Recruitment of the Mammalian Histone-modifying EMSY Complex to Target Genes Is Regulated by ZNF131*[S]

Radhika A. Varier[‡1], Enrique Carrillo de Santa Pau[§2,3], Petra van der Groep[¶3], Rik G. H. Lindeboom[§3], Filomena Matarese[§], Anneloes Mensinga[‡4], Arne H. Smits[‡§], Raghu Ram Edupuganti[§], Marijke P. Baltissen[§], Pascal W. T. C. Jansen[‡§], Natalie ter Hoeve[¶], Danny R. van Weely[‡], Ina Poser[‖], Paul J. van Diest[¶], Hendrik G. Stunnenberg[§], and Michiel Vermeulen[‡§5]

From the ‡Department of Molecular Cancer Research, University Medical Center Utrecht, Universiteitsweg 100, 3584CG Utrecht, The Netherlands, §Department of Molecular Biology, Radboud Institute for Molecular Life Sciences, Geert Grooteplein Zuid 30, 6525GA Nijmegen, The Netherlands, ¶Department of Pathology, University Medical Center Utrecht Heidelberglaan 100, 3584CX Utrecht, The Netherlands, and ‖Max Planck Institute of Molecular Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden, Germany

Recent work from others and us revealed interactions between the Sin3/HDAC complex, the H3K4me3 demethylase KDM5A, GATAD1, and EMSY. Here, we characterize the EMSY/KDM5A/SIN3B complex in detail by quantitative interaction proteomics and ChIP-sequencing. We identify a novel substoichiometric interactor of the complex, transcription factor ZNF131, which recruits EMSY to a large number of active, H3K4me3 marked promoters. Interestingly, using an EMSY knock-out line and subsequent rescue experiments, we show that EMSY is in most cases positively correlated with transcriptional activity of its target genes and stimulates cell proliferation. Finally, by immunohistochemical staining of primary breast tissue microarrays we find that EMSY/KDM5A/SIN3B complex subunits are frequently overexpressed in primary breast cancer cases in a correlative manner. Taken together, these data open venues for exploring the possibility that sporadic breast cancer patients with EMSY amplification might benefit from epigenetic combination therapy targeting both the KDM5A demethylase and histone deacetylases.

In the eukaryotic nucleus, DNA is condensed in the form of chromatin. Chromatin is repressive to processes that require access to the DNA. Chromatin accessibility can be regulated in multiple ways, viz. by ATP-dependent chromatin remodeling and post-translational modifications of core histones. Furthermore, the DNA itself can also be modified. A large number of post-translational modifications on core histones have been identified and characterized, including lysine acetylation and lysine methylation. Some of these modifications are associated with active transcription, whereas others are repressive to transcription. The most studied post-translational modifications on core histones are trimethylations of lysine residues, of which several are known to occur on the N terminus of one of the four core histones, H3. Two of these (H3K4me3 and H3K36me3) are linked to active promoters and gene bodies, respectively, and a third one (H3K27me3) is associated with gene silencing (1). The biology of each of these modifications is believed to be at least partially determined by proteins, so-called chromatin readers, that can specifically recognize these chromatin marks. Lysine trimethylation reader domains include the PHD finger, the chromodomain, and the Tudor domain (2).

Previously, we used quantitative mass spectrometry-based proteomics to identify a large number of chromatin readers for lysine trimethylation sites on histone H3 and H4 (3). In that study, we identified a number of previously unrelated proteins, including GATAD1, EMSY, KDM5A, and SIN3B, as readers for H3K4me3. Subsequent experiments revealed that these proteins may in fact be part of a novel protein complex. Similar observations in *Drosophila* were also reported by others (4). EMSY was originally reported as a transcriptional repressor and breast cancer-associated protein that interacts with the BRCA2 protein (5). KDM5A is a histone lysine demethylase for H3K4me3 (6) containing a PHD finger that directly binds H3K4me3 (7). SIN3B is part of a complex containing histone deacetylase activity (8). These proteins and their associated enzymatic activities therefore suggest that the EMSY complex is a transcriptional repressor.

Here, we applied state-of-the-art quantitative mass spectrometry and ChIP sequencing approaches to characterize the EMSY protein complex and its genome-wide binding profile. Our experiments revealed a striking association between EMSY and H3K4me3-marked, active promoters. Using a CRISPR/Cas9-based knock-out line and rescue experiments we show that in most cases, EMSY expression is positively correlated with expression of its target genes and stimulates cell prolifer-

ation. Furthermore, we identify a novel interactor of the EMSY complex, ZNF131, which recruits EMSY to a subset of its genome-wide targets. Finally, we used immunohistochemistry on primary breast cancer tissues, which revealed a striking expression correlation between EMSY complex subunits, suggesting a functional and pathological link between these proteins in relation to breast cancer.

Experimental Procedures

Generating Stable Cell Lines and Cell Culture—To generate mammalian cells expressing proteins of the EMSY/KDM5A complex at near endogenous levels bacterial artificial chromosomes containing EMSY, GATAD1, PHF12, or KDM5A, respectively were obtained from the BACPAC Resources Center and a GFP cassette was inserted as a C-terminal in-frame fusion using BAC recombineering (9). HeLa or MCF7 cells were transfected with the recombineered BAC using Lipofectamine 2000 or PEImax (for EMSY) and selected with 400 μg/ml geneticin (G418).

Expression of the recombineered ZNF131 BAC in HeLa cells proved unsuccessful hence a full-length mouse ZNF131 construct that was a gift from Juliet Daniel was cloned into the plasmid pBabePuro-C-GFP by gateway cloning that was used to generate retrovirus particles using the Phoenix ecotropic packaging cells. Finally MCF7 cells were transduced with this virus and 24 h after infection, subjected to 1 μg/ml puromycin selection. Monoclonals were isolated, and cell lines were propagated.

Short hairpin knockdown of ZNF131 in the EMSY GFP expressing HeLa cells was mediated by lentiviral transduction. A set of 5 short hairpins purchased from Sigma were transfected into COS7 cells in combination with lentiviral packaging vectors using PEImax. After 24 h, virus generated into the medium was aspirated, filtered, and concentrated. The target cells (EMSY GFP HeLa) were transduced with the virus in the presence of polybrene. shRNA integrated cells were subjected to puromycin selection, 24 h after transduction. Single cells were grown into monoclonals and the knockdown status was assessed by RT-qPCR and Western blotting (Anti-ZNF131, mouse polyclonal #H0000 7690-B0IP, Abnova). The monoclonals selected for this study were shZNF131#1 (TRCN0000254328) and shZNF131#2 (TRCN0000254331).

CRISPR Knock-out of EMSY—The knock-out of EMSY was created using CRISPR-Cas9 system. Ensembl transcript sequence of isoform C11orf30-001 (ENST00000529032) was used as canonical isoform. The guides were designed against an exon common to all isoforms, *i.e.* exon3. The oligos (Biolegio) encoding guide RNAs for inducing a double strand break (DSB)[6] in the genome were designed using the Zhang laboratory genome engineering website CRISPR tool. Guide sequence EMSY-DSB-G2-F: AACGCCACCGTGCTGAAGTT and EMSY-DSB-G2-R: AACTTCAGCACGGTGGCGTT were used for cloning. The guides were cloned into pSpCas9(BB)-2A-Puro (PX459) vector (10). The resulting constructs were sequence verified and used for transfection. The constructs

were transfected into HeLa cells using Lipofectamine LTX&PLUS (Invitrogen). The transfected cells were grown for 2 days and subsequently selected with puromycin for a brief period of 2 days. The surviving cells were clonal diluted into 96 well plates to generate clonal populations from single cells. Genomic DNA from the clonal populations was harvested by using a genomic DNA isolation kit (Invitrogen). The DSB target region was amplified using primers EMSY-forward: GCCTGC-TTGGCAGAGTTCTAT and EMSY-reverse: GACTCAGGA-ATCTGCTACACAA. The amplified regions were Sanger sequenced, and resulting sequences were matched to the wild type transcript. Clones with deletions were tested for knockouts using Western blots (Anti-EMSY, rabbit polyclonal, #A300-253A, Bethyl Laboratories). Re-expression of EMSY-GFP in the EMSY CRISPR knock-out lines was performed using an EMSY-GFP (EX-E2426-M29) from Genecoepia (EX-NEG-M29 was the empty control vector).

GFP Purifications and Label-free Quantitative Mass Spectrometry—Nuclear extract isolation and GFP pull-downs followed by sample preparation for mass spectrometry and data analysis was performed as described in Ref. 11. Briefly, nuclear extracts obtained from GFP-tagged protein-expressing and wild-type cells were subjected to GFP-affinity pull-downs. 1 mg of nuclear extract was used for each pull-down and incubated with GFP chromotek beads for 90 min, and then the beads were washed and subjected to on-bead trypsin digestion. The tryptic peptides were eluted and applied to online nano LC-MS/MS. Raw data were analyzed by Max Quant and *t* test-based statistics was applied on LFQ for interactor identification. iBAQ algorithm was used for determining the stoichiometry of the identified complexes.

Cell Proliferation Assays—Four different cell lines were used for the cell proliferation assays shown in Fig. 4. WT Hela cells with empty vector control and EMSY knock-out cells (CRISPR-mediated knock-out, see Fig. 4*A*) were used in Fig. 4, *B* and *C*. Furthermore, EMSY knock-out cells were stably transfected (PEI transfection reagent, G418 selection) with an EMSY-GFP construct or an empty vector control (Fig. 4, *D* and *E*). The four resulting cell lines were trypsinized into single cells, counted, and 250 or 500 cells were seeded in duplicates in 24-well plates containing DMEM with 10% FBS. The cells were allowed to grow into visible colonies, and 10 days after seeding the colonies were fixed in methanol and stained with crystal violet. The wells were scanned to obtain representative images, and the number of colonies in each well was counted. This was expressed as % colonies survival over the initial number of cells seeded.

DNA and GST Pull-downs—DNA pull-downs were performed mainly as described in Ref. 12. For each DNA pull-down, 2.5 μg of biotinylated oligos containing 3 repeats of the DNA sequence with interspersed spacer (CONS1: GTCGCG, CONS2: GTCGCA, MUT: GCAGCG, SCR:GGGCCT) was immobilized on 25 μl of Dynabeads MyOne C1 (Invitrogen) by incubating for 1 h at room temperature in a total volume of 350 μl of DNA binding buffer (1 M NaCl, 10 mM Tris-HCl, pH 8 , 1 mM EDTA, pH 8, and 0.05% Nonidet P-40). Beads containing immobilized DNA were then incubated with GST-lysates expressing the recombinant ZNF131 domains in a total volume of 600 μl of protein binding buffer (50 mM

---

[6] The abbreviations used are: DSB, double strand break; LFQ, label-free quantitation; ChIP, chromatin immunoprecipitation.
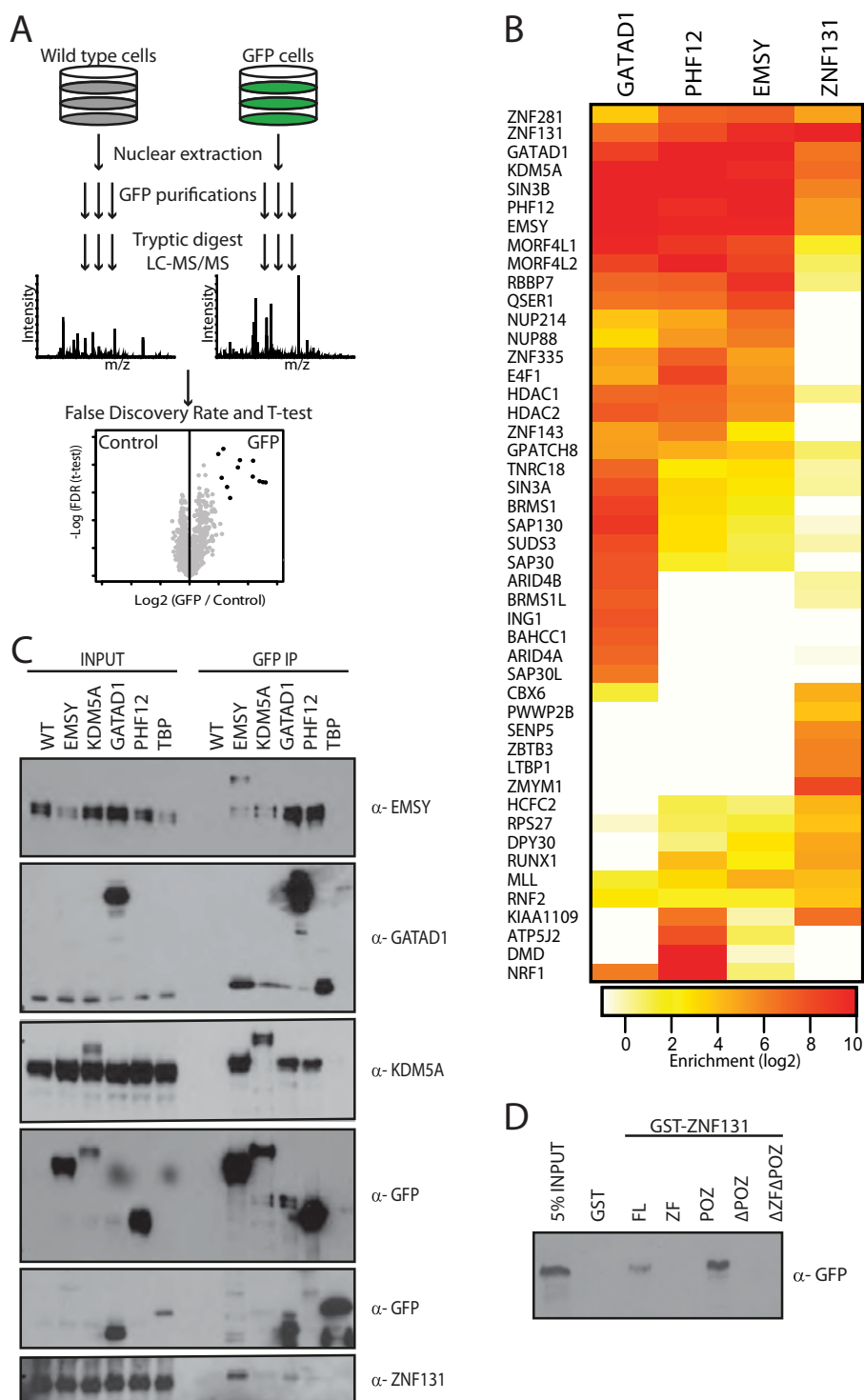
FIGURE 1. **Characterization of EMSY/KDM5A/SIN3B complex.** *A*, schematic representation of the experimental setup. *B*, heatmap showing the average LFQ intensities of the EMSY interactors (indicated to the *left*) in GFP purifications of different EMSY complex subunits (indicated on *top*). In each of these pull-downs, a set of largely overlapping interactors is detected, indicating that these proteins form a stable complex. *C*, co-immunoprecipitation and Western blotting experiments to validate the interactions that were detected by label free mass spectrometry. *WT*, wild-type extract. Since GATAD1 and TBP are smaller proteins, the GFP probing was done on two different blots to detect the higher (*top GFP blot*) and smaller (*bottom GFP blot*) molecular weight proteins, respectively. *D*, ZNF131 enriches GFP-EMSY from HeLa nuclear extracts. GST-tagged ZNF131 fusion proteins were immobilized on glutathione agarose beads and subsequently used for pull-downs in nuclear extracts from HeLa cells expressing GFP-tagged EMSY and probed with GFP antibody. These experiments clearly revealed that the ZNF131 POZ domain is both necessary and sufficient to mediate a direct interaction with EMSY or an EMSY-interacting protein.

Tris-HCl pH 8, 150 mM NaCl, 1 mM DTT, 0.25% Nonidet P-40, and Complete protease inhibitors (Roche, EDTA-free)) in the presence of 10 μg of poly-dAdT for 2 h at 4 °C. Baits were then washed three times with 0.5 ml of protein binding buffer after which protein bound to the beads was eluted and analyzed by Western blotting.

For GST pull-downs, recombinant GST-ZNF131 fusion proteins were expressed in bacteria and lysates were made. Soluble
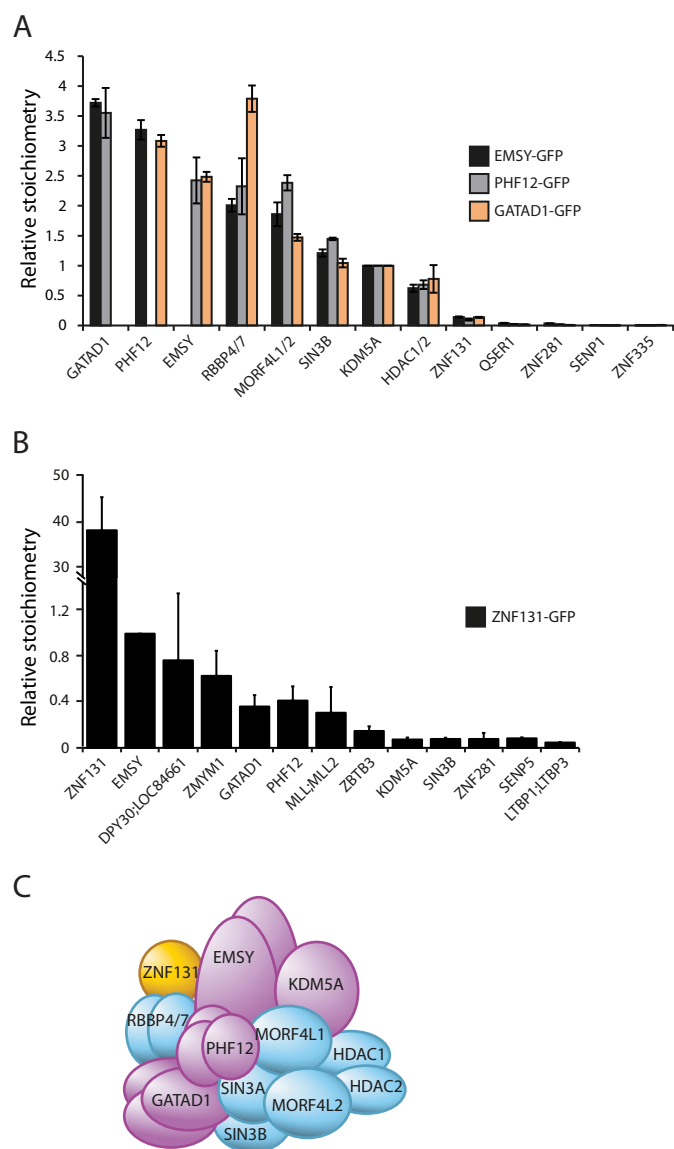
A

B

C

FIGURE 2. **Stoichiometry of the EMSY/KDM5A/SIN3B complex.** *A* and *B*, iBAQ-based stoichiometry estimation of interactors detected in different label-free GFP pull-downs. The *bar plot* shows the stoichiometry of the interactors relative to the levels of KDM5A (*A*) or EMSY (*B*). In *B*, a break in the *y*-axis is used to depict the different scale limits along the *y* axis. *C*, schematic representation of the core EMSY/KDM5A/SIN3B complex.

GST-tagged ZNF131 fusion proteins were then immobilized on glutathione-agarose beads and subsequently used for binding proteins from nuclear extracts of HeLa cells expressing GFP-tagged EMSY. The bound proteins were eluted off the GA beads and analyzed by Western blotting by probing with a GFP antibody.

ChIP-qPCR and ChIP-Seq—HeLa Kyoto wild type (WT), HeLa EMSY GFP, and HeLa EMSY GFP with knockdown of ZNF131 were seeded in 15-cm dishes and 24 h later were formaldehyde cross-linked for 10 min. Chromatin obtained from the cells after sonication was subjected to overnight IP using GFP antibody (Abcam, ab290) followed by extensive washes and decross-linking. DNA was purified using the Qiagen PCR clean-up kit and the bound DNA was analyzed by qPCR using the desired primers. ChIP-Seq was carried out by conventional

ChIP followed by end repair of 15 ng (for protein-GFP fusions) or 30 ng (for the histone modifications) enriched DNA as measured by Qubit fluorometer using the Quant-iT dsDNA HS Assay Kit from (Invitrogen, Q32851). Adaptors were ligated to DNA fragments, which were subsequently size selected (~300 base pair (bp)). The adaptor-modified DNA fragments were subjected to limited PCR amplification (14 cycles) and quality control was made by qPCR (primers sequences are available upon request), as well as by running the PCR products on a Bioanalyzer (Bio-Rad). Finally, cluster generation and sequencing-by-synthesis (36 bp) was performed using the Illumina Genome Analyzer IIx (GAIIx) according to standard protocols of the manufacturer (Illumina). The image files generated by the Genome Analyzer were processed to extract DNA sequence data. Sequences were aligned to the human reference genome (GRCh37/hg19, Feb 2009) with Burrows-Wheeler Aligner1 (bwa,v0.5.9-r16) allowing one mismatch. Uniquely aligned reads were converted to BED format. The total number of sequenced fragments and mapped fragments are shown in supplemental Table S5. All ChIP and input samples were normalized randomly to the same number of reads (10543138). Furthermore, reads were directionally extended to 300 bp, and for each base pair in the genome the number of overlapping sequence reads was determined and averaged over a 10-bp window to create a Wiggle (WIG) file to visualize the data in the University of California Santa Cruz (UCSC) Genome Browser.

Peak Detection—Identification of the binding sites (peak calling) was performed using MACS (13) (version 2.0.9 20111102, tag:alpha) with parameters: -g 2.7e9; -m 10,30; -q 0.05. Genomic regions overlapping to the peaks in the negative control (GFP ChIP-Seq on wild type HeLa cells) were filtered out. Genomic annotation was carried out with Hypergeometric Optimization of Motif EnRichment (HOMER, software v4.2). The tool annotatePeaks was used with parameters by default and defined in the help. A gtf file from GENCODE project based on release 15 (GRCh37/hg19) was used for annotations; the latter included whether a segment is in the TSS/promoter, TTS, exon, 5′ UTR exon, 3′ UTR exon, intron, or is intergenic. Since some annotations overlap, the following priority was assigned: TSS/promoter (from −1 kb to +100 bp), TTS (from −100 bp to +1 kb), CDS exon, 5′ UTR exon, 3′ UTR exon, intron, intergenic. The heatmaps and distribution plot representations were carried out with R package v2.14.1, SeqMiner v1.3.3e (14) and fluff.[7] Heatmaps represent summed reads in 100 bp sliding windows flanking 5 kb to the EMSY- or H3K4me3-bound sites. The overlap between EMSY- and GATAD1-binding sites was calculated with IntervalStats v1.01 (16) using the EMSY-binding sites as reference file. The *p* value applied for overlapping was 0.05.

Pathway Enrichment Analysis—To identify the biological relevant pathways of EMSY-targeted genes we used the functional annotation tool available in the DAVID bioinformatics resources v6.7 (17). DAVID identifies the most relevant pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (18), REACTOME (19) and Gene Ontology (20) data-

---
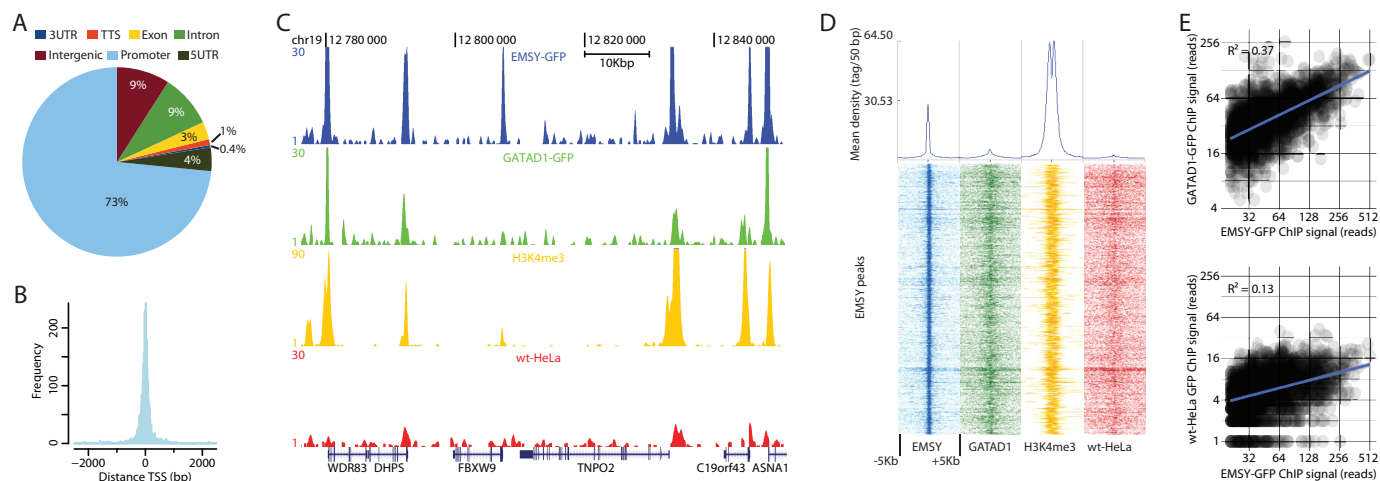
FIGURE 3. **ChIP-sequencing of EMSY.** *A*, *pie-chart* showing the genomic annotation of EMSY-binding sites. *B*, histogram of EMSY peak distribution in relation to the closest annotated transcription start site (TSS) for each detected binding site. *C*, snapshot from the UCSC browser showing co-localization of EMSY-GFP (*blue*), GATAD1-GFP (*green*), and H3K4me3 (*yellow*). GFP ChIP-seq on WT HeLa cells is shown in *red*. *D*, ChIP-sequencing of EMSY-GFP (*blue*), GATAD1-GFP (*green*), H3K4me3 (*yellow*), and GFP in HeLa WT cells (*red*). The shown heatmaps are based on all detected EMSY ChIP-seq peaks. All the distribution plots are −5 to +5 kb flanking EMSY peaks, reads were summed in 100-bp sliding windows. *E*, read count correlation analysis of EMSY-GFP ChIP signal *versus* GATAD1-GFP ChIP signal in identified EMSY peaks (*top panel*) and EMSY-GFP ChIP signal *versus* GFP ChIP signal in WT HeLa cells (*bottom panel*).

bases. Pathways with a Benjamini-Hochberg corrected *p* value lower than 0.1 were considered significant.

Motif Enrichment Analysis—The ZNF131 DNA-binding motifs were obtained from Donaldson *et al.* (21). We calculated the frequency of the motif in the EMSY-binding sites. To determine that the frequency observed is not random chance, we generated 100 simulations of random genomic sequences adjusted by the number and length of it for EMSY-binding sites defined with MACS (see above). For this purpose, we used the subcommands shuffle and getfasta included in bedtools v2.18.2. Further, we tested the differences with the Fisher's exact test (*p* value = 0.004) and generated the distribution plots in R package v2.14.1.

mRNA Expression Analysis—500 ng of RNA extracted from mammalian cells using the QIAgen RNA extraction kit was used for cDNA synthesis using random primers and Superscript II from the Invitrogen first-strand cDNA synthesis kit. mRNA expression of EMSY and the EMSY target genes was analyzed by qPCR using the primers designed for EMSY and EMSY target genes and the values obtained were normalized over actin expression. Gene expression data of 6791 tumor samples analyzed according to the RnaSeqV2′ pipeline was downloaded from The Cancer Genome Atlas (TCGA) data portal at 14-05-2014 to investigate the co-regulation of EMSY subunits across different cancers.

Patient Material—Tissue samples of invasive breast cancer patients were collected between November 2004 and September 2009 at the Department of Pathology of the University Medical Centre in Utrecht (UMCU), The Netherlands. This study selected 103 tissue samples from this consecutive series. According to the Dutch Federation of Medical Scientific Societies (the FEDERA), use of redundant tissue for research purposes does not require informed consent, if the patients are offered the possibilities to refuse this ("opting out" system) and material has been used anonymously or coded. In the UMC Utrecht, all patients are informed that research may happen with their redundant tissue, and are offered to opt out. No

material of patients that have opted out was used in the present study, and all materials were used anonymously. The research protocol for this study was approved by the Scientific Advisory Counsel of the UMC Utrecht Biobank. Grading was performed according to the Nottingham modification of Bloom-Richardson system, and mitoses were counted according to a strict protocol as before to arrive at the mitotic activity index (MAI) (23).

Immunohistochemistry—After deparaffinization and rehydration, antigen retrieval was performed using citrate buffer for EMSY, SIN3B, and ZNF131, and EDTA buffer for GATAD1, KDM5A, at boiling temperature for 20 min. A cooling period of 30 min preceded the incubation of the slides with protein block for EMSY, GATAD1, and KDM5A (Novolink Max Polymer detection system, ready to use, Novocastra Laboratories Ltd, Newcastle Upon Tyne, UK) for 5 min at room temperature. Incubation of the slides with the EMSY antibody (rabbit polyclonal, #A300-253A, Bethyl Laboratories) was done at a dilution of 1:100, for KDM5A (rabbit monoclonal, # 3876S, Cell Signaling) at a dilution of 1:50 and for GATAD1 (mouse monoclonal, # sc-81092, Santa Cruz Biotechnology) at a dilution of 1:50, for 60 min at room temperature. For detection, a polymer (Novolink Max Polymer detection system, ready to use) was used. For ZNF131 (mouse polyclonal, # H0000 7690-B0IP, Abnova) and SIN3B (in-house), the slides were incubated with primary antibodies at a dilution of 1:50 for both of them for 60 min at room temperature. For detection a poly HRP anti-mouse/rabbit/rat IgG (Brightvision, ready to use, ImmunoLogic, Duiven, Netherlands) was used. All slides were developed with diaminobenzidine (5 min for Novolink protocol and 10 min for Brightvision) followed by hematoxylin counterstaining. Before the slides were mounted all sections were dehydrated in alcohol and xylene. Positive controls were used throughout; negative controls were obtained by omission of the primary antibodies from the staining procedure.
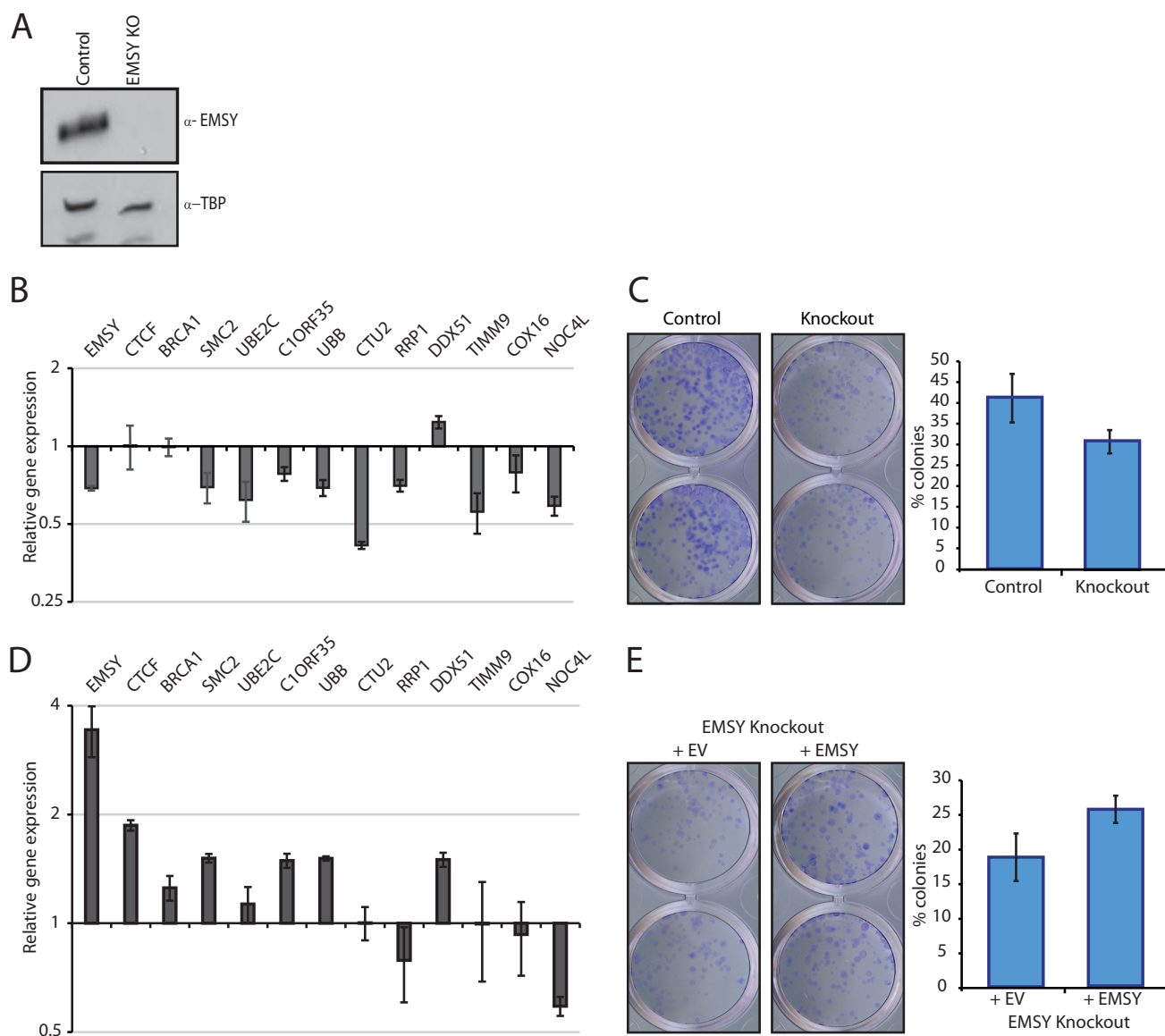
FIGURE 4. **Effect of EMSY perturbation on target gene expression and cell proliferation.** *A*, Western blot depicting the CRISPR-based knock-out of EMSY in HeLa cells. *B*, effect of deletion of EMSY from HeLa cells on mRNA expression of a selection of EMSY target genes. mRNA expression values were normalized over actin and plotted as fold mRNA changes relative to a control knock-out line (empty vector, *EV*). Values are an average from two independent experiments (±S.E. indicated). *C*, effect of deletion of EMSY from HeLa cells on colony-forming ability (*left*). Quantification of the number of colonies from an average of four independent experiments (*right*). *D*, re-expression of EMSY in an EMSY knock-out line partially rescues gene expression profiles and colony forming ability (*E*).

Scoring of immunohistochemistry was performed by one observer (PJvD). For EMSY, KDM5A, SIN3B, ZNF131, GATAD1, the percentage of positively stained nuclei was estimated, furthermore cytoplasmic expression of EMSY was also scored in a scale of 0, 1, 2, and 3. In the statistical analysis we used the product of the EMSY nuclear and cytoplasmic scoring data. Associations among staining were tested by Chi-square analysis.

Results

Recently, we described protein-protein interactions between the H3K4me3 demethylase KDM5A, the protein EMSY, GATAD1, and components of the SIN3/HDAC complex in HeLa cells (3). To further characterize these interactions, we generated BAC-GFP transgenic HeLa cell lines for EMSY,

KDM5A, PHF12, and GATAD1. Nuclear extracts derived from these cells were subjected to single affinity GFP purifications coupled to label-free quantitative mass spectrometry (Fig. 1*A*). Similar, largely overlapping interactions were observed in each case, indicating that these proteins form a large multi-subunit protein complex (Fig. 1*B*, supplemental Table S1, for KDM5A data not shown). This protein complex resembles the recently characterized *Drosophila* Lid/Sin3 complex (4), as well as the SIN3B uniCORE complex reported previously (24). However, neither BRCA2 or HP1 that were initially reported as EMSY interactors (5) were found in our experiments, which we assume is due to differences in experimental systems. Given the reported link between EMSY and breast cancer, we also purified EMSY-GFP from a breast cancer cell line, MCF7 (data not shown) and observed interactions almost identical to those
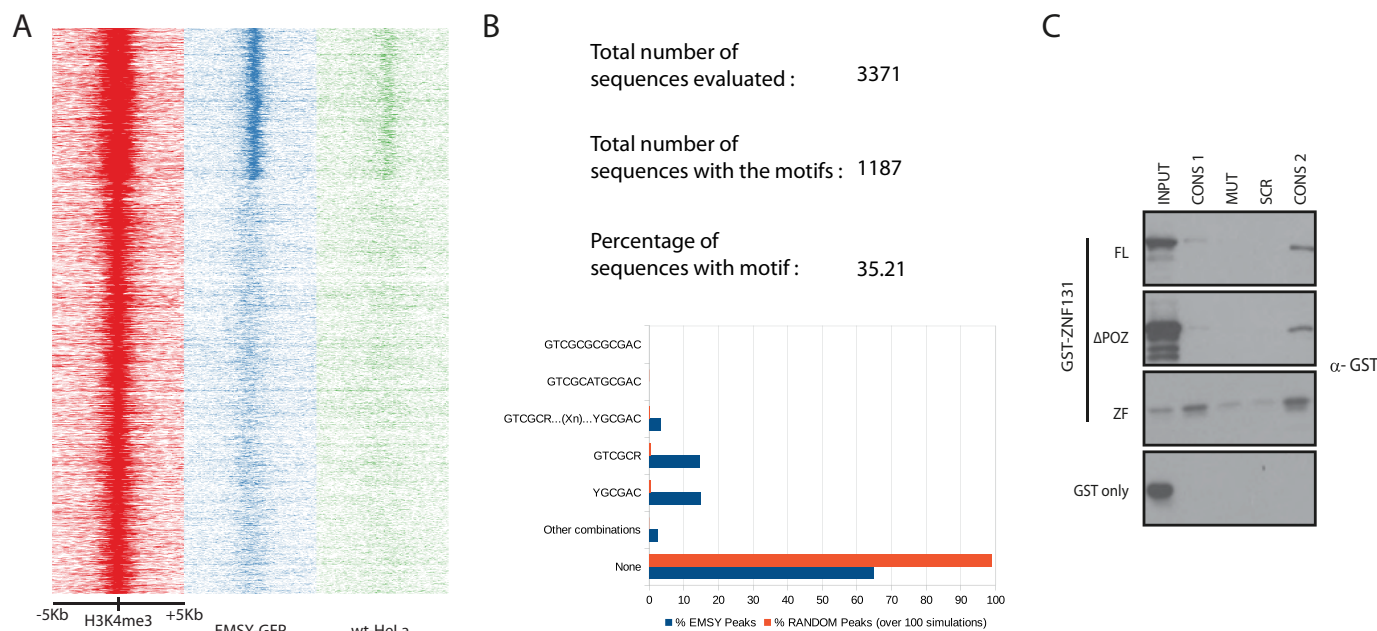
FIGURE 5. **Recruitment of EMSY to its target genes may be regulated by ZNF131.** *A*, heatmap analysis of H3K4me3 (*red*), EMSY-GFP (*blue*), and GFP in WT HeLa cells (*green*). The shown heatmaps are based on all detected H3K4me3 peaks ($\pm$ 5000 bp). EMSY was found to occupy only a subset of these H3K4me3 marked genes. *B*, ZNF131 motif search in EMSY-binding sites. Distribution of complete and partial palindromic motif sequences (defined by Ref. 20 in EMSY-binding sites and average in 100 random sequences. *C*, biotinylated oligos with either a consensus ZNF131 DNA-binding motif (CONS1 and CONS 2), a consensus sequence bearing a single mutation (MUT) or a scrambled sequence (*SCR*) were used in a binding assay with different domains of ZNF131 fused to GST. These experiments revealed that the zinc finger domain of ZNF131 directly interacts with the ZNF131 consensus binding site found in the EMSY ChIP-seq data set.

observed in HeLa cells. Given the very low expression of the EMSY-GFP BAC in MCF7 cells, we used HeLa cells for most of our subsequent analyses.

Interestingly, a DNA-binding zinc finger protein, ZNF131, was consistently identified in the affinity purifications, indicating that this protein may be a novel EMSY/KDM5A interactor. To validate this interaction between ZNF131 and EMSY/KDM5A we generated a stable MCF7 cell line expressing ZNF131-GFP at sub-endogenous levels. As shown in Fig. 1*B*, EMSY/KDM5A components co-purified with GFP-ZNF131, as well as some subunits of the MLL H3K4me3 methyltransferase complex (MLL, DPY-30). We did not, however detect the previously reported ZNF131 interactor Kaiso (21). Nevertheless, a number of other putative DNA-binding proteins were identified as interactors, including ZMYM1, ZBTB3, and ZNF281. Some of these observed interactions were also validated using Western blotting (Fig. 1*C*). ZNF131 contains a number of domains, including an N-terminal POZ domain and five C-terminal zinc fingers (21). The N-terminal POZ domain enriches GFP-EMSY from HeLa nuclear extracts, indicating that this domain interacts either directly with EMSY or plausibly via an EMSY interacting protein (Fig. 1*D*).

iBAQ-based stoichiometry estimation of the detected interactions revealed that the core EMSY complex consists of GATAD1, PHF12, EMSY, RBBP4/7, MORF4L1/2, SIN3B, KDM5A, and HDAC1/2 (Fig. 2*A*). Several subunits in the complex, appear to be multimeric, including GATAD1, PHF12, EMSY, and RBBP4/7. Western blotting also revealed that EMSY-GFP co-purifies endogenous EMSY, thus providing further evidence for multimerization (Fig. 1*C*, *top panel*). Furthermore, approximately equal amounts of endogenous EMSY

co-purifies with GFP-EMSY, illustrating that EMSY-GFP is expressed at near-endogenous levels. ZNF131 and the other zinc finger proteins are substoichiometric, the intensity of ZNF131 being ~10% relative to the core subunits, for example. Stoichiometry analysis of the GFP-ZNF131 pull-down revealed that EMSY and MLL complex subunits as well as ZMYM1 are the major ZNF131 interactors (Fig. 2*B*). Altogether, these results establish a core EMSY interactome and in addition, a number of putative DNA-binding, substoichiometric interactors (Fig. 2*C*).

To gain further insights into the biology of the EMSY complex, we performed ChIP-sequencing on EMSY-GFP cells. In the HeLa genome, 3371 binding sites for EMSY were identified, the large majority of which (73%) are close to transcription start sites (Fig. 3, *A* and *B*, [supplemental Table S2](#)). Interestingly, these EMSY peaks on promoters coincide with high levels of H3K4me3 (profiled in WT HeLa cells) and GATAD1, for which we performed ChIP-sequencing previously ((3), Fig. 3, *C* and *D*). As shown in Fig. 3*E*, we observed a strong genome-wide correlation between the GATAD1-GFP and EMSY-GFP signal on identified EMSY peaks (median length = 484 bp). This correlation ($R^2 = 0.37$; *top panel*) is three times stronger than one would expect from just the increased DNA accessibility of EMSY-bound promoters ($R^2 = 0.13$; *bottom panel*), indicating that the observed genome-wide association between EMSY and GATAD1 represents true co-localization of the two proteins. The genome-wide correlation between EMSY and H3K4me3 is intriguing, given the presence of the H3K4me3 demethylase KDM5A and histone deacetylases, in the EMSY complex. H3K4me3 demethylase activity and histone deacetylation are both enzymatic activities that are generally linked with gene
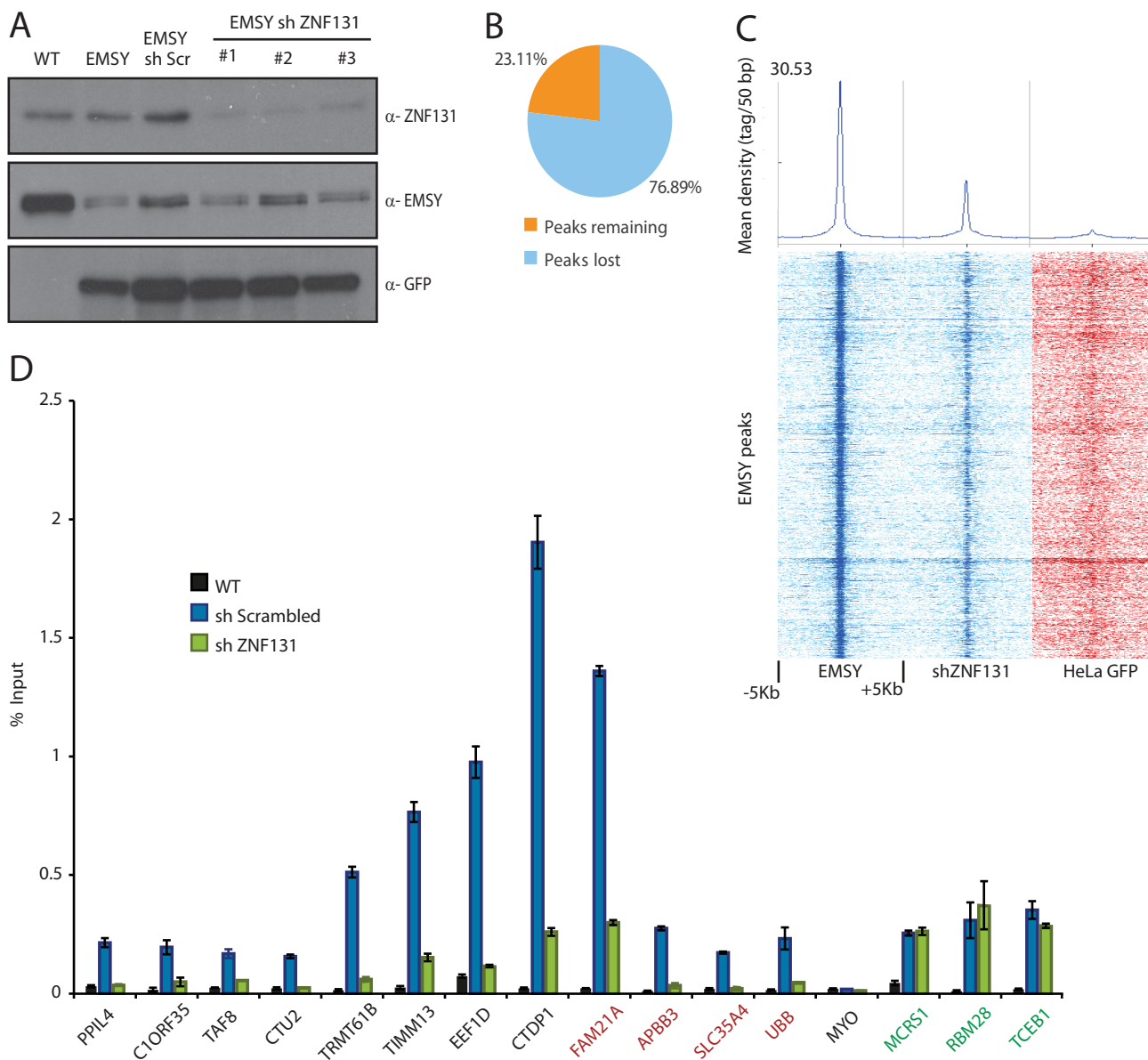
FIGURE 6. **ZNF131 recruits the EMSY complex to a subset of its genome wide target genes.** *A*, knockdown of ZNF131 in HeLa EMSY GFP using short hairpins. Sh ZNF131#1 is used for the experiment shown in *C* and *D*. (Note: We observe a decrease in expression of endogenous EMSY upon EMSY-GFP overexpression (*lane 1 versus lane 2*). *Lane 3* is slightly overloaded compared with other lanes). *B*, representation of the percent-age of EMSY peaks lost after ZNF131 knockdown in EMSY GFP-expressing HeLa cells. *C*, EMSY distribution and heatmap plots at EMSY peaks in WT HeLa and HeLa cells expressing EMSY GFP cells after ZNF131 knockdown. All the distribution plots are −5 to +5 kb flanking EMSY peaks, reads were summed in 100-bp sliding windows. *D*, ChIP-qPCR based validation of the ChIP-seq dataset. EMSY occupancy on target genes was determined by performing GFP ChIPs in WT cells and HeLa cells expressing EMSY-GFP with either a control knockdown (sh Scrambled) or ZNF131 knockdown (sh ZNF131). Genes with a ZNF131 consensus binding motif (*black label*), genes with no ZNF131-binding motif (*red label*), and genes that show no change in EMSY occupancy (*green label*) are shown. MYO (myoglobin gene locus) served as a negative control. These results were substantiated using another independent short hairpin, viz., shZNF131#2 (data not shown).

repression. Apparently, in the context of the EMSY complex, these activities are not strictly associated with stable gene repression. In support of these observations, recent genome wide profiling for histone deacetylases also revealed a correlation with actively transcribed genes (25). To further investigate the effects of the EMSY/KDM5A complex on target gene expression, we generated an EMSY knock-out HeLa cell line using CRISPR-Cas9 (Fig. 4*A*). We then analyzed expression of 12 EMSY target genes, as determined by ChIP-seq (supplemental Table S2), using qPCR. Interestingly, most of the analyzed EMSY target genes showed reduced expression in the absence of EMSY relative to the control cell line (Fig. 4*B*). We also tested

the proliferation capacity of the control and CRISPR knock-out line using a colony formation assay (Fig. 4*C*), which revealed a reduction in colony formation in the absence of EMSY, consistent with reports in (26) and given its reported function as a putative oncogene. To further substantiate these results we conducted rescue experiments by re-expressing EMSY in the CRISPR knock-out line. Re-expression of EMSY in the knock-out line in most cases resulted in up-regulation of EMSY complex target genes relative to a mock transfection (Fig. 4*D*). Furthermore, increased proliferation capacity was observed (Fig. 4*E*). These observations are consistent with our ChIP-seq data, which indicates a strong link between the EMSY/KDM5A com-
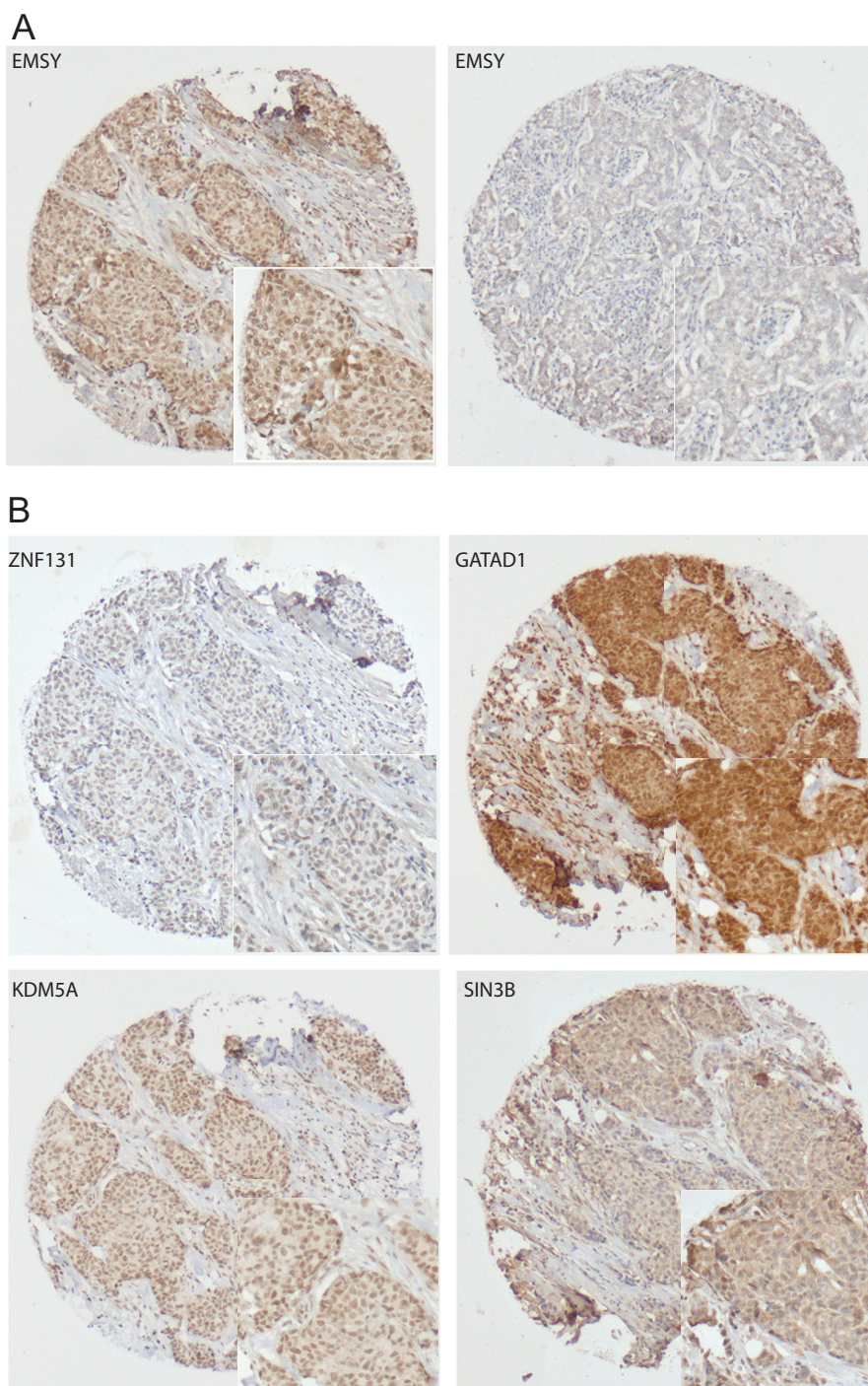
FIGURE 7. **Co-regulation of EMSY complex subunits in breast cancer.** *A*, representative immunohistochemical staining images of a patient with positive EMSY staining (*left*) and another case that stained negative for EMSY (*right*). *B*, case with positive EMSY staining in *A* was stained for other EMSY/KDM5A complex members. ZNF131, GATAD1, and KDM5A correlated positively with the EMSY status whereas SIN3B correlated negatively. The example in *A* that was negative for EMSY expression, was also negative for ZNF131, GATAD1, KDM5A, and SIN3B (data not shown), thus revealing a positive expression correlation between the EMSY/KDM5A complex subunits in primary breast cancer.

plex and genes related to cell proliferation and/or cell cycle regulation (supplemental Table S3). Altogether, these results imply that the EMSY/KDM5A complex is important for regulating target gene expression rather than exclusive transcriptional repression, however this may not be true for all target genes. In addition, the presence of EMSY is important for cell proliferation as also shown in Ref. 26.

Although we observed a strong genome-wide correlation between the EMSY complex and H3K4me3, not all H3K4me3-marked genes are bound by the EMSY complex (Fig. 5A). This implies that other factors, such as DNA-binding proteins in the complex, contribute to achieve genome-wide binding specificity. An ISMARA analysis of the DNA sequences in the EMSY peaks showed enrichment for the DNA-binding motifs of NRF1

		EMSY product		
		neg	pos	p-value
Age	<58	21	23	
	>58	18	18	0.554
Type	Ductal	35	29	
	Ductal-lobular	1	4	
	Lobular	3	4	
	mucinous	0	1	0.322
Grade	1	3	3	
	2	10	12	
	3	25	23	0.876
ZNF131	neg	28	10	
	pos	9	28	0.000
GATAD1	neg	28	11	
	pos	10	27	0.000
KDM5A	neg	25	12	
	pos	13	25	0.004
SIN3B	neg	34	24	
	pos	5	14	0.015

FIGURE 8. **Table showing the statistical significance in the correlation between the expression of EMSY and other complex members in primary breast cancer tissue microarrays.**

FIGURE 9. **EMSY complex subunits are co-regulated across thousands of tumor samples.** *A*, snapshot from the UCSC genome browser showing co-localization of EMSY-GFP and GATAD1-GFP on the KDM5A promoter. GFP ChIP-seq on WT HeLa cells (control) is shown in *red*. *B*, histogram showing genome-wide mRNA expression correlations with EMSY in 6791 TCGA tumor samples. The EMSY complex subunits are among the highest correlating genes, indicating strong co-regulation. The *x*-axis represents the squared Pearson correlation to EMSY expression over 120 equal bins, the *y*-axis represents gene frequency. *C–F*, gene expression correlations between EMSY and *C*) ZNF131, *D*) GATAD1, *E*) KDM5A, and *F*) SIN3B.

and ZNF335 (data not shown). Interestingly, both of these proteins were found as interactors of the EMSY/KDM5A/SIN3B complex in our study (Fig. 1B). However, the most prominent putative DNA-binding protein in the EMSY complex, as revealed by EMSY complex stoichiometry analyses, is ZNF131 (Fig. 2A). The DNA binding specificity of this protein has been previously determined using the CAST approach (21). Interestingly, 35% of the genome wide EMSY peaks contain a motif which resembles this ZNF131 consensus binding site (Fig. 5B). This implies that the ZNF131 protein may recruit the EMSY complex to a subset of its genome wide targets. In support of this, we confirmed that the recombinant zinc finger domain of ZNF131 specifically binds to consensus DNA-binding sites (GTCGCG and GGGCCT) that are enriched in the EMSY ChIP-sequencing peaks (Fig. 5C). To investigate a potential role for ZNF131 in recruiting the EMSY complex to target sites in the genome, we generated a stable shRNA-based ZNF131 knock-down in the GFP-EMSY HeLa BAC line (Fig. 6A). This cell line was subjected to GFP-based ChIP-sequencing. As shown in Fig. 6, *B* and *C*, ZNF131 knock-down resulted in a substantial loss of genome-wide EMSY binding (supplemental Table S4). These observations were validated using ChIP-qPCR (Fig. 6D). In summary, these results imply that the DNA-binding ZNF131 protein regulates recruitment of the EMSY complex to a substantial subset of genome-wide target genes.

Previously, the EMSY protein was shown to be amplified in sporadic breast cancer (5). While mining the EMSY ChIP-seq
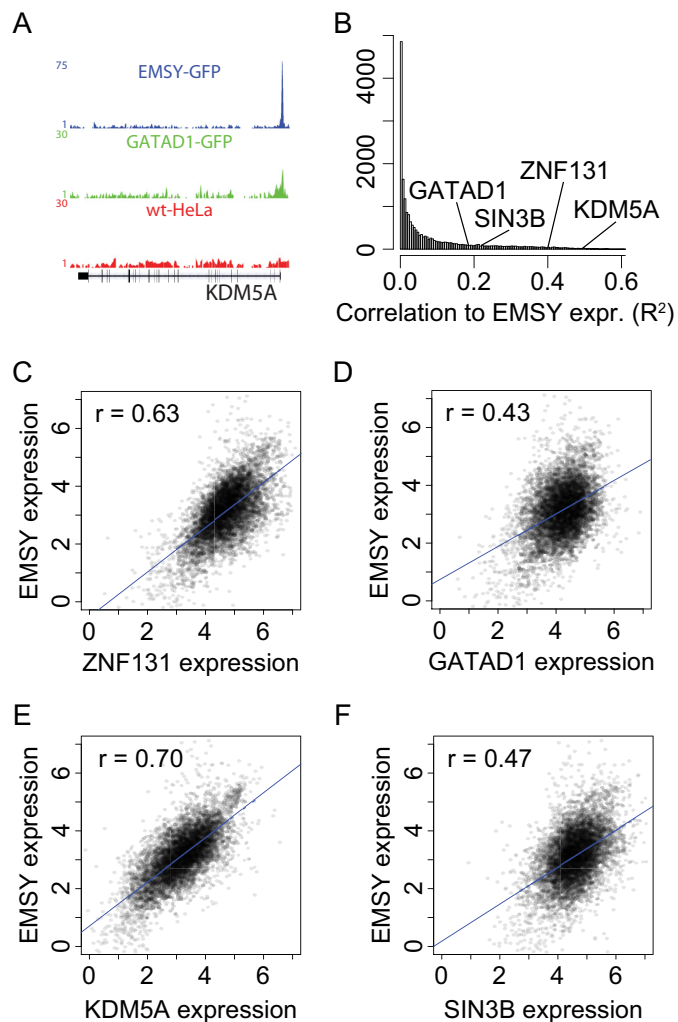
data, we observed that the EMSY protein binds to the promoter of a number of EMSY complex subunits, including KDM5A (Fig. 9A), SIN3B and HDAC1 (data not shown). Given the fact that EMSY expression is in most cases positively correlated with expression of its target genes (Fig. 4B), this prompted us to investigate a potential deregulation of EMSY complex subunits in breast cancer. To this end, we performed immunohistochemistry staining of primary breast tissue microarrays, comprising samples from 103 patients (Fig. 7). We started out with a consecutive cohort of 103 breast cancer cases mostly of ductal type with mild to aggressive morphology (grade 1–3). While screening this cohort for the expression of the markers of interest we lost about 29 cases due to loss of tissue material. The breast cancer patient cohort could be divided into two groups: EMSY positive 51% (39/77) or negative 49% (38/77) breast cancer cases (see Fig. 7A for example staining). For ZNF131 and

GATAD1 expression 51% (37/75 & 37/76) of the cases were positive and 49% (38/75 & 37/75) were negative. Overexpression of KDM5A was seen in 51% (38/75) of the cases. Most of the cases 75% (58/77) were SIN3B negative (see Fig. 7*B* for some example staining). In summary, this data reveals that EMSY overexpression in breast cancer is often accompanied by overexpression of EMSY complex members (Fig. 8), viz. ZNF131 (*p* value <0.001), GATAD1 (*p* value <0.001), KDM5A (0.004) and SIN3B (0.015), suggesting a cooperative function in relation to cancer. Investigating RNA-seq-based expression data for EMSY complex subunits across thousands of tumor samples from the TCGA Research Network further substantiates this observed positive expression correlation between EMSY complex subunits (Fig. 9).

Discussion

Here, we have shown that the EMSY/KDM5A complex mainly binds to H3K4me3-marked, actively transcribed genes in the mammalian genome. Given the enzymatic activities in the complex, a H3K4me3 demethylase and histone deacetylases, these observations are somewhat surprising. However, recent evidence suggests that histone deacetylase activity and co-repressor complexes are not exclusively linked to gene silencing but can also support active transcription (27). The molecular mechanisms underlying this apparent biological versatility are not clear yet. Mammalian transcription has been suggested to involve cyclical waves of acetylation/methylation-deacetylation/demethylation (25, 28) and the EMSY complex may also be involved in this process. Especially our interactomics data tempt to suggest that ZNF131 on one hand, via its association with the MLL complex, could be involved in directing the recruitment of the chromatin writers to genomic loci while on the other hand dictating the targeting of the EMSY/KDM5A eraser complex. Further experiments using synchronized cells and time-course experiments should shed further light on this. It should also be noted that ChIP-sequencing provides correlative data and observed signals represent average binding profiles obtained from thousands of asynchronous cells. Nevertheless, EMSY knock-out/rescue experiments further substantiated a positive link between the EMSY complex and gene expression (Fig. 4, *B* and *D*). In contrast, a recent study linked EMSY to transcriptional repression of a miRNA (26). Our experiments also show that in a minority of cases, EMSY expression is negatively correlated with target gene expression. Further experiments are imperative to decipher the molecular mechanisms of EMSY complex-mediated regulation of gene expression.

ChIP-seq revealed that EMSY mainly binds to transcription start sites and coinciding with H3K4me3. Importantly, knockdown of ZNF131 resulted in a substantial reduction of EMSY peaks in EMSY-GFP ChIP-seq experiments suggesting a requirement for ZNF131 as a recruiting factor. However, in our study, several DNA-binding proteins apart from ZNF131 viz., ZMYM1, ZBTB3, ZNF281, NRF1, and ZNF335 have also been identified as EMSY and/or ZNF131 interacting proteins. Hence it is plausible that some of these DNA-binding proteins may also have a role in recruitment of the EMSY complex to its genomic-binding sites. Consistent with this, we observed that

upon ZNF131 knockdown, EMSY recruitment to a subset of its target genes is not affected (Fig. 6*D*).

Immunohistochemistry experiments on primary human breast cancer cases revealed a striking co-regulation of EMSY/KDM5A complex members in a subset of EMSY-positive breast cancers, which further substantiates our proteomics data and suggests that EMSY complex subunits may functionally cooperate in sporadic breast cancers which are characterized by EMSY amplification. The presence of histone deacetylases, HDAC1 and HDAC2, and a histone lysine demethylase, KDM5A, in the EMSY complex is worth mentioning from a clinical perspective. Histone deacetylase inhibitors are already used in the clinic to treat a variety of malignancies (29), and our data suggest that breast cancer patients that show EMSY amplification may benefit from treatment with histone deacetylase inhibitors as well. Currently, efforts are also being made to develop specific inhibitors targeting histone demethylases, which are frequently deregulated in cancer (30). Combining HDAC inhibitors with specific inhibitors targeting the KDM5A demethylase may be an interesting epigenetic combination therapy in sporadic breast cancers as well as other cancers that show EMSY amplification. This is particularly relevant given the fact the KDM5A demethylase and histone deacetylases have been linked to drug resistance in cancer (31). Treatment of cancers that display EMSY/KDM5A over-expression with HDAC inhibitors and a KDM5A inhibitor might therefore serve a dual purpose: counteracting EMSY/KDM5A/SIN3B overexpression, thereby inhibiting cell proliferation, while at the same time preventing cancer cell drug resistance.

References

1. Kouzarides, T. (2007) Chromatin modifications and their function. *Cell* **128,** 693–705
2. Ruthenburg, A. J., Li, H., Patel, D. J., and Allis, C. D. (2007) Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol.* **8,** 983–994
3. Vermeulen, M., Eberl, H. C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K. K., Olsen, J. V., Hyman, A. A., Stunnenberg, H. G., and Mann, M. (2010) Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* **142,** 967–980
4. Moshkin, Y. M., Kan, T. W., Goodfellow, H., Bezstarosti, K., Maeda, R. K.,

Pilyugin, M., Karch, F., Bray, S. J., Demmers, J. A., and Verrijzer, C. P. (2009) Histone chaperones ASF1 and NAP1 differentially modulate removal of active histone marks by LID-RPD3 complexes during NOTCH silencing. *Mol. Cell* **35,** 782–793

5. Hughes-Davies, L., Huntsman, D., Ruas, M., Fuks, F., Bye, J., Chin, S. F., Milner, J., Brown, L. A., Hsu, F., Gilks, B., Nielsen, T., Schulzer, M., Chia, S., Ragaz, J., Cahn, A., Linger, L., Ozdag, H., Cattaneo, E., Jordanova, E. S., Schuuring, E., Yu, D. S., Venkitaraman, A., Ponder, B., Doherty, A., Aparicio, S., Bentley, D., Theillet, C., Ponting, C. P., Caldas, C., and Kouzarides, T. (2003) EMSY links the BRCA2 pathway to sporadic breast and ovarian cancer. *Cell* **115,** 523–535

6. Kooistra, S. M., and Helin, K. (2012) Molecular mechanisms and potential functions of histone demethylases. *Nat. Rev. Mol. Cell Biol.* **13,** 297–311

7. Wang, G. G., Song, J., Wang, Z., Dormann, H. L., Casadio, F., Li, H., Luo, J. L., Patel, D. J., and Allis, C. D. (2009) Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. *Nature* **459,** 847–851

8. McDonel, P., Costello, I., and Hendrich, B. (2009) Keeping things quiet: roles of NuRD and Sin3 co-repressor complexes during mammalian development. *Int. J. Biochem. Cell Biol.* **41,** 108–116

9. Poser, I., Sarov, M., Hutchins, J. R., Hériché, J. K., Toyoda, Y., Pozniakovsky, A., Weigl, D., Nitzsche, A., Hegemann, B., Bird, A. W., Pelletier, L., Kittler, R., Hua, S., Naumann, R., Augsburg, M., Sykora, M. M., Hofemeister, H., Zhang, Y., Nasmyth, K., White, K. P., Dietzel, S., Mechtler, K., Durbin, R., Stewart, A. F., Peters, J. M., Buchholz, F., and Hyman, A. A. (2008) BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nature Methods* **5,** 409–415

10. Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8,** 2281–2308

11. Smits, A. H., Jansen, P. W., Poser, I., Hyman, A. A., and Vermeulen, M. (2013) Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Res.* **41,** e28

12. Spruijt, C. G., Gnerlich, F., Smits, A. H., Pfaffeneder, T., Jansen, P. W., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., Eberl, H. C., Mensinga, A., Brinkman, A. B., Lephikov, K., Müller, U., Walter, J., Boelens, R., van Ingen, H., Leonhardt, H., Carell, T., and Vermeulen, M. (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152,** 1146–1159

13. Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., and Liu, X. S. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137

14. Ye, T., Krebs, A. R., Choukrallah, M. A., Keime, C., Plewniak, F., Davidson, I., and Tora, L. (2011) seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Res.* **39,** e35

15. Deleted in proof

16. Chikina, M. D., and Troyanskaya, O. G. (2012) An effective statistical evaluation of ChIPseq dataset similarity. *Bioinformatics* **28,** 607–613

17. Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4,** 44–57

18. Kanehisa, M., and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28,** 27–30

19. Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., Jupe, S., Kalatskaya, I., Mahajan, S., May, B., Ndegwa, N., Schmidt, E., Shamovsky, V., Yung, C., Birney, E., Hermjakob, H., D'Eustachio, P., and Stein, L. (2011) Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* **39,** D691–697

20. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics* **25,** 25–29

21. Donaldson, N. S., Nordgaard, C. L., Pierre, C. C., Kelly, K. F., Robinson, S. C., Swystun, L., Henriquez, R., Graham, M., and Daniel, J. M. (2010) Kaiso regulates Znf131-mediated transcriptional activation. *Exp. Cell Res.* **316,** 1692–1705

22. Deleted in proof

23. van Diest, P. J., Fleege, J. C., and Baak, J. P. (1992) Syntactic structure analysis in invasive breast cancer: analysis of reproducibility, biologic background, and prognostic value. *Hum. Pathol.* **23,** 876–883

24. Malovannaya, A., Lanz, R. B., Jung, S. Y., Bulynko, Y., Le, N. T., Chan, D. W., Ding, C., Shi, Y., Yucer, N., Krenciute, G., Kim, B. J., Li, C., Chen, R., Li, W., Wang, Y., O'Malley, B. W., and Qin, J. (2011) Analysis of the human endogenous coregulator complexome. *Cell* **145,** 787–799

25. Wang, Z., Zang, C., Cui, K., Schones, D. E., Barski, A., Peng, W., and Zhao, K. (2009) Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* **138,** 1019–1031

26. Viré, E., Curtis, C., Davalos, V., Git, A., Robson, S., Villanueva, A., Vidal, A., Barbieri, I., Aparicio, S., Esteller, M., Caldas, C., and Kouzarides, T. (2014) The breast cancer oncogene EMSY represses transcription of antimetastatic microRNA miR-31. *Mol. Cell* **53,** 806–818

27. Reynolds, N., O'Shaughnessy, A., and Hendrich, B. (2013) Transcriptional repressors: multifaceted regulators of gene expression. *Development* **140,** 505–512

28. Métivier, R., Penot, G., Hübner, M. R., Reid, G., Brand, H., Kos, M., and Gannon, F. (2003) Estrogen receptor-α directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* **115,** 751–763

29. Falkenberg, K. J., and Johnstone, R. W. (2014) Histone deacetylases and their inhibitors in cancer, neurological diseases and immune disorders. *Nat. Rev. Drug Discov.* **13,** 673–691

30. Højfeldt, J. W., Agger, K., and Helin, K. (2013) Histone lysine demethylases as targets for anticancer therapy. *Nat. Rev. Drug Discov.* **12,** 917–930

31. Sharma, S. V., Lee, D. Y., Li, B., Quinlan, M. P., Takahashi, F., Maheswaran, S., McDermott, U., Azizian, N., Zou, L., Fischbach, M. A., Wong, K. K., Brandstetter, K., Wittner, B., Ramaswamy, S., Classon, M., and Settleman, J. (2010) A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* **141,** 69–80