleosome Occupancy and Methylome Sequencing (NOMe-seq). *Methods in molecular biology* **In press**.

Rozenberg JM, Shlyakhtenko A, Glass K, Rishi V, Myakishev MV, FitzGerald PC, Vinson C. 2008. All and only CpG containing sequences are enriched in promoters abundantly bound by RNA polymerase II in multiple tissues. *BMC genomics* **9**: 67.

Schreiber F, Patricio M, Muffato M, Pignatelli M, Bateman A. 2014. TreeFam v9: a new website, more species and orthology-on-the-fly. *Nucleic acids research* **42**: D922-925.

Shao W, Zeitlinger J. 2017. Paused RNA polymerase II inhibits new transcriptional initiation. *Nature genetics* doi:10.1038/ng.3867.

Sloan CA, Chan ET, Davidson JM, Malladi VS, Strattan JS, Hitz BC, Gabdank I, Narayanan AK, Ho M, Lee BT et al. 2016. ENCODE data at the ENCODE portal. *Nucleic acids research* **44**: D726-732.

Stubbs L, Sun Y, Caetano-Anolles D. 2011. Function and Evolution of C2H2 Zinc Finger Arrays. *Subcell Biochem* **52**: 75-94.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 15545-15550.

Taberlay PC, Statham AL, Kelly TK, Clark SJ, Jones PA. 2014. Reconfiguration of nucleosome-depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer. *Genome research* **24**: 1421-1432.

Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. 2009. A census of human transcription factors: function, expression and evolution. *Nat Reviews Genetics* **10**: 252-263.

Weng H, Wang X, Li M, Wu X, Wang Z, Wu W, Zhang Z, Zhang Y, Zhao S, Liu S et al. 2015. Zinc finger X-chromosomal protein (ZFX) is a significant prognostic indicator and promotes cellular malignant potential in gallbladder cancer. *Cancer Biol Ther* **16**: 1462-1470.

Wingender E, Schoeps T, Haubrock M, Donitz J. 2015. TFClass: a classification of human transcription factors and their rodent orthologs. *Nucleic acids research* **43**: D97-102.

Xi Y, Li W. 2009. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* **10**: 232.

Yan X, Shan Z, Yan L, Zhu Q, Liu L, Xu B, Liu S, Jin Z, Gao Y. 2016. High expression of Zinc-finger protein X-linked promotes tumor growth and predicts a poor outcome for stage II/III colorectal cancer patients. *Oncotarget* **7**: 19680-19692.

Yang F, Ma H, Feng L, Lian M, Wang R, Fan E, Fang J. 2015. Zinc finger protein x-linked (ZFX) contributes to patient prognosis, cell proliferation and apoptosis in human laryngeal squamous cell carcinoma. *Int J Clin Exp Pathol* **8**: 13886-13899.

Yang H, Lu Y, Zheng Y, Yu X, Xia X, He X, Feng W, Xing L, Ling Z. 2014. shRNA-mediated silencing of ZFX attenuated the proliferation of breast cancer cells. *Cancer chemotherapy and pharmacology* **73**: 569-576.

Yao L, Shen H, Laird PW, Farnham PJ, Berman BP. 2015. Inferring regulatory element landscapes and transcription factor networks from cancer methylomes. *Genome biology* **16**: 105.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**: R137.

Zhou Y, Su Z, Huang Y, Sun T, Chen S, Wu T, Chen G, Xie X, Li B, Du Z. 2011. The Zfx gene is expressed in human gliomas and is important in the proliferation and apoptosis of the human malignant glioma cell line U251. *J Exp Clin Cancer Res* **30**: 114.