10856–10868 *Nucleic Acids Research, 2014, Vol. 42, No. 16*
doi: 10.1093/nar/gku708

Analysis of an artificial zinc finger epigenetic modulator: widespread binding but limited regulation

Matthew R. Grimmer[1,2], Sabine Stolzenburg[3,4], Ethan Ford[5], Ryan Lister[5], Pilar Blancafort[3,4,6,*] and Peggy J. Farnham[1,*]

[1]Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA 90089, USA, [2]Integrated Genetics and Genomics, University of California-Davis, Davis, CA 95616, USA, [3]Department of Pharmacology, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA, [4]School of Anatomy, Physiology and Human Biology, M309, The University of Western Australia, Crawley, WA 6009, Australia, [5]Australian Research Council Centre of Excellence in Plant Energy Biology, The University of Western Australia, Perth, WA 6009, Australia and [6]Cancer Epigenetics Group, Harry Perkins Institute of Medical Research, Nedlands, WA 6008, Australia

Received June 14, 2014; Revised July 18, 2014; Accepted July 21, 2014

ABSTRACT

Artificial transcription factors (ATFs) and genomic nucleases based on a DNA binding platform consisting of multiple zinc finger domains are currently being developed for clinical applications. However, no genome-wide investigations into their binding specificity have been performed. We have created six-finger ATFs to target two different 18 nt regions of the human *SOX2* promoter; each ATF is constructed such that it contains or lacks a super KRAB domain (SKD) that interacts with a complex containing repressive histone methyltransferases. ChIP-seq analysis of the effector-free ATFs in MCF7 breast cancer cells identified thousands of binding sites, mostly in promoter regions; the addition of an SKD domain increased the number of binding sites ∼5-fold, with a majority of the new sites located outside of promoters. *De novo* motif analyses suggest that the lack of binding specificity is due to subsets of the finger domains being used for genomic interactions. Although the ATFs display widespread binding, few genes showed expression differences; genes repressed by the ATF-SKD have stronger binding sites and are more enriched for a 12 nt motif. Interestingly, epigenetic analyses indicate that the transcriptional repression caused by the ATF-SKD is not due to changes in active histone modifications.

INTRODUCTION

Dysregulation of normal epigenetic control of DNA methylation and histone modifications at gene promoters and enhancers is a hallmark of cancer (1). These regulatory layers are thought to cooperatively affect nucleosome positioning, chromatin structure and subsequent transcription factor accessibility to the genome (2). Epigenetic drugs targeting DNA methylation (3) and histone modifications (4–6) have shown promise in restoring some normal epigenetic features, but because they affect epigenome-regulating proteins, they are not sequence-specific. Just as modern small-molecule drugs can minimize classic chemotherapeutic side effects by specifically targeting disease-related proteins (7), artificial transcription factors (ATFs) may soon allow precise regulation of disease-related genetic loci. ATFs are proteins that can be engineered to bind and modulate the expression of specific genes through epigenetic effector domains (8,9). Epigenetic activation domains can be used to demethylate DNA (TET), recruit transcriptional machinery (VP64) or remove active histone marks (LSD). Conversely, epigenetic repression domains directly methylate DNA (DNMT3a) or recruit repressive histone modifiers (KRAB, SKD) (10). Genomic nucleases targeted by C2H2 zinc finger (ZNF) DNA binding domains were the first to be developed (11) and proteins composed of ZNF DNA binding domains fused to a nuclease have been employed in clinical trials (NCT00842634, NCT01044654 and NCT01252641).

Endogenous ZNF proteins, which served as a model for ATF development, typically bind CG-rich regions, underwent a drastic expansion in vertebrates, and are the largest family of transcription factors in humans (12). ZNF DNA binding domains are comprised of tandem arrays of several ZNFs, each specifically interacting with ∼3 nucleotides (nt) of DNA in an antiparallel orientation. Observations of endogenous ZNF binding specificities and *in vitro* refinements lead to the development of the ZNF code for targeting DNA sequences of choice (13–18). ATFs typically contain six C2H2 domains intended to target a unique 18 nt

*To whom correspondence should be addressed. Tel: 323 442 8015. Email: pfarnham@usc.edu
Correspondence may also be addressed to Pilar Blancafort. Tel: + 61 8 6151 0990. Email: pilar.blancafort@uwa.edu.au

genomic sequence. The majority of endogenous ZNFs also contain an amino-terminal effector domain, facilitating interactions with epigenetic modifiers; the most common ZNF effector domain in humans is the Krüppel-associated box (KRAB) domain (12). For example, the KRAB domain of ZNF274 recruits a complex, comprised of KAP1 and SETDB1, enriching the repressive histone modification, H3K9me3, at ZNF274 binding sites (19,20). Similarly, the addition of epigenetic effector domains to artificial ZNFs, creating ZNF 'toggle switches', allows modulation of the local chromatin structure at the intended target site. Unlike designer genomic nucleases or RNAi-based approaches, toggle switches can also upregulate silenced gene expression. For example, KRAB and VP64 effector domains were able to downregulate or upregulate the EGP-2 promoter, respectively (21).

A handful of genes have been successfully targeted in cancer cells and in tumors using ZNF toggle switches. The silenced tumor suppressor, *Maspin*, was upregulated using a VP64 activation domain (22–24) and the TET DNA demethylase was able to re-activate the *ICAM-1* promoter (25). A ZNF-DNMT3a fusion was used to force methylation and downregulate expression at the promoters of the *Maspin* gene and the oncogene *SOX2* (26). Stolzenburg *et al.* also targeted the *SOX2* promoter employing a super KRAB domain (SKD) to recruit repressive histone complexes (27). Effects from toggle switches alone have been fairly modest and no pattern has emerged to predict the efficacy of repression or activation. However, treatment of cells with broadly acting epigenetic drugs and toggle switches can have synergistic effects on target gene regulation (28). Recent studies of Cas9-based ATF toggle switches suggest that simultaneous epigenetic targeting of several neighboring sites may also have synergistic effects on gene regulation (29). This synergism suggests off-target effects of a single ATF may be minimal, but using multiple ATFs clearly requires monitoring genome-wide ATF binding patterns *in vivo*. Recent ChIP-seq studies of Transcription Activator-Like Effector (TALE)-based (30) and Cas9-based (31) ATFs have been informative, but to date, no genome-wide binding analysis of a ZNF-based ATF has been reported.

In this study, we examine the genome-wide binding patterns of two previously reported ZNF effectors, 598-SKD and 552-SKD, which target the *SOX2* promoter (26,27). As 598-SKD was more effective in downregulating *SOX2* expression (27), we also profiled effects of this ATF on histone modifications and genome-wide expression. We found that ZNF-based ATFs bind very broadly, each one binding to thousands of promoters. However, the activity of only ∼10% of bound promoters was significantly altered by 598-SKD and changes in histone marks do not correlate with expression changes. Our results demonstrate a large disconnect between binding and activity for ZNF-based ATF toggle switches.

MATERIALS AND METHODS

Cell growth conditions

ATF plasmids were designed and stably integrated into MCF7 cells as described in (27). Stable lines were grown at 30–80% confluency in Dulbecco's Modified Eagle's Medium (Corning, Corning, NY) supplemented with 10% heat-inactivated fetal bovine serum (Invitrogen, Life Technologies, Grand Island, NY) and 1% penicillin/streptomycin; cells were selected using 5 μg/ml puromycin (VWR, Radnor, PA) and 200 μg/ml G418 (VWR, Radnor, PA). ATF expression was induced by treatment with media containing 1μg/ml doxycycline (VWR, Radnor, PA) at 0 h, doxycycline media was refreshed at 48 h, and cells were harvested at 72 h. ATF expression was confirmed by hemagglutinin (HA) tag western blot prior to HA ChIP-seq, histone ChIP-seq and RNA-seq analysis.

ChIP-seq

For HA-tag ChIP-seq, stable MCF7 cell lines were induced using 100 ng/ml doxycycline (Sigma) at 0 h, doxycycline media was refreshed at 48 h, and cells were harvested at 72 h by crosslinking in a final concentration of 1% formaldehyde. Crosslinking was stopped after 5 min by adding glycine to a final concentration of 125 mM. Crosslinked cells were washed in cold phosphate buffered saline, lysed using 1 ml low-salt IP buffer (150 mM NaCl, 50 mM Tris-HCl (pH7.5), 5 mM EDTA, NP-40 (0.5%), Triton X-100 (1%) containing protease inhibitors) and aliquoted at 1 × 10^7 cells/ml. Cells are then sonicated to a fragment size range of 500–800 bp. Samples were then diluted in 1 ml low-salt buffer and incubated with 3 μl of anti-HA antibody (Covance, Princeton, NJ). Three-hundred microliter Sepharose A beads (GE Healthcare Life Science) were used for pull-down. Samples were sequenced at the UNC-CH Genome Analysis Facility (Chapel Hill, NC) on an HiSeq (Illumina, San Diego, CA) to read counts of 4.1–67.3 M total reads. For histone ChIP-seq, antibodies to H3K4me3 (Cell Signaling Technologies CST9751), H3K9Ac (Cell Signaling Technologies CST9649) and H3K9me3 (Diagenode pAb-056-050) were used; samples were prepared as previously described (32).

ChIP-seq analysis

Each sequenced HA ChIP-seq biological replicate for HA-598 (two replicates), HA-598-SKD (three replicates), HA-552 (two replicates), HA-552-SKD (three replicates) and pT3-empty-vector (two replicates) yielded standard fastq file outputs. Fastq files were mapped to human genome version GRCh37/hg19 using Bowtie 2.0.1 (33) with default parameters to produce SAM files. SAM files were converted to BAM files using SAMTools (34). BAM files were converted to BED files using BEDTools (35). In order to call ChIP-seq peaks, BED files were submitted individually (each with a similarly mapped MCF7 input BED file as a background reference) to the Sole-Search Tool (36) with the following parameters: permutation number, 5; average chromatin fragment size, 150; alpha value, 0.001; peak merge distance, 0; histone blur, no. Resulting GFF format peak files were sorted by peak height and truncated to the number of peaks in the smallest replicate for each set. The top 40% of peaks for each truncated replicate were tested for statistically significant overlap versus corresponding replicate total peak files using the Sole-Search GFF-overlap Tool. All 40% replicate peak

files satisfied the ENCODE standard of 80% overlap with truncated corresponding replicates (https://genome.ucsc.edu/ENCODE/protocols/dataStandards/2010-05-30_modENCODE_TF_Chrom_Data_Standards.pdf). Corresponding replicates were then submitted simultaneously as merged sets to the Sole-Search Tool for peak calling, producing sets of merged GFF peak files and SGR visualization files. Again using the Sole-Search GFF-overlap Tool, peaks for each ATF variation that overlapped with pT3-empty-vector GFF files or with regions of aneuploidy in MCF-7 cells were removed to eliminate possible artifacts not resulting from ATF binding. Remaining ATF peak sets were sorted by peak height and plotted against peak rank to select high-confidence (HC) peaks (Supplementary Figures S1A and B), along with visual inspection of SGR format visualization files. The Homer genomic suite of tools (37) was used for tag density plots (annotatePeaks.pl), scatter plots (annotatePeaks.pl), overlap analysis (mergePeaks), genomic locations (annotatePeaks.pl) and *de novo* motif finding (findMotifsGenome.pl). *De novo* motif finding for the top 4000 peaks from each HC ATF peak set was performed using all hg19 RefSeq gene promoters as a background for statistical comparison and all nucleotides within the peaks, as called by the Sole-Search tool, were subjected to the analysis. The resulting predominant motif position weight matrix files for each ATF were used to search for the presence of significant motif matches among the top 2500 HC peaks for each ATF. The predominant motif position weight matrix files were restricted to single nucleotide tolerances to represent partial target sequences in order to search the entire genome for instances of partial target sites using annotatePeaks.pl. Histone ChIP-seq enrichment values for all RefSeq promoter regions (TSS ± 1 kb) were calculated from the average tags per bp using annotatePeaks.pl within each promoter region. Tag density plots for histone enrichment at different promoter categories (downregulated, unchanged, upregulated and unbound) were calculated using the -hist 200 option of annotatePeaks.pl. H3K4me3 plots and enrichments before and after induction were compared to ENCODE datasets for H3K4me3 in MCF7 (wgEncodeEH000967) to ensure that uninduced profiles are representative of MCF7 cells.

Whole genome bisulfite sequencing

MCF7:598-SKD cells were either not induced or induced with doxycycline as described above. Two microgram of genomic DNA was isolated using a QIAeasy DNA mini kit (Qiagen, Venlo, The Netherlands) and sonicated using a Covaris to an average molecular weight of 150 bp. Achievement of the desired size range was verified by Bioanalyzer (Agilent) analysis. Fragmented DNA was repaired to generate blunt ends using the END-It kit (Epicentre Biotechnologies, Madison, WI) according to manufacturer's instructions. Following incubation, the treated DNA was purified using AmpureX beads from Agencourt. In general, magnetic beads were employed for all nucleic acid purifications in the following protocol. Following end repair, A-tailing was performed using the NEB dA-tailing module according to manufacturer's instructions (New England Biolabs, Ipswich, MA). Adapters with a 3′ 'T' overhang were then ligated to the end-modified DNA. For whole genome bisulfite sequencing, modified Illumina paired-end (PE) adapters were used in which cytosine bases in the adapter are replaced with 5-methylcytosine bases. Depending on the specific application, we utilized either Early Access Methylation Adapter Oligos that do not contain barcodes, or the adapters present in later versions of the Illumina DNA Sample Preparation kits, which contain both indices and methylated cytosines. Ligation was carried out using ultrapure, rapid T4 ligase (Enzymatics, Beverly, MA) according to manufacturer's instructions. The final product was then purified with magnetic beads to yield an adapter-ligation mix. Prior to bisulfite conversion, bacteriophage lambda DNA that had been through the same library preparation protocol described above to generate adapter-ligation mixes was combined with the genomic sample adapter ligation mix at 0.5% w/w. Adapter-ligation mixes were then bisulfite converted using the Zymo DNA Methylation Gold kit (Zymo Research, Orange, CA) according to the manufacturer's recommendations. Final modified product was purified by magnetic beads and eluted in a final volume of 20 μl. Amplification of one-half the adapter-ligated library was performed using Kapa HiFi-U Ready Mix for the following protocol: 98° 2′; then six cycles of 98° 30″, 65° 15″, 72° 60″; with a final 72° 10′ extension, in a 50 μl total volume reaction. The final library product was examined on the Agilent Bioanalyzer, then quantified using the Kapa Biosystems Library Quantification kit according to manufacturer's instructions. Optimal concentrations to get the right cluster density were determined empirically but tended to be higher than for non-bisulfite libraries. Libraries were plated using the Illumina cBot and run on the Hi-Seq 2000 according to manufacturer's instructions using HSCS v 1.5.15.1. Image analysis and basecalling were carried out using RTA 1.13.48.0; deconvolution and fastq file generation were carried out using CASAVA_v1.7.1a5. Raw reads were mapped with [programs/parameters], and percent methyl-C/C was calculated for every CpG dinucleotide in the human genome. All CpG dinucleotides with a minimum sequencing coverage of 3× were used for downstream analyses. To determine the average DNA methylation surrounding a region of interest, the annotatePeaks tool of the HOMER suite was used with the '-ratio' option and a bin size of 200 bp.

RNA-seq

MCF7 cell lines were grown as described above. Each experiment (no induction control and doxycycline induction) was performed in triplicate. RNA was harvested using the Qiagen RNeasy kit and Illumina libraries were made with the Illumina TruSeq Stranded mRNA Sample Prep kit according to the manufacturer's instructions. DNA libraries were sequenced on an Illumina HiSeq1500 with read lengths of 96 bp and read counts of 26–57 million. Prealignment QA/QC and base trimming were performed in Partek Flow v3.0.14.0321 using fastqc and quality_trim. Reads were mapped to the Gencode version 19 of the human hg19 reference genome using Tophat (38). Quantification to the transcriptome was performed by Partek expectation maximization (EM). The Partek Flow software

suite was made available through the USC Norris Medical Library Bioinformatics Services (http://norris.usc.libguides.com/nml-bioinfo). Genes with a false discovery rate less than 0.05 and a change in expression greater than 2-fold were counted as differentially expressed.

Data access

The MCF7 DHS data (UCSC accession #wgEncodeEH000502) and H3K27ac MCF7 (UCSC accession # wgEncodeEH002872) are from the ENCODE Consortium (39–42) and are available at http://genome.ucsc.edu; all ENCODE data used in this study are past the 9 month moratorium. The other ChIP-seq data, the DNA methylation data, and the RNA-seq data have been submitted to GEO (GSE59980).

RESULTS

Six-finger ATFs display widespread binding to the human genome

For our studies, we used MCF7 cell lines stably expressing four different six-finger ATFs under the control of a doxycycline-inducible promoter (27). Each ATF should specifically recognize a unique 18 nt region of the human *SOX2* promoter located 552 or 598 bp upstream of the transcription start site (TSS) (Figure 1A). Due to the nature of the ZNF code (15,43–44), ATFs based on ZNF domains recognize GC-rich motifs. The ATFs either lack or contain an SKD, a transcriptional repression domain that has been shown to recruit the KAP1/SETDB1 histone methyltransferase which trimethylates histone H3 on lysine 9 (19). Each ATF also contains a HA-tag, which serves as an antibody recognition domain for chromatin immunoprecipitation (ChIP) assays. Previous studies had used ChIP-PCR to demonstrate that 552-SKD and 598-SKD can bind to the *SOX2* promoter (26,27). However, the previous ChIP-PCR was limited to specific genomic regions and the overall specificity of binding was not examined for the ATFs.

Therefore, we performed multiple, independent ChIP-seq assays for each of the four ATFs (552, 552-SKD, 598 and 598-SKD) using an antibody that recognizes the HA-tag. As a control, a fifth MCF7 line with stable integration of the empty ATF vector was also induced with doxycycline (Dox) and used for background peak subtraction. As expected, each of the ATFs bound to the correct site within the *SOX2* promoter (Figure 1B). However, binding was not restricted to this promoter region. In fact, thousands of binding sites were detected for the 552 and 598 ATFs, and the addition of the SKD effector domain increased the number of sites for each ATF by about 5-fold (Figure 1C). Although the peak at the *SOX2* promoter was among the top of the ranked peaks (e.g. having a rank of 258 out of the 25 000 598-SKD peaks), there were thousands of sites of similar height. The ChIP-seq peak sets for the four ATFs, including the new effector-specific sites, were robust and reproducible (Supplementary Figures S1 and S2). Conservative, HC peak sets for each ATF (cutoffs indicated by the colored circles in Figure 1C) were chosen for further analysis. A comparison of the 552 versus 552-SKD and 598 versus 598-SKD sites revealed that ∼90% of the sites bound by the effector-free ATFs were also bound by the SKD-containing ATFs (Figure 1D), again showing the reproducibility of the ChIP-seq signals.

All ATF variants bind partial target sequences

As noted above, the 18 nt target sites were expected to be uniquely bound in the human genome. Comparison of these sequences to the hg19 reference genome revealed that 552 occurs only at chr3:181429514–181429714 and 598 occurs at chr3:181429469–181429669 and chr11:82443421–82443621. However, we identified thousands of binding sites for each ATF. We note that both 552-SKD and 598-SKD have more sites than either of the effector-free parental ZNFs. We realize that it is difficult to know whether the thousands of extra sites bound by the 552-SKD and 598-SKD proteins are a result of differences in the expression level and/or DNA binding ability of the ZNFs that contain or lack an effector domain. We reasoned that if all the 'extra' peaks in the ZNF-SKD constructs were due to higher expression levels allowing occupancy of thousands of additional low affinity sites, perhaps the SKD-unique peaks would be smaller than those peaks also bound by the effector-free ZNFs (i.e. the common peaks). However, this is not the case; ∼3000 of both the 552-SKD unique peaks and 598-SKD unique peaks are equivalent to or higher than the median peak height from the 5138 common peaks. Therefore, it is possible that at least some of the ZNF-SKD peaks could bind to sites other than the *SOX2* target site by recruitment to the genome via protein–protein interactions of the SKD domain with complexes containing other site-specific DNA binding proteins. As noted above, the SKD domain is a potent KRAB domain taken from the KRAB-ZNF transcription factor KOX1 (45). A well-characterized function of the KRAB domains from KRAB-ZNF proteins is to interact with KAP1. Although there are many potential KRAB-ZNFs that might interact with KAP1, in fact, only one (ZNF274) has been identified to co-localize with KAP1 on the genome (19–20,46). The most robust ZNF274/KAP1 recruitment sites are at the 3′ exons of ZNF genes on chromosome 19. Interestingly, KAP1 functions as a trimer and therefore could possibly interact with ZNF274 and with the ZNF-SKD proteins simultaneously. If so, then perhaps the KAP1 protein was interacting with both the ATF-SKD and ZNF274, with the DNA-binding properties of ZNF274 bringing the ATF-SKD to genomic regions lacking the 598 or 552 motifs. Therefore, we examined whether the ZNF-SKD proteins can be found at the normal KAP1 binding locations by comparing the binding patterns of SKD-552, SKD-598 and ZNF274. ZNF274 ChIP-seq has not been performed on MCF7 cells. However, ZN274 sites vary minimally across all analyzed cell lines. Therefore, we used a pan-ZNF274 site dataset, encompassing all reproducible ZNF274 binding sites from 11 different normal and cancer cell lines. We found that only three out of 25 000 552-SKD peaks and two out of 25 000 598-SKD peaks overlap with ZNF274 binding sites at ZNF 3′ exons (Supplementary Figure S3). Therefore, the large number of observed effector-specific sites were not due to co-localization with the ZNF274/KAP1 complex. However, this does not eliminate the possibility that the ATF-SKDs are being recruited

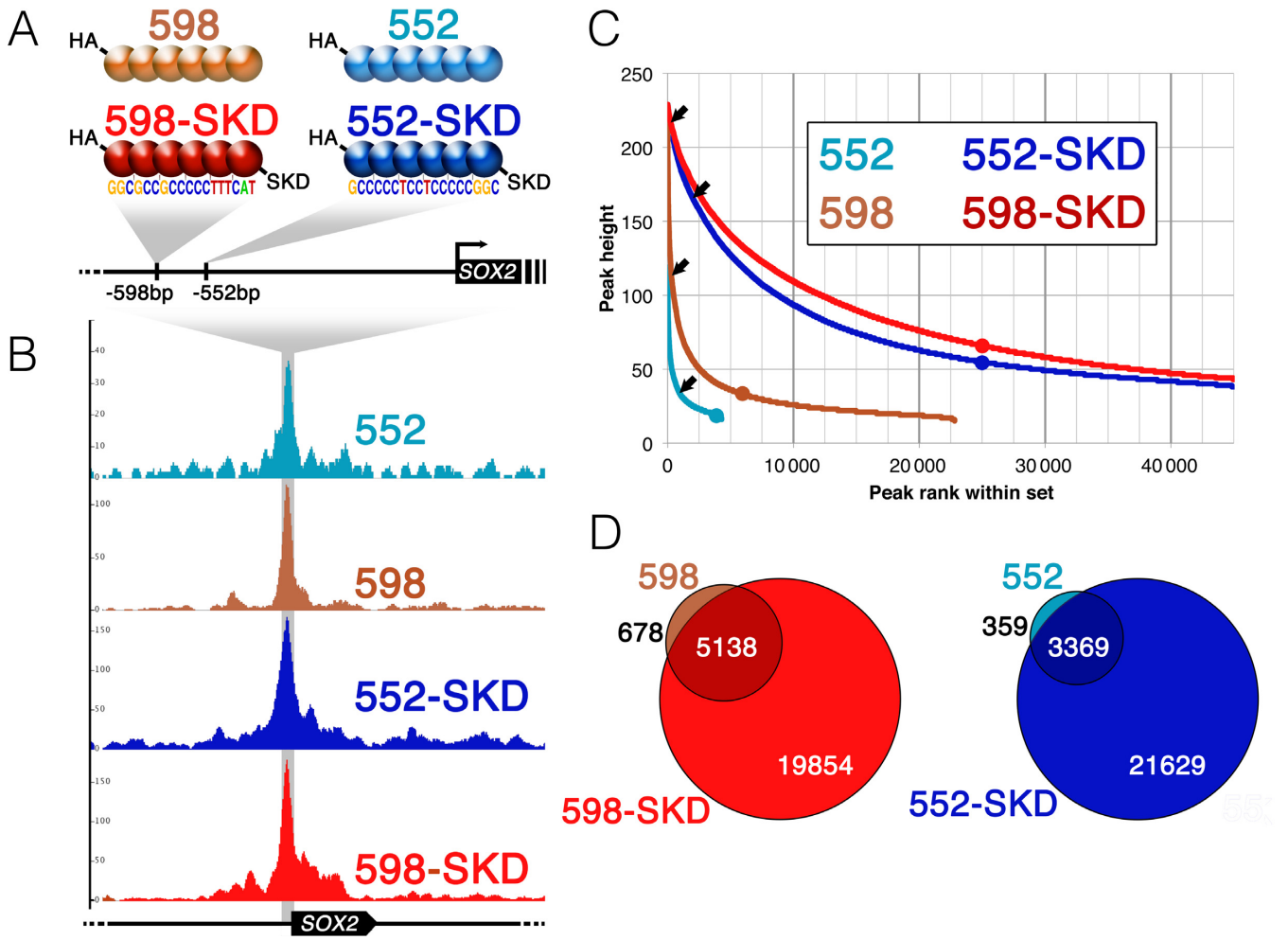

Figure 1. ChIP-seq analysis of ATFs directed to the *SOX2* promoter. (**A**) Tandem arrays of six zinc finger DNA binding domains were engineered to bind to two different sites, 598 or 552 bp upstream of the transcription start site of the *SOX2* gene. Each zinc finger array carries a carboxy-terminal HA-tag and lacks or contains an amino-terminal SKD repression domain. (**B**) ChIP-seq peaks for the four ATFs at the *SOX2* gene are shown; the position of the *SOX2* gene is indicated below the tracks and the peak height (number of sequenced tags) is indicated on the Y-axis. (**C**) ChIP-seq inflection curves for the four ATFs are shown; peak height is shown on the Y-axis and peak rank is shown on the X-axis. The approximate position of the peak at the target site in the *SOX2* promoter is indicated by a black arrow for each ATF. For example, the *SOX2* peak was ranked 275 out of the entire >45 000 sites in the 598-SKD peak set. High-confidence peak cutoffs are indicated by colored circles along the curves. (**D**) Venn diagrams comparing the high-confidence peak sets for the no-effector versus SKD domain-containing 552 and 598 ATFs.

by KAP1 in complex with other (to date uncharacterized) cellular KRAB-ZNFs to sites that are mutually exclusive from ZNF274-KAP1 genomic sites.

Another possibility for promiscuous binding is that the ATFs are using only a subset of their six fingers to contact the genomic DNA. To address this possibility, we performed *de novo* motif analysis of the top 4000 peaks for each ATF, identifying a predominant motif for each peak set (Figure 2A). Clearly, the 552- and 598-based ATFs frequently bound to sequences resembling a subset of the 18 nt in the target sequences and the addition of the SKD domain did not greatly affect the predominant motif for either 552 or 598. A majority of the top ranked sites for each ATF contain the identified predominant submotif; no other significant motifs were detected, even in the sites that lacked a strong match to the predominant motif (Supplementary Figure S4). The percentages of total binding sites corresponding to the partial motifs are 59%, 50%, 39% and 30% for 598, 552, 598-SKD and 552-SKD, respectively. Effector-free ATF peak height correlated more strongly with partial motif presence, suggesting greater dependence on the partial motif for binding than the ATF-SKDs. Comparison to the target sequence suggests that the 552 ATFs were mainly utilizing fingers 2–5 and the 598 ATFs were mainly utilizing fingers 3–6. Unlike the high specificity of the 18 nt motifs to which the ATFs were designed to bind, these 12 nt partial target subsequences occur respectively 1171 and 6532 times in the genome (Figure 2B). A small number of these 12-mers are found in regions of open chromatin, identified as a DNAse I hypersensitive site (DHS). Of these, between 40–70% are bound by an ATF. However, the vast majority of the 12-mer motifs are found in a non-DHS, which are usually inaccessible to most transcription factors. In fact, an analysis of 69 TFs in six cell lines by the ENCODE Consortium indicated that over 98% of all transcription factor ChIP-seq peaks fall within a DHS (47,48).

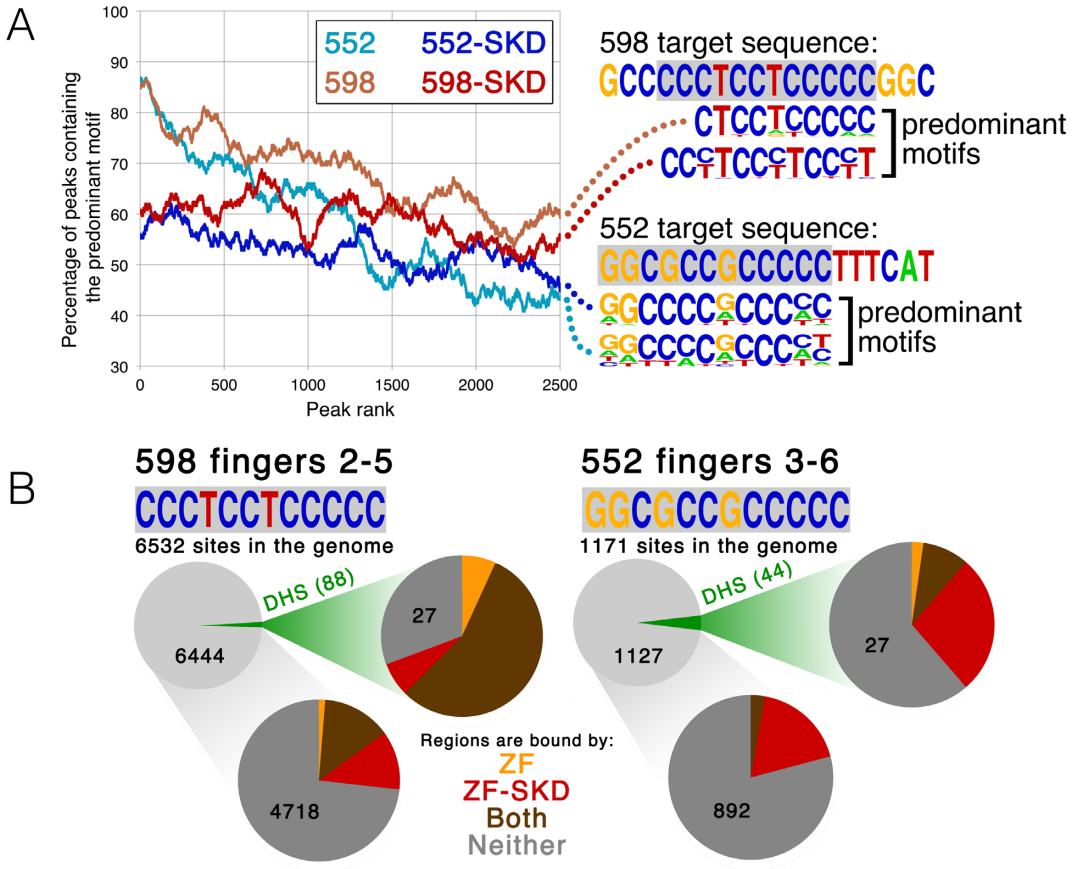

Figure 2. The ATFs bind to partial target sequence motifs. (**A**) Motif density plots of the percentage of the top 2500 peaks that contain the predominant motif (as determined by *de novo* motif analysis) for each ATF. 598 and 552 design target sequences are shown, compared to the predominant motifs identified by *de novo* motif analysis of the top 4000 peaks for each ATF. (**B**) Analysis of ATF binding to the 598 and 552 partial target sequences found throughout the genome, divided into regions of open chromatin (DHS) or other regions not overlapping open chromatin.

Surprisingly, 21% of all possible 12-mer partial 552 target sequences in the genome and 27% of all partial 12-mer 598 target sequences in the genome found in non-DHS regions are bound by at least one form of the corresponding ATF. For the case of 598-SKD, more than 1700 binding sites corresponded to regions that contained the 12-mer but were not identified as a DHS.

ATF binding in relation to genomic and epigenomic structure

Our analyses which revealed that some of the bound 12-mer motif-containing sites are not within a DHS suggested that the ATFs have the ability to bind to nucleosome-free (characterized by DHS) and nucleosome-containing sites. To further characterize all binding sites of the ATFs, we determined their genomic locations relative to promoter regions, enhancer regions and DHS (Figure 3A). The majority of the binding sites for 552 and 598 were in promoter regions (as defined by ± 1000 nt from a TSS); we note that the binding site at the *SOX2* promoter was ranked 55 out of the 5917 promoters bound by 598-SKD. Because the majority of the promoters in the human genome are GC-rich, we thought that perhaps the large number of promoter peaks in the ChIP-seq data (ranging from ∼2500 in 552 to ∼8000 in 552-SKD) was due to the presence of multiple 12-mer binding sites and therefore multiple peaks in a given promoter region. However, as shown in Figure 3B, the number of uniquely identified promoters was similar to the total number of promoters identified for each ATF. This suggests that the ATFs have the potential to regulate up to 8000 promoters in MCF7 cells.

Interestingly, although the number of total binding sites increased ∼5-fold when the effector domain was added, most of the newly acquired sites were not in promoter regions. Furthermore, promoter peaks were distributed somewhat evenly among peaks, with little correlation to peak rank (Supplementary Figure S5), suggesting that many robust binding sites were outside of promoters. TFs can bind to both promoters and enhancers and we therefore hypothesized that the new sites were located at enhancer regions. Using H3K27Ac ChIP-seq data from the ENCODE project, we compared the location of the binding sites for the four ATFs to active enhancers in MCF7 cells (49). However, very few of the ATF binding sites were in active enhancer regions. As noted above, most transcription factor binding sites fall within a DHS (47,48). DHS regions encompass not only promoters and active enhancers, but also include a large set of regions of open chromatin that are available for transcription factor binding but the transcription factors that are bound are not sufficient to recruit activating

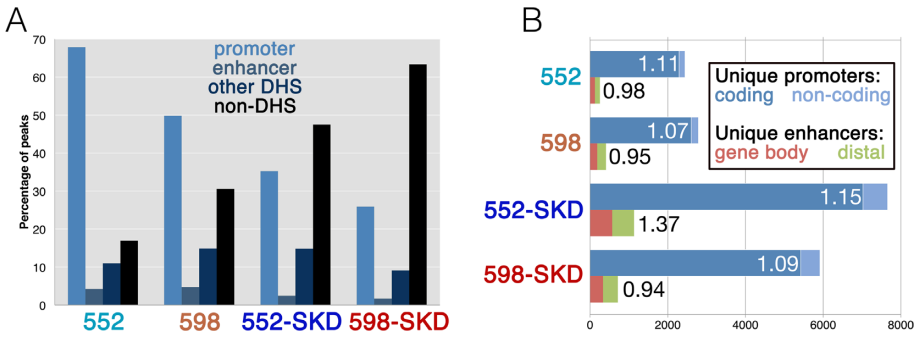

Figure 3. Addition of the SKD domain results in more distal binding sites for both ATFs. (**A**) For each of the four ATFs, the percentage of binding sites located in promoters, enhancers (H3K27ac-enriched regions) and other regions of open chromatin not corresponding to promoters or enhancers (denoted as 'other DHS'), and non-DHS regions. (**B**) Location analysis for all four ATF variants displaying the number of uniquely bound genetic elements; the average number of peaks per unique element is shown.

histone acetylation complexes (thus, the sites are not classified as active enhancers). We expected that most of the ATF binding sites that are not in promoters or active enhancers would fall within the category of non-promoter, non-enhancer DHS (which we denote as 'other DHS' regions). However, we found that very few of the ATF peaks were within this category. In fact, the largest set of non-promoter peaks for each ATF lies outside of DHS (denoted "non-DHS' regions). The surprising finding that the ATFs could bind to non-DHS regions suggested that perhaps these ATFs could bind within regions of methylated DNA (since most of the genome is methylated). Indeed, we found that many of the peaks are within heavily methylated regions, with 552-SKD and 598-SKD non-DHS peaks binding almost exclusively in methylated regions (Figure 4A). It is possible that these methylated regions are not as heavily condensed as regions considered to be heterochromatin (e.g. silenced promoters) and thus the ATFs can access their motifs. We also stress that we have not yet shown that a specific CpG in the genomic sequence directly contacted by the factors is methylated. A browser snapshot example of a promoter peak and an 'other' peak is shown in Figure 4B and C; the promoter peak has low DNA methylation and high H3K4me3 whereas the 'other' peak has high DNA methylation and no H3K4me3 mark.

598-SKD affects activity of only a subset of bound promoters

To determine the functional consequences of binding of an ATF having a domain that recruits a repressive histone-modifying complex to thousands of promoters, we performed ChIP-seq for histone modifications before and after induction of 598-SKD. We found that, in general, the level of H3K4me3 did not significantly change upon binding by 598-SKD when all 5915 bound promoters are considered together (Figure 5). These ChIP-seq analyses suggested that the recruitment of 598-SKD was not causing changes in gene expression at most of the bound promoters. To test this hypothesis, we performed RNA-seq analysis before and after induction of the 598-SKD ATF. As shown in Figure 6A, 769 genes were upregulated and 1066 were downregulated upon induction of 598-SKD. To separate primary, direct effects from secondary, downstream effects on gene regulation, we selected for further analysis only those genes whose promoters were in the set of 5915 promoters bound by 598-SKD (Figure 6B). In this subset of genes, we found that only ∼10% of the bound promoters showed changes in gene expression (416 were downregulated and 264 were upregulated). Representative 598-SKD promoter binding sites and histone modification profiles for a downregulated gene and an upregulated gene are shown in Figure 6C and D, respectively.

As noted above, in general, the set of ∼6000 promoters bound by 598-ATF did not show differences in H3K4me3 patterns. However, it was possible that promoters of genes having expression changes would show differences in histone marks. Therefore, we compared histone modifications at the set of all bound promoters, bound and downregulated promoters, bound and upregulated promoters and all unbound promoters. While H3K4me3 and H3K9ac levels show slight overall decreases genome-wide, we found essentially no differences in histone marks at any subset of bound promoters as compared to unbound promoters (Figure 7). CpG density and peak width at bound promoters also showed no correlation with the ability to downregulate gene expression (Supplementary Figure S6). However, we did find that there was a higher motif density in the set of repressed peaks than in the other two sets (Figure 8A) and that the overall peak height was higher for both 598-SKD and 598 for the set of bound and repressed promoters (Figure 8B). Because 598-SKD could possibly downregulate genes by blocking elongating RNA Polymerase II, we compared the position of the 598-SKD peaks relative to the TSS and expression changes at those genes (Figure 8C). We found that 598-SKD binding was only enriched at the most 5′ end of gene bodies (10–15% of the length of the gene), suggesting that effects on expression are largely dictated by interactions at gene promoters.

DISCUSSION

ATFs and genomic nucleases based on a DNA binding platform consisting of multiple ZNF domains are currently being developed for clinical applications. However, no genome-wide investigations into their binding specificity have been performed. We have created six-finger ATFs that should each specifically recognize a different 18 nt region of the human *SOX2* promoter; each ATF is constructed

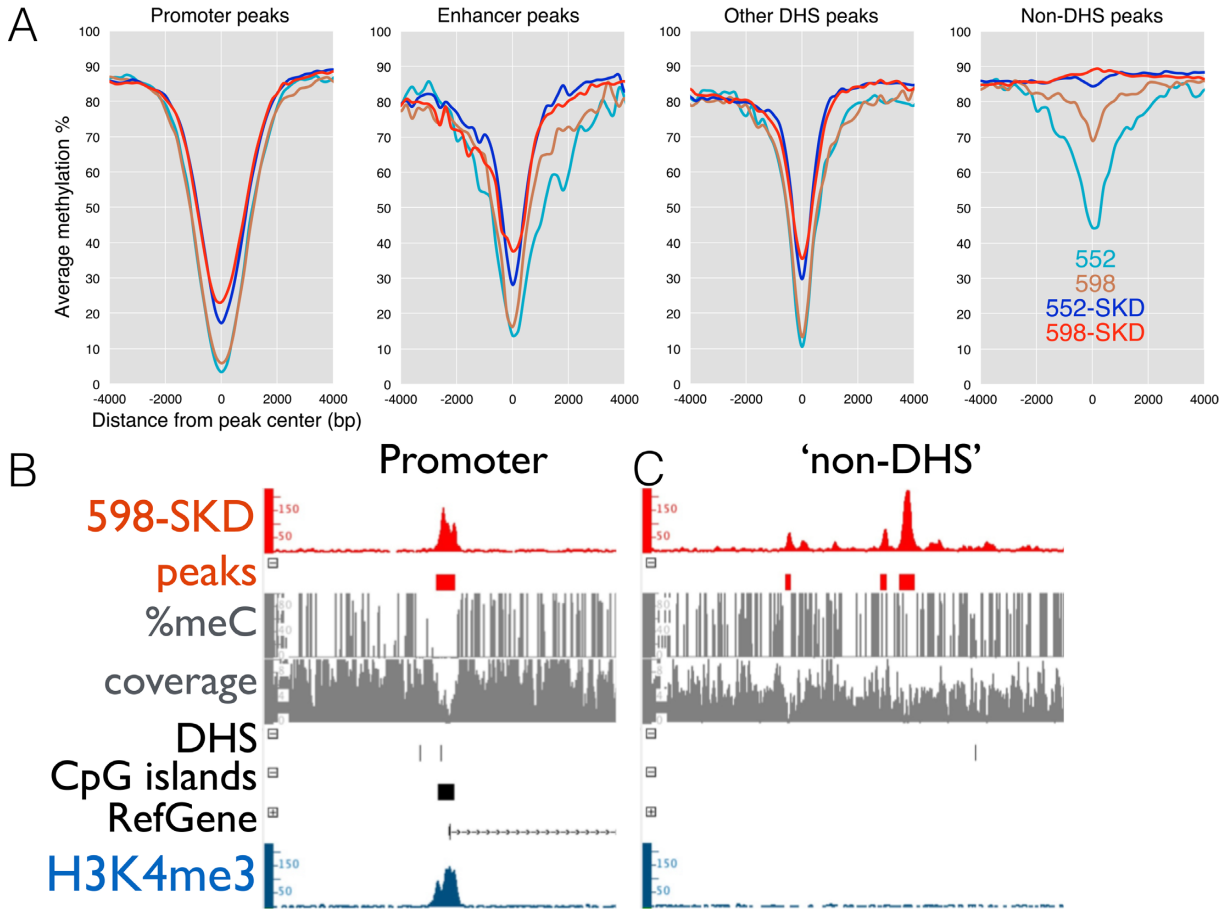

Figure 4. The ATFs can bind to regions of methylated DNA. (**A**) DNA methylation tag density plots for the four categories of ATF binding sites, displaying the average genomic methylation ± 4 kb from peak centers of each ATF variant. Genome browser pile-up tracks for a representative, robust promoter peak (**B**) and (**C**) a 'non-DHS' peak. Histone tracks for each display DNA methylation coverage and levels and H3K4me3 peaks from MCF7 cells.

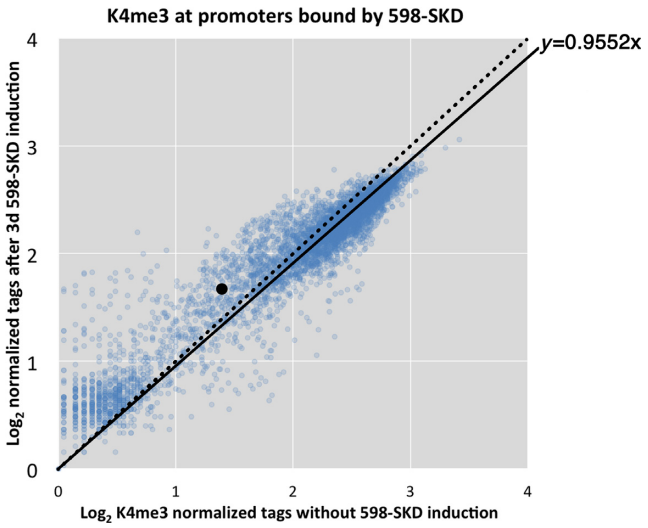

Figure 5. Binding of 598-SKD does not affect H3K4me3 levels at promoters. ChIP-seq scatter plot for normalized H3K4me3 tags at 598-SKD-bound promoters (TSS ± 1 kb) at 0 days and 3 days of 598-SKD induction. The dashed line is $x = y$, the solid line is the best-fit slope of the plotted data (with slope equation), and the *SOX2* promoter is represented as a black circle.

such that it contains or lacks an SKD domain that interacts with a transcription repression complex. ChIP-seq analysis of the effector-free ATFs in MCF7 breast cancer cells identified thousands of binding sites, mostly in promoter regions; the addition of an SKD domain increased the number of binding sites ∼5-fold, with a majority of the new sites located outside of promoters. *De novo* motif analyses suggest that the lack of binding specificity is due to subsets of the finger domains being used for genomic interactions. Although the ATFs display widespread binding, few genes showed expression differences; genes repressed by the ATF-SKD have stronger binding sites and are more enriched for a 12 nt partial target motif. Interestingly, epigenetic analyses indicate that the transcriptional downregulation caused by the ATF-SKD is not due to loss of H3K4me3 on the repressed promoters.

Potential off-target effects due to 598-SKD binding to 25 000 genomic sites are somewhat muted by the observation that the majority of these sites are not at promoter or enhancer regions, suggesting that the majority of the off-target binding would not affect gene expression. However, we did identify ∼6000 promoters that were bound by 598-SKD. At first glance, this may suggest that off-target binding would be a serious problem in that thousands of genes would show differences in expression. To the contrary, we found that

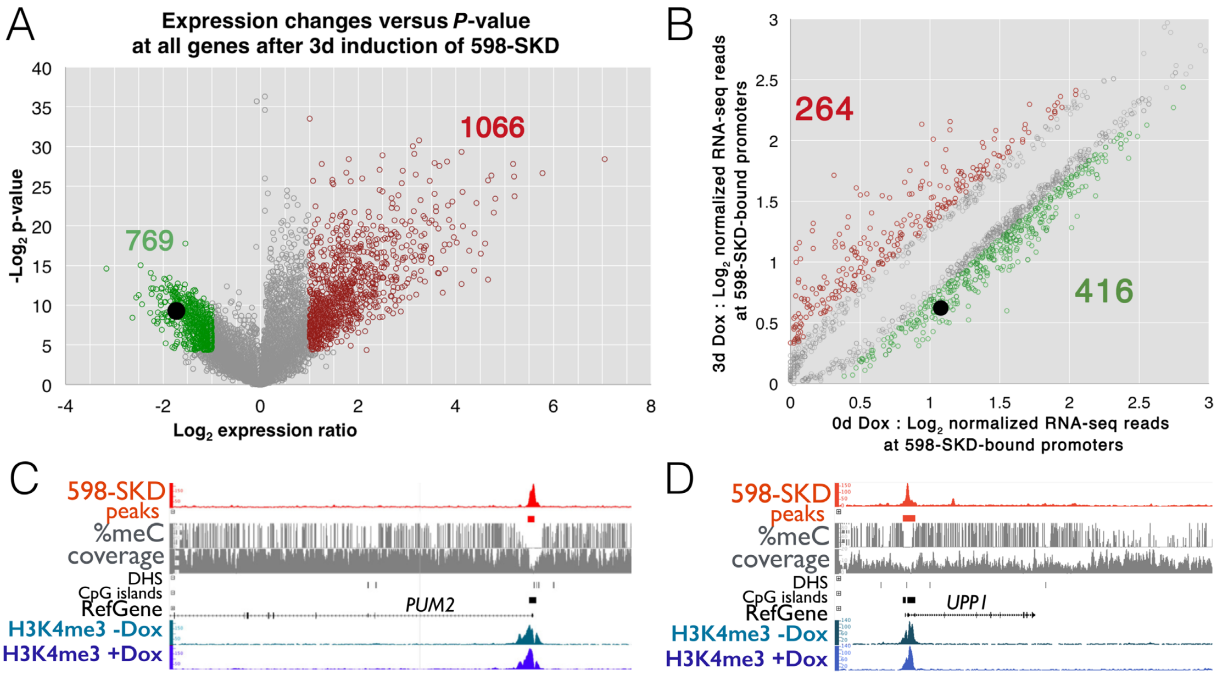

Figure 6. 598-SKD affects activity of only a subset of bound promoters. (**A**) A volcano plot of genome-wide expression changes and upon 3 day induction of 598-SKD. Significantly downregulated genes and upregulated genes (≥2-fold, $P < 0.05$) and counts are shown in green and red, respectively. *SOX2* is represented as a black circle. (**B**) Scatter plot for all significant ($P < 0.05$) expression changes at genes with promoters bound by 598-SKD, comparing 0 days and 3 days of 598-SKD induction. Significantly downregulated and upregulated genes ($P < 0.05$ and expression changed ≥2-fold) with counts are shown in green and red, respectively. *SOX2* is represented as a black circle. Expression changes that were <2-fold, but with $P > 0.05$, are shown in gray. (**C**) Genome browser pile-up tracks for a high-confidence 598-SKD peak at a representative downregulated gene promoter and (**D**) an upregulated gene promoter. Histone tracks for each display H3K4me3 at 0 days and 3 days of 598-SKD induction.

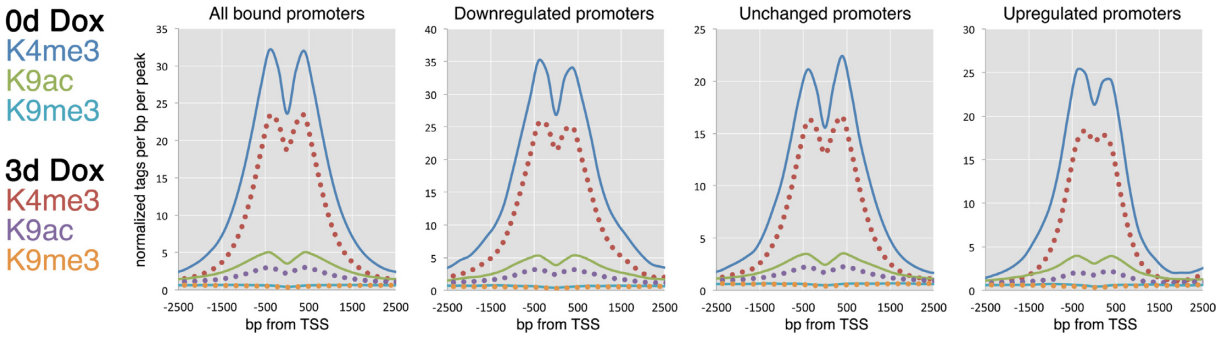

Figure 7. Changes in gene expression are not correlated with changes in histone marks. Tag density plots for K4me3, K9ac and K9me3 modifications of histone H3 at 598-SKD-bound promoters, bound and downregulated promoters, bound and unchanged promoters, bound and upregulated promoters and unbound promoters at 0 and 3 days of 598-SKD induction.

less than 700 genes whose promoters were bound by 598-SKD showed expression changes (Figure 9). It is possible that this subset of genes is just the earliest to respond and that continued 598-SKD induction would eventually drive more widespread downregulation. Nevertheless, our results suggest that off-target binding is perhaps less of an issue than is lack of function upon binding.

The SKD domain is an optimized KAP1-interaction domain from the KRAB-ZNF Kox1 (45). KAP1 is a large protein that serves as a scaffold for components of a complex that is thought to be involved in heterochromatin formation (50). Components of the NuRD histone deacetylase complex and the H3 lysine 9-specific histone methyltransferase SETDB1 bind to KAP1 (51) and are thought to work cooperatively to form condensed heterochromatin. We have previously used ChIP-seq to perform a genome-wide functional analysis of KAP1, identifying thousands of promoter regions and intragenic regions to which KAP1 is recruited (20). However, although the identified binding sites are quite strong and found in multiple cell types, KAP1 did not appear to regulate the expression of most genes that are near its binding sites (20). We note that KAP1 fused to DNA binding domains has been effectively used to regulate artificial promoter constructs containing multiple binding sites for the fusion protein (52–56). Work with other ATFs has also shown increased activity when multiple copies of the ATF are brought to a single promoter (57,58). Taken together, such studies suggest that perhaps an ATF-SKD

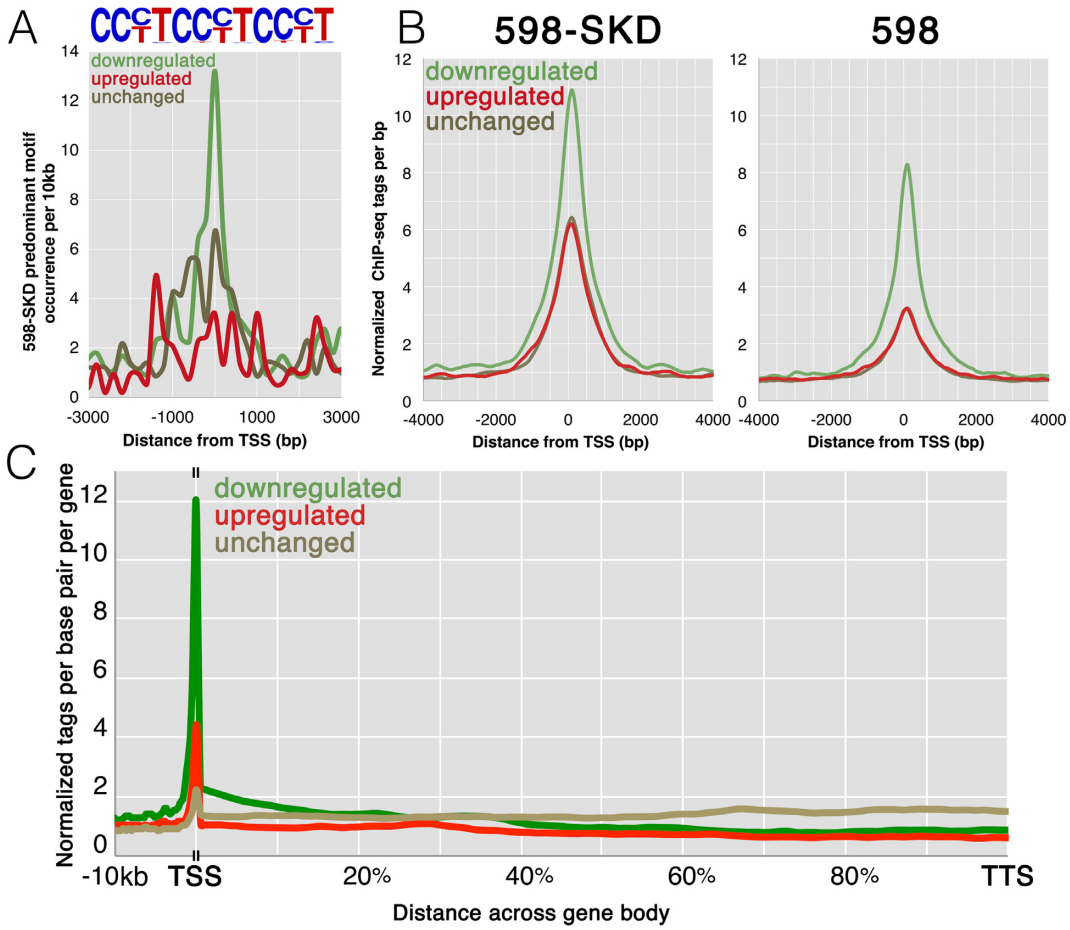

Figure 8. Analysis of activated versus repressed promoters. (**A**) Tag density plot for the frequency of the predominant 598-SKD-bound motif (shown above the plot, representing a partial design target motif) at bound promoters. Colored plot lines correspond to 598-SKD-bound promoters at genes with unchanged, downregulated or upregulated activity. (**B**) Tag density plots for 598-SKD and 598 ChIP-seq tags at 598-SKD target promoters across expression groups; plot colors correspond to those in (A). (**C**) Tag density plots for 598-SKD across promoters and gene bodies; plot colors correspond to those in (A).

acquires the ability to act as a transcriptional repressor only under certain conditions, such as if multiple complexes are recruited via high-affinity DNA binding proteins and are properly positioned near a TSS.

We did find that the promoters that were repressed by 598-SKD had a higher enrichment of motifs. Such studies suggest that one option might be to target a given promoter with multiple ATFs that recognize binding sites spaced appropriately apart to allow simultaneous binding. Although many thousands of off-target binding events for each of the ATFs may occur, perhaps the only promoters that would be highly regulated would be those that have the sites for all of the ATFs. One recent study targeted over 300 different sites across five gene promoters (*OCT4*, *SOX2*, *KLF4*, *c-MYC* and *miR302/367*) using five- or six-fingered artificial ZNFs with p65, VP16 and VP64 activation domains (59). They found that different effector domains were better suited to upregulate each gene and that not all sites resulted in significant target gene expression changes. Genome-wide binding analyses were not performed and therefore off-target binding was not investigated.

We note that some genes driven by promoters that were bound by 598-SKD showed increased RNA levels. Although it is certainly possible that many of these changes are indirect responses to upstream effects on signaling pathways, KAP1 has been previously associated with transcriptional activation. For example, using reporter-promoter assays, KAP1 was shown to function as a co-activator for NGFI-B and CCAAT-enhancer-binding protein (60–62). Also, the effective re-activation of the endogenous Oct4 promoter by a six ZNF ATF required a KRAB domain (63). The degree of activation varied in different cell types, in accordance with previous studies that have suggested that KAP1 function is cell type-specific (50). Thus, although KRAB domain-containing ATFs are traditionally described as transcriptional repressors, they might also function as activators; the mechanisms by which KAP1 can activate transcription are not known.

In conclusion, although we were initially dismayed by the identification of thousands of off-target binding sites, such a large number of binding events have provided us with the opportunity to gain insight into several features of ATF-mediated gene regulation. For example, it was ex-

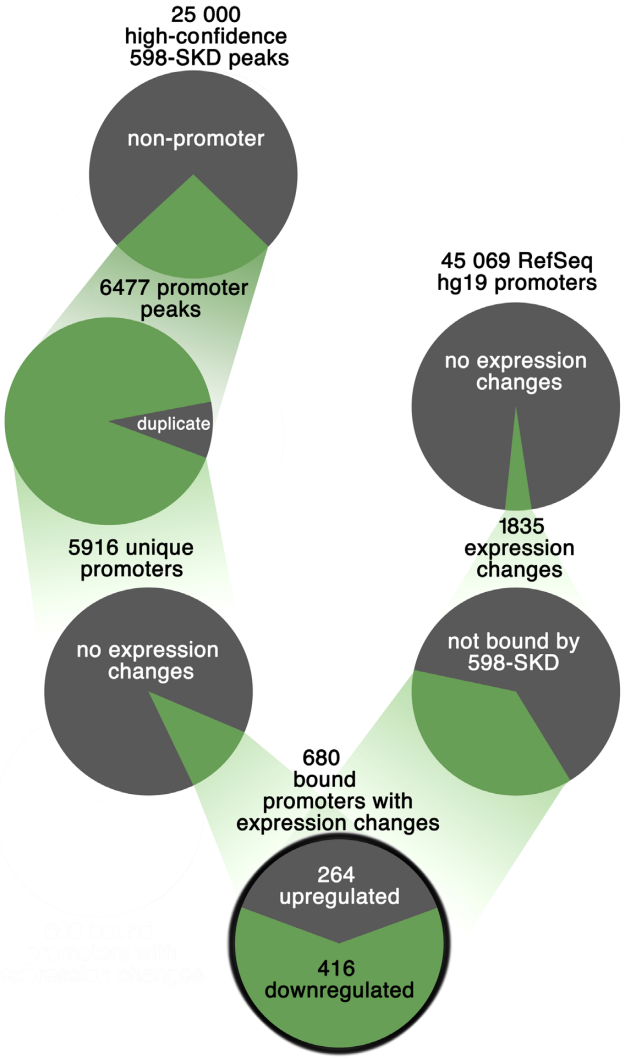

Figure 9. Widespread binding but limited regulation. Overall starting numbers for 598-SKD peaks (25 000) and RefSeq hg19 promoters (45 069) converge on 416 genes as each set is filtered for 598-SKD binding, the number of unique promoters bound, and gene downregulation.

citing to observe binding of the ATFs to regions of highly methylated DNA, suggesting that they may be useful as pioneering factors to target (and subsequently activate) a silenced promoter. Few characterized transcription factors have this ability (in fact, ZNF274 is the only factor studied by ENCODE that shows binding outside of a DHS) and thus ATFs based on ZNFs may provide a unique opportunity to remodel heterochromatic genomic regions. Also, our studies suggest that the outside fingers do not always contribute to binding specificity; perhaps constructing eight-finger ATFs would allow the internal six fingers to contribute to binding specificity and would result in fewer off-target sites. Finally, the fact that only a small percentage of promoters bound by 598-SKD showed gene expression changes, that bound promoters could be either repressed or activated, and the fact that previous studies have suggested that KAP1-mediated regulation is cell type-specific, we suggest that before use in clinical settings, preliminary studies in cell culture should be performed to test whether an ATF-SKD, or more likely a combination of ATF-SKDs, will activate or repress your gene of interest. We also are aware that other genome targeting platforms, such as clustered regularly interspaced short palindromic repeats (CRISPRs), may provide more specificity (31,64–66). In fact, after learning of our studies, others have directly compared the binding specificity of a CRISPR and a ZNF targeted to the same genomic region; the ZNFs had many more off-target binding sites (O'Geen and Segal, personal communication).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGMENTS

The ATF ChIP-seq samples were sequenced at the University of North Carolina at Chapel Hill Genome Analysis Facility whereas the histone ChIP-seq and WGBS samples were sequenced at the USC NCCC Molecular Genomics Sequencing Core. The MCF7 DHS data was produced by the lab of John Stamatoyannopoulos (UCSC accession # wgEncodeEH000502) and H3K27ac MCF7 ChIP-seq was produced by the laboratory of Peggy Farnham (UCSC accession # wgEncodeEH002872), both as a part of the ENCODE Consortium (39–42), and are available at http://genome.ucsc.edu. Chris Benner provided the Homer suite of tools and added a tool adaptation for methylation analysis upon request (Salk Institute, La Jolla, CA).

FUNDING

National Cancer Institute [P30CA014089]; Australian Research Council Future Fellowship [FT130101767, FT120100862]; Cancer Council of Western Australia Research Fellowship; National Health and Medical Research Council [APP1069830]; National Institutes of Health [R01CA170370, R01DA036906, R21HG006761]; National Breast Cancer Foundation Novel Concept award NC-14-024; Raine Medical Research Foundation; Western Australia Centre of Excellence for Computational Systems Biology. Funding for open access charge: National Institutes of Health [R21HG006761].
Conflict of interest statement. None declared.

REFERENCES

1. Lee, T.I. and Young, R.A. Lee, T.I. and Young, R.A. (2013) Transcriptional regulation and its misregulation in disease. *Cell*, **152**, 1237–1251.
2. Farnham, P.J.Farnham, P.J. (2009) Insights from genomic profiling of transcription factors. *Nature Rev. Genet.*, **10**, 605–616.
3. Yang, X., Lay, F., Han, H., and Jones, P.A.Yang, X., Lay, F., Han, H., and Jones, P.A. (2010) Targeting DNA methylation for epigenetic therapy. *Trends Pharmacol. Sci.*, **31**, 536–546.
4. Li, X., Zhang, J., Xie, Y., Jiang, Y., Yingjie, Z., and Xu, W.Li, X., Zhang, J., Xie, Y., Jiang, Y., Yingjie, Z., and Xu, W. (2014) *Curr. Drug Targets*, **15**, 622–634.
5. West, A.C. and Johnstone, R.W.West, A.C. and Johnstone, R.W. (2014) New and emerging HDAC inhibitors for cancer treatment. *J. Clin. Invest.*, **124**, 30–39.
6. Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A.Lovén, J., Hoke,

H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013) Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*, **153**, 320–334.
7. Dobbelstein, M. and Moll, U.Dobbelstein, M. and Moll, U. (2014) Targeting tumour-supportive cellular machineries in anticancer drug development. Nature reviews. *Drug Discov.*, **13**, 179–196.
8. Segal, D.J. and Meckler, J.F.Segal, D.J. and Meckler, J.F. (2013) Genome engineering at the dawn of the golden age. *Annu. Rev. Genomics Hum. Genet.*, **14**, 135–158.
9. Grimmer, M.R. and Farnham, P.J.in pressGrimmer, M.R. and Farnham, P.J. (2014) Can genome engineering be used to target cancer-associated enhancers?. *Epigenomics*,.
10. Verschure, P.J., Visser, A.E., and Rots, M.G. Verschure, P.J., Visser, A.E., and Rots, M.G. (2006) Step out of the groove: epigenetic gene control systems and engineered transcription factors. *Adv. Genet.*, **56**, 163–204.
11. Liu, Q., Segal, D.J., Ghiara, J.B., and Barbas, C.F. IIILiu, Q., Segal, D.J., Ghiara, J.B., and Barbas, C.F. III (1997) Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 5525–5530.
12. Tadepally, H., Burger, G., and Aubry, M.Tadepally, H., Burger, G., and Aubry, M. (2008) Evolution of C2H2-zinc finger genes and subfamilies in mammals: species-specific duplication and loss of clusters, genes and effector domains. *BMC Evol. Biol.*, **8**, 176.
13. Segal, D.J., Beerli, R.R., Blancafort, P., Dreier, B., Effertz, K., Huber, A., Koksch, B., Lund, C.V., Magnenat, L., and Valente, D. *et al.*Segal, D.J., Beerli, R.R., Blancafort, P., Dreier, B., Effertz, K., Huber, A., Koksch, B., Lund, C.V., Magnenat, L., and Valente, D. (2003) Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry*, **42**, 2137–2148.
14. Beerli, R.R., Segal, D.J., Dreier, B., and Barbas, C.F. IIIBeerli, R.R., Segal, D.J., Dreier, B., and Barbas, C.F. III (1998) Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 14628–14633.
15. Bhakta, M.S. and Segal, D.J.Bhakta, M.S. and Segal, D.J. (2010) The generation of zinc finger proteins by modular assembly. *Methods Mol. Biol.*, **649**, 3–30.
16. Segal, D.J. and Barbas, C.F.Segal, D.J. and Barbas, C.F. (1999) Design of novel sequence-specific DNA-binding proteins. *Curr. Opin. Chem. Biol.*, **4**, 34–39.
17. Dreier, B., Beerli, R.R., Segal, D.J., Flippin, J.D., and Barbas, C.F. IIIDreier, B., Beerli, R.R., Segal, D.J., Flippin, J.D., and Barbas, C.F. III (2001) Development of zinc finger domains for recognition of the 5′-ANN-3′ family of DNA sequences and their use in the construction of artificial transcription factors. *J. Biol. Chem.*, **276**, 29466–29478.
18. Elrod-Erickson, M., Rould, M.A., Nekludova, L., and Pabo, C.O.Elrod-Erickson, M., Rould, M.A., Nekludova, L., and Pabo, C.O. (1996) Zif268 protein-DNA complex refined at 1.6 A: a model system for understanding zinc finger-DNA interactions. *Structure*, **4**, 1171–1180.
19. Frietze, S., O'Geen, H., Blahnik, K.R., Jin, V.X., and Farnham, P.J.Frietze, S., O'Geen, H., Blahnik, K.R., Jin, V.X., and Farnham, P.J. (2010) ZNF274 recruits the histone methyltransferase SETDB1 to the 3′ ends of ZNF genes. *PloS One*, **5**, e15082.
20. Iyengar, S., Ivanov, A.V., Jin, V.X., Rauscher, F.J. III, and Farnham, P.J.Iyengar, S., Ivanov, A.V., Jin, V.X., Rauscher, F.J. III, and Farnham, P.J. (2011) Functional analysis of KAP1 genomic recruitment. *Mol. Cell. Biol.*, **31**, 1833–1847.
21. Gommans, W.M., McLaughlin, P.M., Lindhout, B.I., Segal, D.J., Wiegman, D.J., Haisma, H.J., van der Zaal, B.J., and Rots, M.G.Gommans, W.M., McLaughlin, P.M., Lindhout, B.I., Segal, D.J., Wiegman, D.J., Haisma, H.J., van der Zaal, B.J., and Rots, M.G. (2007) Engineering zinc finger protein transcription factors to downregulate the epithelial glycoprotein-2 promoter as a novel anti-cancer treatment. *Mol. Carcinog.*, **46**, 391–401.
22. Beltran, A.S. and Blancafort, P.Beltran, A.S. and Blancafort, P. (2011) Reactivation of MASPIN in non-small cell lung carcinoma (NSCLC) cells by artificial transcription factors (ATFs). *Epigenetics*, **6**, 224–235.
23. Beltran, A., Parikh, S., Liu, Y., Cuevas, B.D., Johnson, G.L., Futscher, B.W., and Blancafort, P.Beltran, A., Parikh, S., Liu, Y., Cuevas, B.D., Johnson, G.L., Futscher, B.W., and Blancafort, P. (2007) Re-activation of a dormant tumor suppressor gene maspin by designed transcription factors. *Oncogene*, **26**, 2791–2798.
24. Lara, H., Wang, Y., Beltran, A.S., Juarez-Moreno, K., Yuan, X., Kato, S., Leisewitz, A.V., Cuello Fredes, M., Licea, A.F., and Connolly, D.C. *et al.*Lara, H., Wang, Y., Beltran, A.S., Juarez-Moreno, K., Yuan, X., Kato, S., Leisewitz, A.V., Cuello Fredes, M., Licea, A.F., and Connolly, D.C. (2012) Targeting serous epithelial ovarian cancer with designer zinc finger transcription factors. *J. Biol. Chem.*, **287**, 29873–29886.
25. Chen, H., Kazemier, H.G., de Groote, M.L., Ruiters, M.H., Xu, G.L., and Rots, M.G.Chen, H., Kazemier, H.G., de Groote, M.L., Ruiters, M.H., Xu, G.L., and Rots, M.G. (2013) Induced DNA demethylation by targeting ten-eleven translocation 2 to the human ICAM-1 promoter. *Nucleic Acids Res.*, **42**, 1563–1574.
26. Rivenbark, A.G., Stolzenburg, S., Beltran, A.S., Yuan, X., Rots, M.G., Strahl, B.D., and Blancafort, P.Rivenbark, A.G., Stolzenburg, S., Beltran, A.S., Yuan, X., Rots, M.G., Strahl, B.D., and Blancafort, P. (2012) Epigenetic reprogramming of cancer cells via targeted DNA methylation. *Epigenetics*, **7**, 350–360.
27. Stolzenburg, S., Rots, M.G., Beltran, A.S., Rivenbark, A.G., Yuan, X., Qian, H., Strahl, B.D., and Blancafort, P.Stolzenburg, S., Rots, M.G., Beltran, A.S., Rivenbark, A.G., Yuan, X., Qian, H., Strahl, B.D., and Blancafort, P. (2012) Targeted silencing of the oncogenic transcription factor SOX2 in breast cancer. *Nucleic Acids Res.*, **40**, 6725–6740.
28. Beltran, A.S., Sun, X., Lizardi, P.M., and Blancafort, P.Beltran, A.S., Sun, X., Lizardi, P.M., and Blancafort, P. (2008) Reprogramming epigenetic silencing: artificial transcription factors synergize with chromatin remodeling drugs to reactivate the tumor suppressor mammary serine protease inhibitor. *Mol. Cancer Ther.*, **7**, 1080–1090.
29. Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., and Doudna, J.A. *et al.*Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., and Doudna, J.A. (2013) *Cell*, **154**, 442–451.
30. Mendenhall, E.M., Williamson, K.E., Reyon, D., Zou, J.Y., Ram, O., Joung, J.K., and Bernstein, B.E.Mendenhall, E.M., Williamson, K.E., Reyon, D., Zou, J.Y., Ram, O., Joung, J.K., and Bernstein, B.E. (2013) Locus-specific editing of histone modifications at endogenous enhancers. *Nat. Biotechnol.*, **31**, 1133–1136.
31. Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., Cheng, A.W., Trevino, A.E., Konermann, S., and Chen, S. Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., Cheng, A.W., Trevino, A.E., Konermann, S., and Chen, S. (2014) Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. In: *Nature Biotech.*, pp. 670–676.
32. O'Geen, H., Echipare, L., and Farnham, P.J.O'Geen, H., Echipare, L., and Farnham, P.J. (2011) Using ChIP-seq technology to generate high-resolution profiles of histone modifications. *Methods Mol. Biol.*, **791**, 265–286.
33. Langmead, B. and Salzberg, S.L.Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
34. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
35. Quinlan, A.R. and Hall, I.M.Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
36. Blahnik, K.R., Dou, L., O'Geen, H., McPhillips, T., Xu, X., Cao, A.R., Iyengar, S., Nicolet, C.M., Ludaescher, B., and Korf, I. *et al.*Blahnik, K.R., Dou, L., O'Geen, H., McPhillips, T., Xu, X., Cao, A.R., Iyengar, S., Nicolet, C.M., Ludaescher, B., and Korf, I. (2010) Sole-search: an integrated analysis program for peak detection and functional annotation using ChIP-seq data. *Nucleic Acids Research*, **38**, e13.
37. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K.Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre,

C., Singh, H., and Glass, C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
38. Trapnell, C., Pachter, L., and Salzberg, S.L.Trapnell, C., Pachter, L., and Salzberg, S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**, 1105–1111.
39. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
40. (2011) A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.*, **9**, e1001046.
41. Consortium, T.E.Consortium, T.E. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799–816.
42. (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*, **306**, 636–640.
43. Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D.Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D. (2010) Genome editing with engineered zinc finger nucleases. Nature reviews. *Genetics*, **11**, 636–646.
44. Ramirez, C.L., Foley, J.E., Wright, D.A., Muller-Lerch, F., Rahman, S.H., Cornu, T.I., Winfrey, R.J., Sander, J.D., Fu, F., and Townsend, J.A. *et al.*Ramirez, C.L., Foley, J.E., Wright, D.A., Muller-Lerch, F., Rahman, S.H., Cornu, T.I., Winfrey, R.J., Sander, J.D., Fu, F., and Townsend, J.A. (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat. Methods*, **5**, 374–375.
45. Margolin, J.F., Friedman, J.R., Meyer, W.K., Vissing, H., Thiesen, H.J., and Rauscher, F.J. IIIMargolin, J.F., Friedman, J.R., Meyer, W.K., Vissing, H., Thiesen, H.J., and Rauscher, F.J. III (1994) Kruppel-associated boxes are potent transcriptional repression domains. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 4509–4513.
46. Blahnik, K.R., Dou, L., Echipare, L., Iyengar, S., O'Geen, H., Sanchez, E., Zhao, Y., Marra, M.A., Hirst, M., and Costello, J.F. *et al.*Blahnik, K.R., Dou, L., Echipare, L., Iyengar, S., O'Geen, H., Sanchez, E., Zhao, Y., Marra, M.A., Hirst, M., and Costello, J.F. (2011) Characterization of the contradictory chromatin signatures at the 3′ exons of zinc finger genes. *PloS One*, **6**, e17121.
47. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
48. Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., and Vernot, B. *et al.*Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., and Vernot, B. (2012) The accessible chromatin landscape of the human genome. *Nature*, **489**, 75–82.
49. Frietze, S., Wang, R., Yao, L., Tak, Y.G., Ye, Z., Gaddis, M., Witt, H., Farnham, P.J., and Jin, V.X.Frietze, S., Wang, R., Yao, L., Tak, Y.G., Ye, Z., Gaddis, M., Witt, H., Farnham, P.J., and Jin, V.X. (2012) Cell type-specific binding patterns reveal that TCF7L2 can be tethered to the genome by association with GATA3. *Genome Biol.*, **13**, R52.
50. Iyengar, S. and Farnham, P.J.Iyengar, S. and Farnham, P.J. (2011) KAP1: an enigmatic master regulator of the genome. *J. Biol. Chem.*, **286**, 26267–26276.
51. Ivanov, A.V., Peng, H., Yurchenko, V., Yap, K.L., Negorev, D.G., Schultz, D.C., Psulkowski, E., Fredericks, W.J., White, D.E., and Maul, G.G. *et al.*Ivanov, A.V., Peng, H., Yurchenko, V., Yap, K.L., Negorev, D.G., Schultz, D.C., Psulkowski, E., Fredericks, W.J., White, D.E., and Maul, G.G. (2007) PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing. *Mol. Cell*, **28**, 823–837.
52. Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., and Rauscher, F.J. IIISchultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., and Rauscher, F.J. III (2002) SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev.*, **16**, 919–932.
53. Sripathy, S.P., Stevens, J., and Schultz, D.C.Sripathy, S.P., Stevens, J., and Schultz, D.C. (2006) The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression. *Mol. Cell. Biol.*, **26**, 8623–8638.
54. Groner, A.C., Meylan, S., Ciuffi, A., Zangger, N., Ambrosini, G., Denervaud, N., Bucher, P., and Trono, D.Groner, A.C., Meylan, S., Ciuffi, A., Zangger, N., Ambrosini, G., Denervaud, N., Bucher, P., and Trono, D. (2010) KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. *PLoS Genet.*, **6**, e1000869.
55. Ayyanathan, K., Lechner, M.S., Bell, P., Maul, G.G., Schultz, D.C., Yamada, Y., Tanaka, K., Torigoe, K., and Rauscher, F.J. IIIAyyanathan, K., Lechner, M.S., Bell, P., Maul, G.G., Schultz, D.C., Yamada, Y., Tanaka, K., Torigoe, K., and Rauscher, F.J. III (2003) Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. *Genes Dev.*, **17**, 1855–1869.
56. Barde, I., Laurenti, E., Verp, S., Groner, A.C., Towne, C., Padrun, V., Aebischer, P., Trumpp, A., and Trono, D.Barde, I., Laurenti, E., Verp, S., Groner, A.C., Towne, C., Padrun, V., Aebischer, P., Trumpp, A., and Trono, D. (2009) Regulation of episomal gene expression by KRAB/KAP1-mediated histone modifications. *J. Virol.*, **83**, 5574–5580.
57. Perez-Pinera, P., Ousterout, D.G., Brunger, J.M., Farin, A.M., Glass, K.A., Guilak, F., Crawford, G.E., Hartemink, A.J., and Gersbach, C.A.Perez-Pinera, P., Ousterout, D.G., Brunger, J.M., Farin, A.M., Glass, K.A., Guilak, F., Crawford, G.E., Hartemink, A.J., and Gersbach, C.A. (2013) Synergistic and tunable human gene activation by combinations of synthetic transcription factors. *Nat. Methods*, **10**, 239–242.
58. Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., and Leong, K.W. *et al.*Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., and Leong, K.W. (2013) RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods*, **10**, 973–976.
59. Ji, Q., Fischer, A.L., Brown, C.R., Eastlund, E.R., Dvash, T., Zhong, B., Gerber, M.A., Lyons, I., Knight, S.W., and Kreader, C.A.Ji, Q., Fischer, A.L., Brown, C.R., Eastlund, E.R., Dvash, T., Zhong, B., Gerber, M.A., Lyons, I., Knight, S.W., and Kreader, C.A. (2014) Engineered zinc-finger transcription factors activate OCT4 (POU5F1), SOX2, KLF4, c-MYC (MYC) and miR302/367. *Nucleic Acids Res*, **42**, 6158–6167.
60. Chang, C.J., Chen, Y.L., and Lee, S.C.Chang, C.J., Chen, Y.L., and Lee, S.C. (1998) Coactivator TIF1beta interacts with transcription factor C/EBPbeta and glucocorticoid receptor to induce alpha1-acid glycoprotein gene expression. *Mol. Cell. Biol.*, **18**, 5880–5887.
61. Rooney, J.W. and Calame, K.L.Rooney, J.W. and Calame, K.L. (2001) TIF1beta functions as a coactivator for C/EBPbeta and is required for induced differentiation in the myelomonocytic cell line U937. *Genes Dev.*, **15**, 3023–3038.
62. Rambaud, J., Desroches, J., Balsalobre, A., and Drouin, J.Rambaud, J., Desroches, J., Balsalobre, A., and Drouin, J. (2009) TIF1beta/KAP-1 is a coactivator of the orphan nuclear receptor NGFI-B/Nur77. *J. Biol. Chem.*, **284**, 14147–14156.
63. Juarez-Moreno, K., Erices, R., Beltran, A.S., Stolzenburg, S., Cuello-Fredes, M., Owen, G.I., Qian, H., and Blancafort, P.Juarez-Moreno, K., Erices, R., Beltran, A.S., Stolzenburg, S., Cuello-Fredes, M., Owen, G.I., Qian, H., and Blancafort, P. (2013) Breaking through an epigenetic wall: Re-activation of Oct4 by KRAB-containing designer zinc finger transcription factors. *Epigenetics*, **8**, 164–176.
64. Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., and Shalem, O. *et al.*Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., and Shalem, O. (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.*, **31**, 827–832.
65. Kuscu, C., Arslan, S., Singh, R., Thorpe, J., and Adli, M.Kuscu, C., Arslan, S., Singh, R., Thorpe, J., and Adli, M. (2014) Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.*, **32**, 677–683.
66. Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J.S.Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J.S. (2014) Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.*, **24**, 132–141.