Supporting Information

Beshiri et al. 10.1073/pnas.1216724109

SI Materials and Methods

Cell Culture, Reagents for Transfections, and Immunoblotting. U937 cells were grown in cell suspension at 37 °C in RPMI medium 1640 (Mediatech) supplemented with 10% (vol/vol) FBS (HyClone) and differentiated with 12-O-tetradecanoylphorbol-13-acetate (TPA; Sigma). Mouse embryonic fibroblasts (MEFs) were isolated from the mice of corresponding genotypes and used to prepare cell lysates at passage 6. MEFs, 293T cells, T98G, and SAOS-2 human osteosarcoma cells were grown in DMEM (Mediatech) supplemented with 10% FBS. For cell cycle analysis, human glioblastoma T98G cells were synchronized in G_o phase by serum starvation and restimulated as previously described (1). Nucleofection of U937 cells was performed with Nucleofector II using Nucleofection reagents from kit C or V using programs W-001 or V-001, respectively (Amaxa Biosystems), and SAOS-2 cells were transfected using Lipofectamine 2000 (Invitrogen) with lysine (K)-specific demethylase 5A (KDM5A) siRNA duplex sequences no. 4 and control siRNAs *GL3* and no. 1sc, as previously shown (2), or with *E2F4* siRNA 5′-GGCAGAGAUCGAGGAGCUGdTdT-3′ (Sigma) and siGENOME *LIN9* siRNA (M-018918-01; Dharmacon). For stable KDM5A knockdown, low passage 293T cells at 80% confluency were transfected by Lipofectamine 2000 with lentiviral expression and packaging constructs, pMD2.G, psPAX2, and pLVTHM-shKDM5A-GFP or the control pLVTHM-GFP vector. After 7 h, the transfection media was replaced with growth media (DMEM/10% FBS), 7 μg/mL polybrene, and 5×10^5 U937 cells per milliliter growth media. Following coculturing for 24 h, the transduced U937 cells were resuspended in RPMI/10% FBS in a T75 flask.

Protein lysates were prepared in 62 mM Tris•HCl (pH 6.8), 1.5% (wt/vol) SDS buffer followed by boiling. Next, 50 μg of protein was loaded onto a 6.25% (wt/vol) SDS/PAGE gel. Blots were developed using ECL and images were analyzed using ImageJ 1.42q software.

ChIP and Gene Expression Analyses. The generated *Kdm5a*$^{f/f}$ and *Kdm5a*$^{-/-}$ ES cells were maintained on mitomycin-treated MEFs and expanded for three passages on gelatinized-tissue culture plates without MEFs before proceeding to growing cells for ChIP. Cells were grown under standard mESC conditions, as described previously (3).

Two separate *Kdm5a*$^{f/f}$ clones and a single *Kdm5a*$^{-/-}$ clone were analyzed. Total genomic DNA sample was prepared from DNA of the two *Kdm5a*$^{f/f}$ clones mixed in a 1:1 ratio. Reference DNA was either the total genomic DNA sample or the *Kdm5a*$^{-/-}$ clone ChIP-seq sample. The genomic library was applied to the flow-cell of the Cluster Station from Illumina. After priming, flow cells were loaded into the Illumina Genome Analyzer II. Images acquired from the Illumina/Solexa sequencer were processed through the bundled Solexa image extraction pipeline, which performed base calling, and generated quality control statistics. Nine- and 24-million reads were generated for the two *Kdm5a*$^{f/f}$ clones, respectively, 22-million reads were generated for the *Kdm5a*$^{-/-}$ clone, and 25-million reads for the total genomic DNA. Sequenced 36-bp short reads were uniquely aligned allowing at best two mismatches to the University of California at Santa Cruz (The Genome Sequencing Consortium) reference genome (mm9) using the BOWTIE program (v0.12.2) (4). Sequences matched exactly more than one place with equal quality were discarded to avoid bias. Peak caller algorithm MACS (v1.3.7.1) (8) was used to determine the enriched peak region against reference DNA as background. Data for two *Kdm5a*$^{f/f}$ clones were combined. Peaks were modeled using default parameter with "futurefdr" flags on with "mfold" 10. Average ChIP signal along the 3-kb metagene (RefSeq) together with 1 kb upstream from the transcription start site (TSS) and 1 kb downstream from the transcription termination site (TTS) were plotted using CEAS software (6). Our data highly overlapped with the previously determined KDM5A binding sites in ES cells (7) (Dataset S2). From our 3,093 peaks and 1,764 peaks in the published study (GSM480163) (7), we found 1,065 overlap. MACS has been shown to be performing better than QuEST (8). Remapping GSM480163 data using BOWTIE and MACS, the same algorithms that we used for our data, resulted in almost half of our peaks (1,452 of 3,093 peaks) overlap with GSM480163 data, suggesting that KDM5A binding was reproducible at many regions. Using BOWTIE and MACS for the analysis GSM480163, we identified many more peaks (7,511 peaks) than this group reported using ELAND-QuEST.

Significance of expression changes in KDM5A genes was approached using the Wilcoxon rank sum test. In the previous ChIP-on-chip experiment, we used a self-printed array containing PCR generated probes representing the proximal promoter regions spanning between −500 bp and +500 bp of the TSS. Detailed information and all primer sequences are provided in Datasets S6, S7, and S8.

Location Analysis Data and Enrichment Analysis of Targets. E2F4 genomic locations were taken from the authors' analyzed genomic location BED file (9) and annotated to the closest gene (i.e., TSS) by ChIPpeakAnno. To analyze expression of genes displaying KDM5A binding (ChIP-seq data) in particular regions relative to the TSS, the following steps were performed: (*i*) the gene closest to the peak was determined by ChIPpeakAnno; (*ii*) KDM5A peaks were then classified according to distance from TSS and location of peak summit; (*iii*) differential expression of those genes was plotted. Overlap of KDM5A and E2F4 targets was presented using a Venn diagram and significance of overlap was determined by standard χ^2 test using R statistical programming language.

Functional annotation of target genes is based on Gene Ontology (GO) (10) (http://www.geneontology.org) as extracted from the EnsEMBL (11) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (12). We have taken only the GO/pathway categories that have at least 10 genes as targets annotated. Resulting *P* values were adjusted for multiple testing using the Benjamin and Hochberg's method of false-discovery rate (13).

1. Litovchick L, Chestukhin A, DeCaprio JA (2004) Glycogen synthase kinase 3 phosphorylates RBL2/p130 during quiescence. *Mol Cell Biol* 24(20):8970–8980.
2. Benevolenskaya EV, Murray HL, Branton P, Young RA, Kaelin WG, Jr. (2005) Binding of pRB to the PHD protein RBP2 promotes cellular differentiation. *Mol Cell* 18(6):623–635.
3. Lin W, et al. (2011) Loss of the retinoblastoma binding protein 2 (RBP2) histone demethylase suppresses tumorigenesis in mice lacking Rb1 or Men1. *Proc Natl Acad Sci USA* 108(33):13379–13386.
4. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.
5. Zhang Y, et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9(9):R137.
6. Shin H, Liu T, Manrai AK, Liu XS (2009) CEAS: *cis*-regulatory element annotation system. *Bioinformatics* 25(19):2605–2606.
7. Peng JC, et al. (2009) Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* 139(7):1290–1302.

8. Wilbanks EG, Facciotti MT (2010) Evaluation of algorithm performance in ChIP-seq peak detection. *PLoS ONE* 5(7):e11471.
9. Kim J, et al. (2010) A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. *Cell* 143(2):313–324.
10. Ashburner M, et al.; The Gene Ontology Consortium (2000) Gene Ontology: Tool for the unification of biology. *Nat Genet* 25(1):25–29.
11. Hubbard TJ, et al. (2007) Ensembl 2007. *Nucleic Acids Res* 35(Database issue): D610–D617.
12. Kanehisa M, et al. (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res* 36(Database issue):D480–D484.
13. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc, B* 57:289–300.

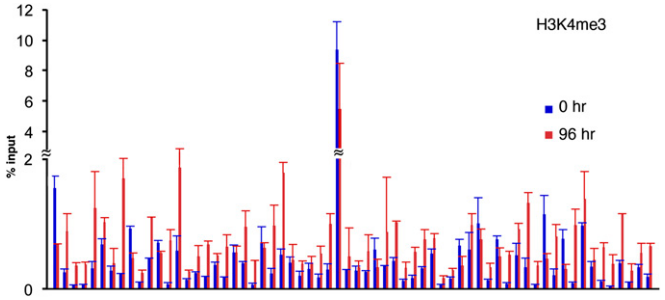

Fig. S1. Analysis of histone H3 lysine 4 trimethylation (H3K4me3) changes at KDM5A target genes during differentiation. The regions bound by KDM5A ($P < 0.002$) at 0 and 96 h after TPA addition (Dataset S8) were analyzed at these time points for H3K4me3. H3K4me3 enrichment for these regions was determined by ChIP-qPCR and shown as percent of input in sets of two, corresponding to 0 and 96 h. Error bars: means ± SEM, $n = 2$.

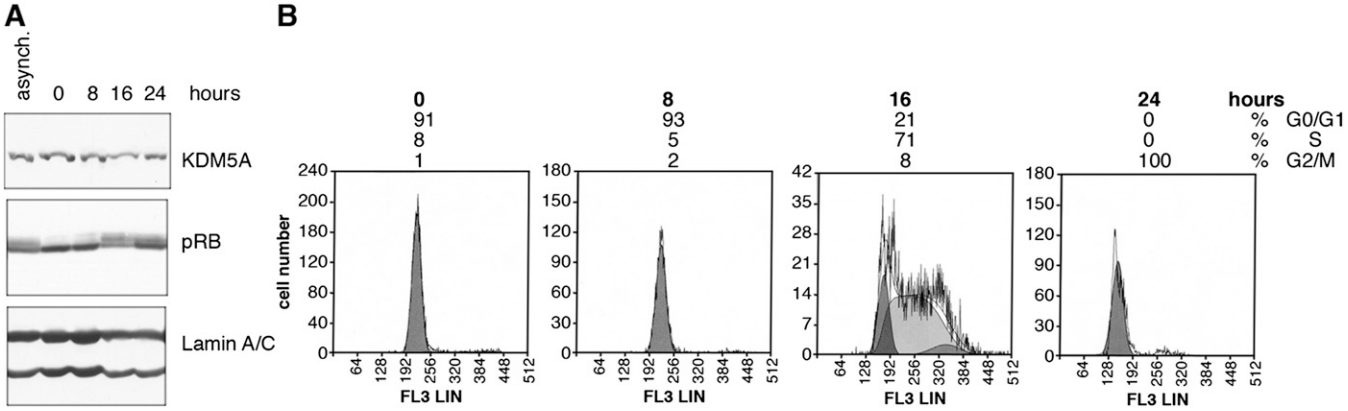

Fig. S2. (*A* and *B*) KDM5A protein level does not change during the cell cycle. Immunoblotting and FACS analysis was performed in T98G glioblastoma cells at different stages of the cell cycle. The retinoblastompa protein (pRB) underwent phosphorylation events, which is consistent with pRB function. We were unable to detect any changes in the KDM5A protein level.

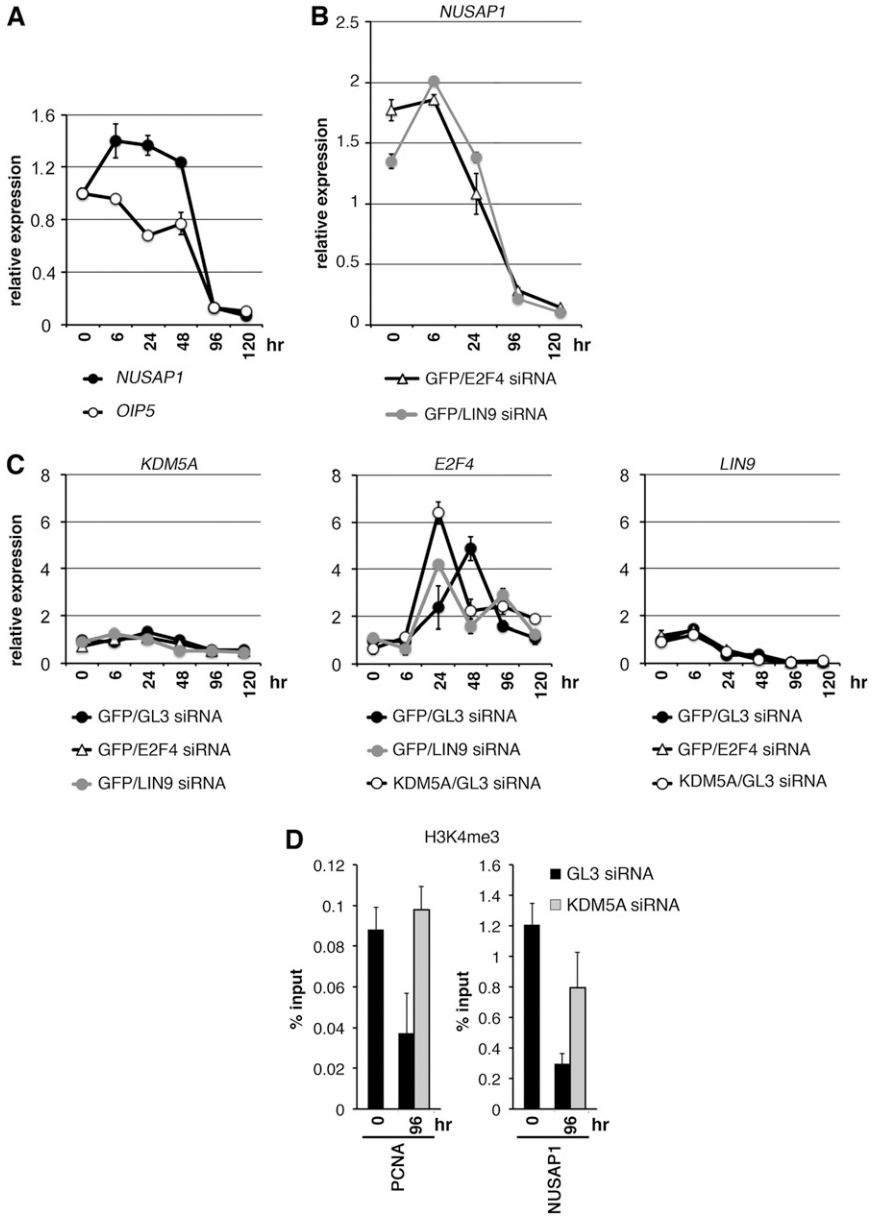

Fig. S3. Effect of gene knockdowns on gene expression. (*A*) Consistent with the notion that bidirectional genes contain *cis*-regulatory elements regulating both genes (1), *NUSAP1* (*nucleolar spindle-associated protein*) and *OIP5* are corepressed in the course of TPA treatment. (*B*) In differentiating cells, the effect of *LIN* siRNA on *NUSAP1* expression is similar to the effect of E2F4 siRNA. (*C*) Relative transcript level of *KDM5A*, *E2F4*, and *LIN9* genes in control cells and cells treated with siRNAs to *E2F4*, *LIN9*, and shRNA to *KDM5A*. (*D*) *KDM5A* knockdown results in increased methylation at both the *PCNA* (*proliferating cell nuclear antigen*) and *NUSAP1* promoters. The experiment for *NUSAP1* is a separate biological experiment from Fig. 4*G*. Analysis in *A*–*C* was performed at timed intervals after TPA addition, normalized to the level of *B2M*. In *D*, results are shown as a percent of input. Error bars: means ± SEM, $n = 3$.

1. Trinklein ND, et al. (2004) An abundance of bidirectional promoters in the human genome. *Genome Res* 14(1):62–66.

Dataset S1. List of KDM5A targets determined by ChIP-seq in mouse ES cells

Dataset S1

Dataset S2. List of overlapping ChIP-seq KDM5A/JARID1A peaks with the previous study

Dataset S2

Venn diagrams show the overlap of our data with peaks reported in GSM480163 (mapped by authors using ELAND and QuEST) in *A* and with peaks that we remapped from GSM480163 data using BOWTIE and MACS in *B*. The gene list for the overlap generated from BOWTIE and MACS data are presented. The data from the previous study by Peng et al. (7) are highlighted in yellow.

Dataset S3. Classification of KDM5A peaks based on peak distance from TSS and expression of associated genes in *Kdm5a* knockout cells

Dataset S3

Distance of peaks was measured from peak summit to TSS and classified as follows: TSS peaks: if peak summit to TSS distance is within −200 bp to +1,100 bp; promoter: if peak summit is in region from −3,100 to −201 bp upstream of TSS; distal: if peak summit is in region beyond −3,101 bp upstream of TSS; gene body: if peak summit is located within +1101 to TTS+100 bp region; downstream: if distance of peak summit is more than TTS+100 bp from its closest gene.

Dataset S4. Wilcoxon rank sum test for gene expression changes in *Kdm5a* knockout cells during differentiation

Dataset S4

Comparison with the gene expression changes of non-KDM5A targets is shown for each set of KDM5A targets.

Dataset S5. χ^2 Test on overlap between KDM5A and E2F4: 2 × 2 contingency table

Dataset S5

Dataset S6. List of differentially bound KDM5A target genes in U937 cells

Dataset S6

Venn diagram shows the overlap between the KDM5A target genes within each group and the DREAM complex targets. The DREAM targets were determined as the genes bound by any three of the following four regulators: p130, E2F4, LIN9, and LIN52 (1).

1. Litovchick L, et al. (2007) Evolutionarily conserved multisubunit RBL2/p130 and E2F4 protein complex represses human cell cycle-dependent genes in quiescence. *Mol Cell* 26(4):539–551.

Dataset S7. Primers used in quantitative PCR

Dataset S7

Dataset S8. List of genes displaying enrichment of KDM5A in different conditions in U937 cells (to accompany Fig. S1)

Dataset S8