Supporting Information

Guo et al. 10.1073/pnas.1208807109

SI Materials and Methods

Cell Culture and Reagents. Human colorectal carcinoma cell line HCT116 (American Type Culture Collection) was cultured in McCoy's 5A medium (Gibco) with 10% (vol/vol) FBS (HyClone). Cells were grown up to 70% confluency in 100 × 20 mm cell-culture dishes (Corning) for genomic DNA, RNA, or nuclear protein extraction. HCT116-*MLL2*$^{-/-}$ cells were maintained in complete medium containing 0.5 mg/mL Geneticin. HEK 293FT cells (Invitrogen) were cultured in DMEM with high glucose (Gibco) containing 10% (vol/vol) FBS and used for rAAV targeting virus production as previously described (1). Exponentially growing cells were used for ChIP and microarray gene expression profiling. For retinoic acid treatment, cells in six-well plates were treated with 1 μM *all-trans* retinoic acid (Sigma) for 48 h before cells were harvested for RNA preparation. The anti-Flag monoclonal antibody was from Sigma (F3165); the anti–mixed-lineage leukemia (MLL)2 antibody was described previously (2).

Genetic Targeting. The rAAV vector used for the somatic knockin of an epitope tag into the carboxyl terminus of MLL2 was previously described (3, 4). The DNA fragment, which was constructed as previously described, was inserted into the position immediately upstream of the stop codon of the *MLL2* gene and contained a streptavidin binding peptide, 3× Flag-tag coding sequences, and a neomycin (neo) selection gene sequence (1, 3, 4). The homology arms were ∼1–1.2 kb in length and surrounded the insertion site. They were amplified from HCT116 genomic DNA using Hotstart High-Fidelity Platinum Taq polymerase (Invitrogen). Using Lipofectamine 2000 (Invitrogen), the vector was cotransfected with pAAV-RC and pHelper plasmids (Invitrogen) into HEK 293FT cells to generate rAAV targeting virus using a modified procedure based on the previously described protocol (1). Virus was harvested 2 d after transfection. HCT116 cells were infected with targeting virus for 5 h. Cells were allowed to recover in complete culture medium for 2 d, and were serially diluted into 96-well plates in culture media containing 0.5 mg/mL Geneticin. Geneticin-resistant cells were grown for 14–21 d, harvested for Lyse-*N*-Go genomic DNA extraction (Thermo Scientific), and screened by PCR using two sets of primers (see *Primers* for primer information). The first set of primers has the forward primer, P1, located in the genome, upstream of the 3× Flag-tag insertion, and the reverse primer, NR, is located in the 3× Flag-tag insertion region. The second set of primers has the forward primer, NF, located in the insertion region, and reverse primer, P2, located downstream of the insertion in the genome. The screening PCR products that identified positive clones were confirmed by sequencing (Genewiz). Positive clones were then infected with Cre recombinase-expressing adenovirus (Vector Biolabs) to remove the drug-resistance selection marker. Cells were serially diluted and cultured in complete media for 10–14 d, and clones were picked for PCR screening using primers P3 and P4 (see *Primers* for primer information). MLL2-Flag expression was confirmed by sequencing the fusion transcript (RT-PCR) and by anti-Flag immunoblot analysis of cell nuclear extract.

To generate *MLL2*-knockout cell lines, the above strategy was modified such that the DNA fragment was inserted right before amino acid 5391Leu. After the generation of *MLL2*$^{+/-}$ cell lines, two independent *MLL2*$^{+/-}$ lines were used for targeting the second *MLL2* allele in exactly the same manner to generate *MLL2*$^{-/-}$ cell lines. The neomycin selection marker in the second *MLL2* allele was not removed. *MLL2* knockout (*MLL2*$^{-/-}$) was confirmed by genotyping, sequencing the mutant transcript (RT-PCR), and anti-MLL2 immunoblot.

Nuclear Extraction, Immunoprecipitation, and Western Blot. The protocol for nuclear extraction was previously described (2). Chemicals were ordered from Sigma unless otherwise stated. Briefly, parental HCT116 cells or HCT116 cells expressing Flag-MLL2 fusion protein were harvested, washed with buffer A [10 mM Hepes·KOH (EMD), pH 7.9, 1.5 mM magnesium chloride, 10 mM potassium chloride, 0.5 mM DTT, 5 mM sodium fluoride, 2 μM sodium orthovanadate (MP Biomedical), protease inhibitor mixture (Roche)], and lysed in buffer A with 0.1% Nonidet P-40. The nuclei were collected by centrifugation (4 °C, at 18,000 × *g* for 30 min) and lysed in high-salt buffer C [20 mM Hepes·KOH, pH 7.9, 1.5 mM magnesium chloride, 420 mM sodium chloride, 0.2 mM EDTA (Gibco), 25% (vol/vol) glycerol, 0.5 mM DTT, 5 mM sodium fluoride, 2 μM sodium orthovanadate, protease inhibitor mixture] at 4 °C. Nuclear lysate was diluted 1:2 (vol/vol) with low-salt buffer D [20 mM Hepes·KOH, pH 7.9, 1.5 mM magnesium chloride, 50 mM potassium chloride, 0.2 mM EDTA, 20% (vol/vol) glycerol, 0.5 mM DTT, 5 mM sodium fluoride, 2 μM sodium orthovanadate, protease inhibitor mixture] for direct immunoblot or immunoprecipitation.

One milligram of nuclear extract was diluted with low-salt buffer D in the volume ratio of 1:2 to obtain a final salt concentration of ∼150 mM and supplemented with 0.1% Nonidet P-40. Immunoprecipitation was done using anti-Flag magnetic M2-agarose beads (Sigma) by incubation overnight at 4 °C. Beads were washed with immunoprecipitation buffer 12 times at 4 °C and eluted with 0.25 mg/mL 3× Flag peptide (Sigma) for 2 h at 4 °C. Eluate was resolved on 4–12% SDS/PAGE. The blot was transferred using a TE22 tank transfer unit (Hoefer) for anti-Flag antibody immunoblot (F3165; Sigma).

Chromatin Immunoprecipitation. HCT116 and MLL2-Flag knockin cells were cultured using McCoy's 5A medium. Cells were harvested at 60–80% confluency and washed three times with PBS; 2×10^8 cells were used for each ChIP. Cells ($1–2 \times 10^8$) were cross-linked with 2 mM disuccinimidyl glutarate (Pierce) in PBS for 45 min and further cross-linked with 1% (vol/vol) formaldehyde (Sigma) for 8–10 min. Glycine (1.25 M) was used to stop the cross-linking. Cells were then lysed with Hepes cell-lysis buffer (Santa Cruz Biotechnology) for 15 min, centrifuged for 5 min at 3,000 rpm, and then further lysed with RIPA buffer (Santa Cruz Biotechnology) for 15 min. Afterward, cell lysates were subjected to sonication with a Branson sonifier (25% output, 50% duty time, 15 min per cycle for two cycles) and further sonicated with a Bioruptor (15 s on/15 s off, high power, 5 min per cycle for six cycles) (Diagenode). The size of the sheared chromatin was verified by agarose gel electrophoresis after reverse cross-linking. For immunoprecipitation, mouse anti-Flag antibody (Sigma) was conjugated to protein-G magnetic beads (New England BioLabs) overnight in PBS containing protease inhibitor. Conjugated beads were added to the cell lysates and incubated at 4 °C overnight. Afterward, beads were washed four times with LiCl washing buffer [100 mM Tris, pH 7.5, 500 mM LiCl, 1% (vol/vol) Nonidet P-40, 1% (wt/vol) sodium deoxycholate] and twice with TE buffer (10 mM Tris·HCl, pH 8.0, 1 mM EDTA). To verify the specificity of ChIP, Western blot was performed to detect MLL2-Flag fusion protein signal. Briefly, 5% of the precipitated beads was aliquoted before reverse cross-linking and mixed with 1× SDS sample buffer [5% (vol/vol) glycerol, 2.5%

(vol/vol) 2-mercaptoethanol, 1% (wt/vol) SDS, 0.1% (wt/vol) bromophenol blue, 30 mM Tris·HCl, pH 6.8] and boiled for 10 min. Samples were loaded onto SDS/PAGE for electrophoresis. Western blot was performed as described above. To recover DNA from the immunoprecipitation fraction for library preparation, reverse cross-linking was achieved by treating precipitated beads at 65 °C for 6 h with protease K and 0.2 M NaCl. DNA was purified with a MinElute PCR Cleanup Kit (Invitrogen).

ChIP for anti-H3K4me3 (Millipore; 07-473) and anti-H3K27me3 (Millipore; 07-449) was performed essentially following the same procedure as described above, except that cells were cross-linked only with formaldehyde and sonicated with a Branson sonifier (25% output, 50% duty time, for 15 min). The DNA recovered from ChIP was diluted 1:20 and used as quantitative (q)PCR template. ChIP versus input DNA enrichment was determined by qPCR and normalized by LINE-1 (see *Primers* for primer information).

Library Preparation for Sequencing and ChIP-Sequencing Data Analysis. Libraries were prepared using a NEBNext DNA Sample Preparation Kit (New England BioLabs). Briefly, DNA fragments from ChIP were blunt-ended with T4 DNA polymerase, Klenow polymerase, and T4 polynucleotide kinase; dA tailing was achieved by Klenow polymerase; Illumina library construction adaptors were ligated to the DNA fragments; PCR amplification (18 or 20 cycles) was used to amplify the libraries. Sequencing of the library was performed following Illumina's Solexa protocol using the Illumina GAII platform. The sequence reads (tags of 35 bases each) obtained were aligned with Eland software using the eland_extended module, which matches 32 bases and then explores the alignment of the remaining 3 bases of the 35 base tags. The following criteria were used to filter the tags for further analysis: Each tag was required (*i*) to pass the Illumina chastity filter and (*ii*) to match uniquely to the human genome (hg18) with no more than two mismatches. In total, a range of ~14–22 million qualified sequencing tags were obtained for the parental, MLL2-Flag1, and MLL2-Flag2 cell-derived ChIP-sequencing (seq) libraries. Each of these tags was uniquely mapped to a genomic location. The numbers of nonrepetitive tags (hereafter named as distinct tags) were normalized by the total tags obtained for each library in further analyses.

We modified a previously described differential tag-density analysis approach for ChIP-seq data analysis (5, 6). To identify MLL2-targeted regions, we applied a previously established, relatively stringent criterion by which at least fivefold enrichment (MLL2-Flag cells versus parental cells) was required to identify differential tag density (6). In this approach, we generated bins of 1,000 bp (1 kb) each across the human genome, assigned each distinct tag to a bin, and compared the distinct tag density for each bin between the parental and MLL2-Flag–expressing cell lines. In addition, to minimize low sequencing coverage artifacts, we analyzed only those bins that had a coverage of at least fivefold of the average tag number that would have been expected assuming an evenly random tag distribution for the library derived from the MLL2-Flag–expressing cells. By this criterion, a total of 2,060 MLL2 binding events was identified. A subset of identified binding events was confirmed by ChIP-qPCR in the originally sequenced libraries and in independently prepared ChIP libraries to validate the approach. To characterize the global distribution of these binding events, the relative position of each enriched region (bin) to a gene was determined. The list of genes and their coordinates is from the Ensembl ensGene database (http://hgdownload.cse.ucsc.edu/goldenPath/hg18/database/ensGene.txt.gz). Distal promoter and terminator regions of the genes were defined as those sequences that mapped 10 kb upstream or 10 kb downstream from transcription "start" and "end" sites, respectively. For any given gene with single or multiple transcripts, the minimal transcript start position and the maximal transcript end position were used to define a gene boundary. Statistical analysis for the concentration of binding events in the intragenic region was performed by χ^2 test.

To evaluate the specificity of ChIP-identified MLL2-bound loci (Fig. 2*A*), 48,003 qualified loci (bins) with at least 36 distinct tags in MLL2-Flag–expressing cell line 1 were surveyed. For each locus, we plotted the ratio of the normalized distinct tag number in the MLL2-Flag–expressing cells to the distinct tag number in the parental cells. The ratios of normalized distinct tag number (parental to MLL2-Flag1) for 18,331 qualified loci (those with at least 36 unique tags in the parental cell line) were included for plotting to confirm the lack of parental cell-specific enrichment.

ChIP-Quantitative PCR. To confirm the loci identified by ChIP-seq, primers were designed according to the enriched locus coordinates, and SYBR green-based real-time qPCR was performed with an Applied Biosystems 7900 platform (Life Technologies). A LINE-1 amplicon was used for normalization (refer to *Primers* for a list of primers used for ChIP-qPCR). ChIP and ChIP DNA library preparation was done as described above. Library DNA (2.5 μL) was used as qPCR template. Quantitative PCR was done in 20-μL reactions containing 1× PCR buffer (67 mM Tris·HCT, pH 8.8, 6.7 mM $MgCl_2$, 16.6 mM NH_4SO_4, 10 mM 2-mercaptoethanol), 2 mM dNTPs, 0.5 μM forward and reverse primers, 6% (vol/vol) DMSO, 1:2,000 SYBR green, and 1 unit Platinum Taq (Invitrogen). The PCR program was 95 °C, 10 min for hotstart, amplification for 38 cycles (95 °C, 15 s to 60 °C, 15 s to 72 °C, 15 s), and a dissociation curve analysis (95 °C, 15 s to 60 °C, 15 s to 95 °C, 15 s, ramp rate 2%). Three independent ChIP assays were done with the parental cell line and MLL2-Flag knockin cell line (cell line 1). Another MLL2-Flag knockin cell line (cell line 2) was also used for further confirmation.

Microarray Assay. Total RNA was prepared from $MLL2^{-/-}$ cells (two pairs of cell lines, clone 1/2 and cell line clone 3/4, which were derived from two independent $MLL2^{+/-}$ progenitor lines, were included for two separate microarray experiments) and HCT116 parental cells. The quality of the RNA was assessed with an 2100 Bioanalyzer G2939A (Agilent Technologies) and Nanodrop 8000 spectrophotometer (Thermo Scientific). Hybridization targets were prepared with a MessageAmp Premier RNA Amplification Kit (Applied Biosystems/Ambion) from total RNA, hybridized to GeneChip U133 Plus 2.0 arrays in Affymetrix GeneChip Hybridization Oven 645, washed in Affymetrix GeneChip Fluidics Station 450, and scanned with Affymetrix GeneChip Scanner 7G according to standard Affymetrix GeneChip hybridization, wash, and stain protocols. (Affymetrix). Each cell line was done in triplicate. Statistical analysis for false discovery rate was performed by significance analysis of microarrays (7). Statistical analysis for the enrichment of down-regulated genes in the ChIP-identified genes was performed by χ^2 test.

Confirmation of Microarray Gene Expression. The expression change of selected genes was confirmed by RT-qPCR in the original microarray-analyzed pair of $MLL2^{-/-}$ cell lines (clones 1 and 2) and in additional $MLL2^{-/-}$ clones (clones 3 and 4) derived from an independent $MLL2^{+/-}$ progenitor line. Total RNA was prepared from parental cell lines and $MLL2^{-/-}$ cells using a total RNA kit (Omega Bio-Tek) according to the manufacturer's instructions. One microgram of total RNA was used for reverse transcription using a SuperScript II First-Strand Synthesis Kit (Life Technologies) according to the kit manual. After transcription, the cDNA was diluted 1:20 and used as template for qPCR. Quantitative PCR was done as described for ChIP-qPCR. Gene expression levels were normalized to the housekeeper gene GAPDH (see *Primers* for list of primers used for ChIP-qPCR).

Ingenuity Pathway Analysis. The web-based Ingenuity Pathway Analysis (IPA) (www.ingenuity.com) tool was used to analyze the

genes down-regulated by >50% in the microarray and the putative target genes identified by ChIP. Top affected canonical pathways were identified for the down-regulated genes (Table 1). Putative MLL2 target genes were submitted to IPA for transcription factor analysis (Dataset S3). Statistical analysis for transcription factor and pathway enrichment was performed with software provided by IPA.

Retinoic Acid Treatment. HCT116 parental cells or *MLL2*$^{-/-}$ cells were treated with retinoic acid (1 μM) for 48 h. Afterward, RNA preparation, reverse transcription to generate cDNA, and real-time qPCR were performed as described above to measure the expression of ASB2 as a readout for retinoic acid treatment.

Primers

Name	Direction	Sequence (5′-3′)
MLL2-Flag somatic knockin genomic screening		
P1	Forward	AGTTGCTTCCTCACCAGAGC
NR	Reverse	AGTCGAGCTCACCCTGAAAA
NF	Forward	AGGGGATCGGCAATAAAAAG
P2	Reverse	CAGCAGTTTGGTGACTGAGG
P3a	Forward	TTGAGGACGATCAGCACAAG
P3b	Forward	CTGCAAGCTGGGCGATTAT
P4	Reverse	CAGGGAAGAGGTTGTGGGTA
MLL2-knockout genomic DNA screening		
P1	Forward	TCTGTGTACCAGGGACCTTAGTC
NR	Reverse	GGTCTTCTCGTCCATACTAGTGC
NF	Forward	AGGGGATCGGCAATAAAAAG
P2	Reverse	GCTACCTCTCTTCCCCCTCA
P3	Forward	CTTGGTCTGAGGGAGAGCTG
P4	Reverse	TGATGGTGCCAATGTACTCG
RT-PCR primers		
MLL2-Flag knockin	Forward	GCCCTGCCAGGTACATTAAC
	Reverse	AGTCGAGCTCACCCTGAAAA
MLL2 knockout	Forward	CACTACAAACGGCCCCATAC
	Reverse	TTCTAGGTCCTTGGCTGCAT

ChIP-qPCR primers for confirmation of ChIP peaks

Name	Putative target gene	Direction	Sequence
ChIP-qPCR primers for confirmation of ChIP peaks			
Chr1:151805001–151806000	S100A2	Forward	CCCCTCTACCTCAGCCCTAA
		Reverse	GGTCTCTGTGCCAGTGCTTT
Chr8:145086000–145087000	S100A2	Forward	GCACTTCACTGCACACCAAA
		Reverse	GAGGGAGCTCCAAAGAGAGG
Chr11:118288001–118289000	BCL9L	Forward	AGTCTCCGAAGACACCTGGA
		Reverse	GGTACTAGCAAGCCCCCTTC
Chr1:153429001–153430000	MUC1/MIR92b	Forward	AGCCTCAGGGCAACAACTAA
		Reverse	GGAAGTGGAGTGGGAGACCT
Chr14:93495001–93496000	ASB2	Forward	GAATGTGTCACCGTCACCTG
		Reverse	CGAGAGGCTAGGAGCACTGT
Chr16:88511001–88512000	TUBB3	Forward	GCAGCTGTAAATGGGACGAG
		Reverse	GCATCTTCCTGGGTTCACTC
Chr12:88270001–88271000	DUSP6	Forward	GCTGGAACAGGTTGTGTTGA
		Reverse	TGTTGCAGCTTGTTTGCACT
Chr1:151775000–151776000	S100A6	Forward	TGACGAGTCTCCCCTCATCT
		Reverse	GAGGGTCACATTCACGGTCT
Chr1:151788000–151789000	S100A3	Forward	TTCCTGGAATTGCTCACCTC
		Reverse	ACTGGGCTAGCATGGAAAAA
Chr1:151784000–151785000	S100A4	Forward	CATTCTTTTCCCTCCCAGAA
		Reverse	CTGCAGCTTCTCTTCCAACC
Chr1:151848000–151849000	S100A16	Forward	CTGAGCAAGCCACAGTCAGA
		Reverse	GGAGTAGGAATGGCCACAAG
Chr1:151847000–151848000	S100A16	Forward	CCCCTTATGGGAAAGGACAA
		Reverse	GTTCTTGGGTCCTGCTGTGT
Internal control	LINE-1	Forward	AAGGCCGCTCAACTACATGG
		Reverse	TGCTTTTGAATGCGTCCCAGAG

RT-qPCR for detecting gene expression levels

Target gene	Direction	Sequence (5′-3′)
S100A2	Forward	AGCTTTGTGGGGGAGAAAGT
	Reverse	CAGTGATGAGTGCCAGGAAA
S100A3	Forward	TGTGGGGACAAATACAAGCTC
	Reverse	AGTCCACCTCGCAGTCCTT
S100A4	Forward	CAAGTACTCGGGCAAAGAGG
	Reverse	GCTGTCCAAGTTGCTCATCA
S100A5	Forward	AGAGCTGTGTCTTGGGGAGA
	Reverse	CACAGCATGGTCAGGAACAC
S100A6	Forward	GAAGGAGCTGAAGGAGCTGA
	Reverse	TGGAAGTTCACCTCCTGGTC
S100A14	Forward	CTGACCCCTTCTGAGCTACG
	Reverse	TCCAATCAGCTCCCAGAAAC
S100A16	Forward	GCGAGATGCTCCAGAAAGAG
	Reverse	CCGCCTATCAAGGTCCAGTA
ASB2	Forward	CCGGACATCTCCAACAAATC
	Reverse	GTTGGTGTCTGCATTGTGCT
GAPDH	Forward	AGCCACATCGCTCAGACA
	Reverse	GAGGCATTGCTGATGATCTTG

1. Rago C, Vogelstein B, Bunz F (2007) Genetic knockouts and knockins in human somatic cells. *Nat Protoc* 2(11):2734–2746.
2. Cho YW, et al. (2007) PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *J Biol Chem* 282(28):20395–20406.
3. Zhang X, et al. (2008) Epitope tagging of endogenous proteins for genome-wide ChIP-chip studies. *Nat Methods* 5(2):163–165.
4. Wang Z (2009) Epitope tagging of endogenous proteins for genome-wide chromatin immunoprecipitation analysis. *Methods Mol Biol* 567:87–98.
5. Robertson G, et al. (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* 4(8):651–657.
6. Johnson DS, Mortazavi A, Myers RM, Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316(5830):1497–1502.
7. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 98(9):5116–5121.

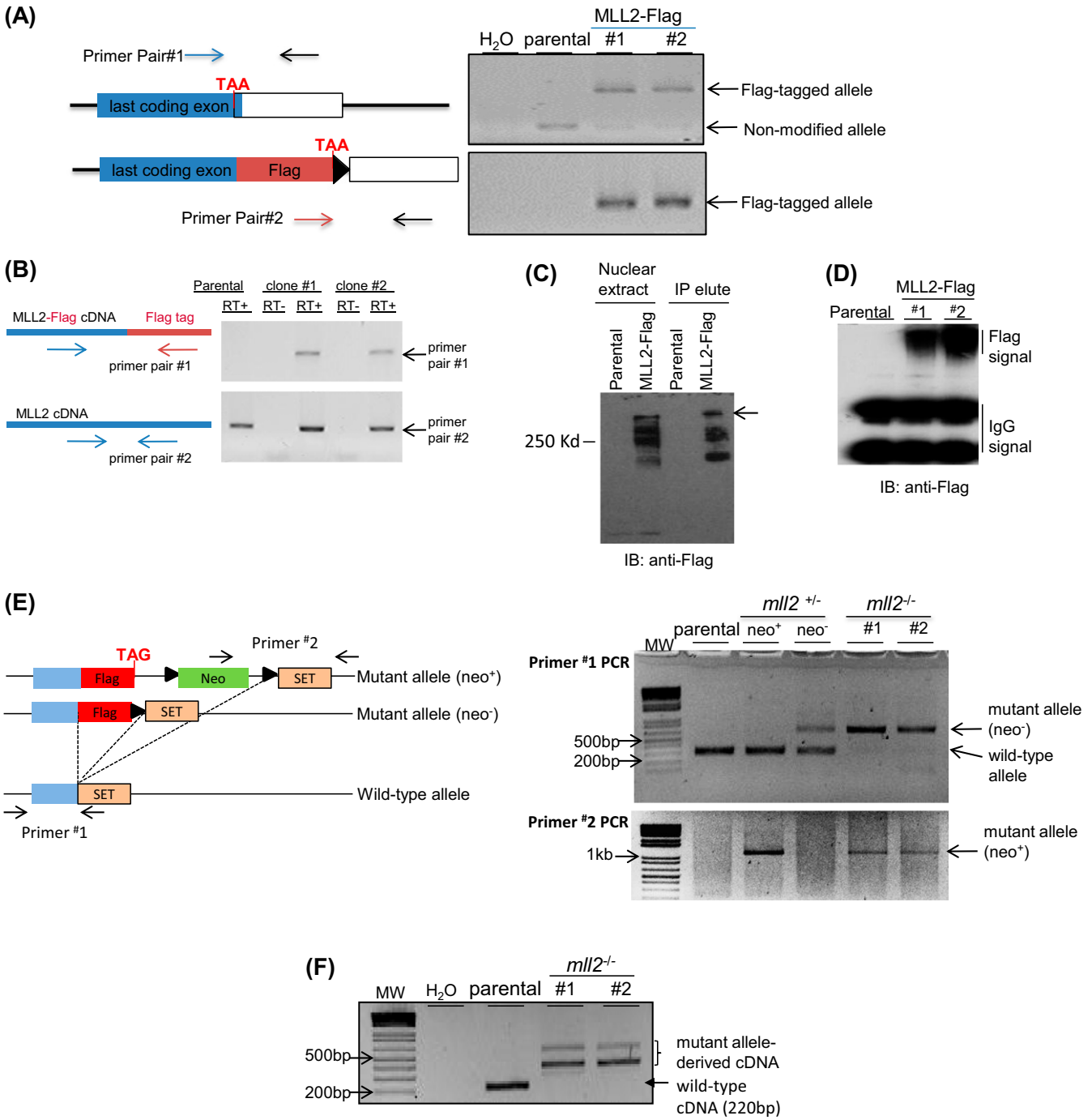

Fig. S1. Somatic targeting of MLL2. (*A*) Genotype confirmation of the endogenous *MLL2-Flag* allele. Primer pairs used for PCR, wild-type allele, and targeted allele are depicted. (*B*) RT-PCR was performed to detect both the *MLL2-Flag* transcript and the nonmodified *MLL2* transcript in the recombinant clones (1 and 2); in the parental cells, only the latter was detectable. (*C*) Anti-Flag antibody pulled down the MLL2-Flag polypeptide from the nuclear extract. IB, immunoblot. (*D*) Anti-Flag antibody pulled down the MLL2-Flag–associated complex from cross-linked chromatin. (*E*) Genotype confirmation of the *MLL2* gene knockout by insertion (note the larger neo$^+$ mutant allele was not amplified well by primer 1 when a smaller amplicon was present). (*F*) RT-PCR confirms the absence of the wild-type transcript in *MLL2*$^{-/-}$ cells. The longer transcripts detected in the knockout cells are mutant allele-derived transcripts, as expected, or with out-of-frame alternative splicing.

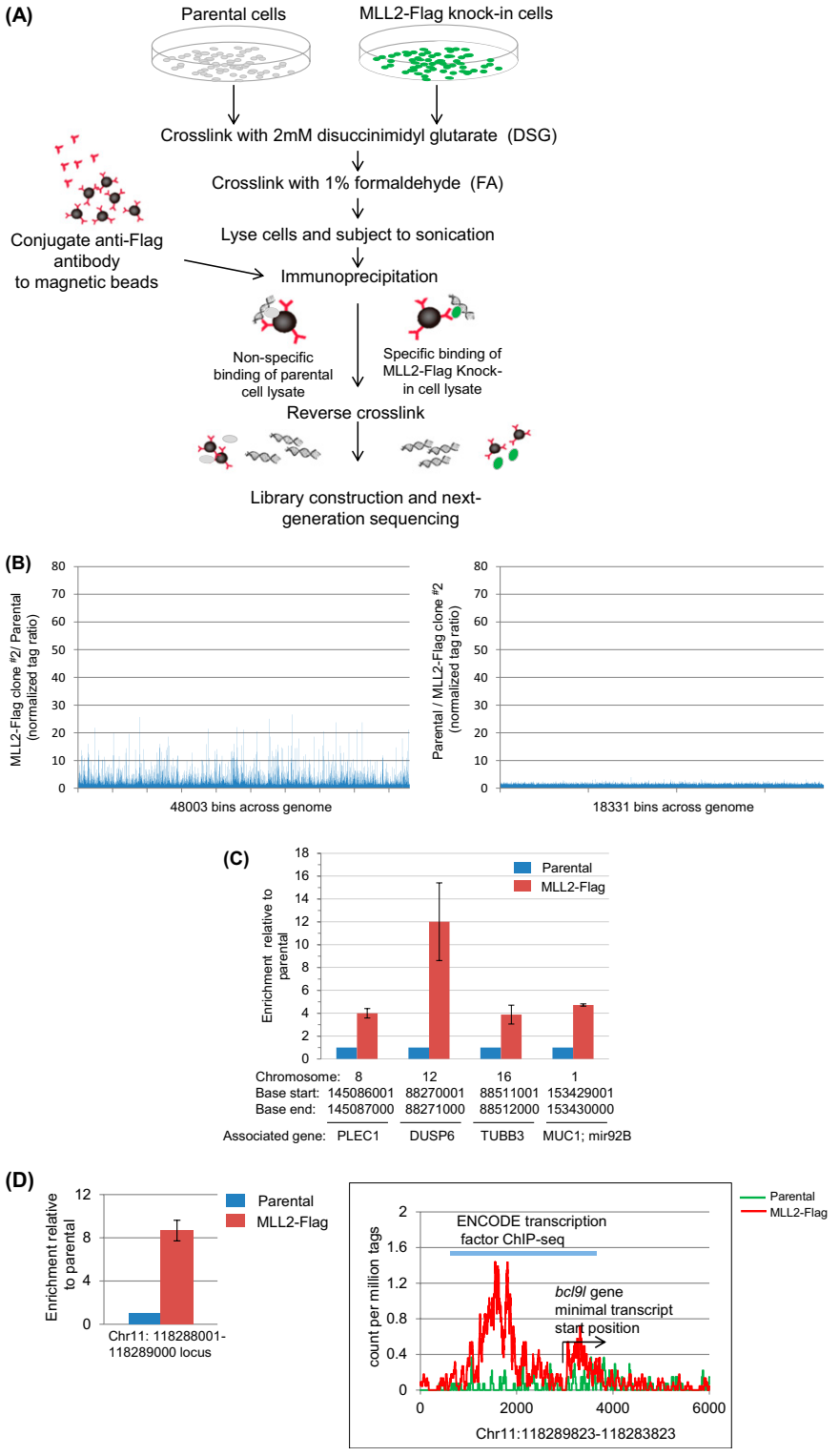

Fig. S2. Anti-Flag ChIP-seq for identification of MLL2 complex binding loci. (*A*) Schematic of the ChIP procedure. DNA derived from ChIP was used for library preparation for high-throughput sequencing with the Illumina GAII system. (*B*) Differential anti-Flag ChIP enrichment of MLL2-Flag cell line 2 confirmed the enriched loci identified by MLL2-Flag line 1. (*Left*) Analysis of the same 48,003 bins identified as enriched loci by MLL2-Flag line 1 as shown in Fig. 1*C*. For each bin, the ratio of the normalized distinct tag number in MLL2-Flag cell line 2 to the distinct tag number in the parental cells was plotted. (*Right*) Plot of the converse ratio of normalized distinct tag number (parental to MLL2-Flag2) of 18,331 bins with at least 36 unique tags in the parental cell line. (*C*) Confirmation of ChIP-seq–identified loci by ChIP coupled with quantitative PCR. (*D*) An MLL2-Flag–enriched locus. (*Left*) ChIP coupled with qPCR to confirm the locus identified by ChIP-seq. Results are from three independent ChIP experiments. (*Right*) The distinct coverage of each base in a 6-kb region is plotted. The minimal transcript start site of the gene that is most closely associated with the enriched locus and the corresponding transcription factor binding site identified by the ENCODE Project (1) are indicated.

1. Myers RM, et al.; ENCODE Project Consortium (2011) A user's guide to the Encyclopedia of DNA Elements (ENCODE). *PLoS Biol* 9(4):e1001046.

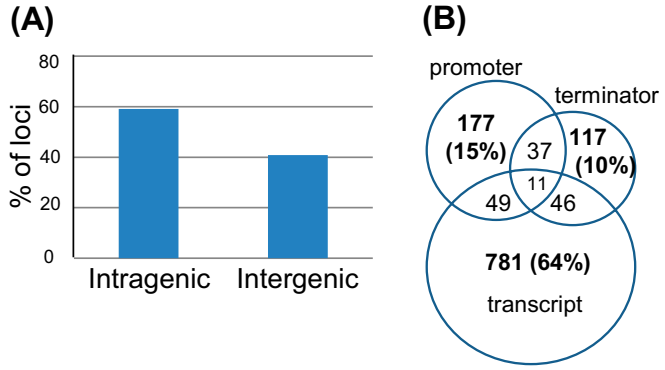

Fig. S3. Global distribution of MLL2 binding loci. (*A*) Enrichment of loci in intragenic regions ($P < 0.0001$). Among the 2,060 loci that were analyzed, 1,218 (59%) were in the intragenic region [defined as from the minimal transcript start to the maximal transcript end plus the upstream and downstream distal distance (±5 kb) boundary extended from both ends]; the remaining fraction (842; 41%) was defined as intergenic. The intragenic region encompassed the boundary of a gene, defined by the minimal transcript start to the maximal transcript end, plus the 5-kb upstream (promoter) region and 5-kb downstream (terminator) region. (*B*) The distribution of 1,218 intragenic loci within a gene, a promoter region, or a terminator region, as defined in *A*.

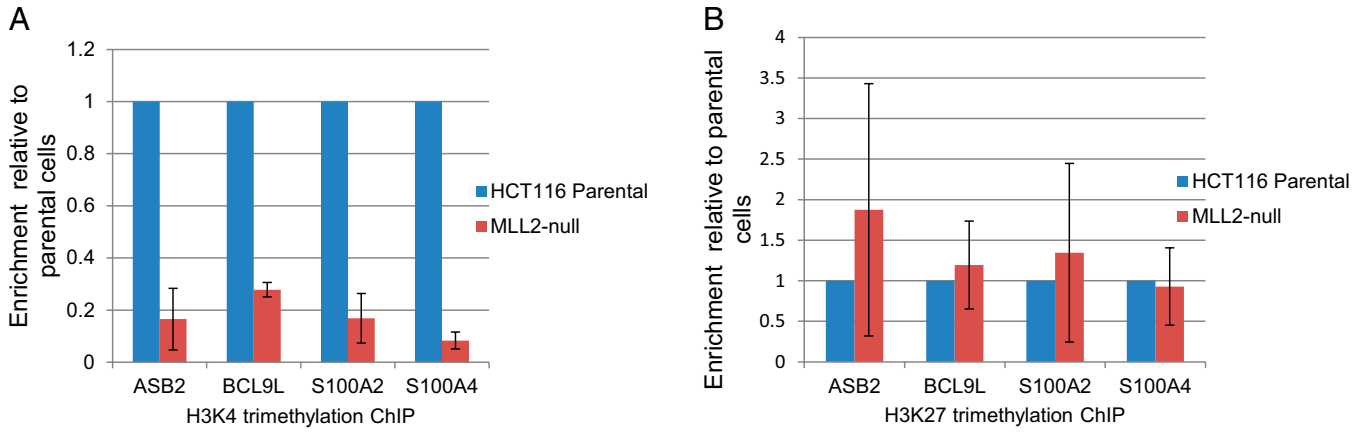

Fig. S4. Reduced H3K4 trimethylation in MLL2 binding loci in *MLL2*$^{-/-}$ cells. ChIP coupled with quantitative PCR revealed reduced H3K4 trimethylation (*A*) and minor or no change in H3K27 trimethylation (*B*) in selected loci bound by MLL2 in *MLL2*$^{-/-}$ cells. Four genes were included for analysis. Results were from three independent ChIP experiments in an *MLL2*$^{-/-}$ HCT116 clone.

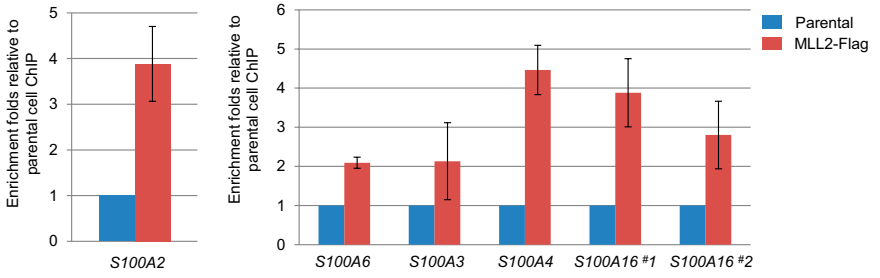

Fig. S5. MLL2 binding loci identified by ChIP-seq. ChIP coupled with quantitative PCR confirmed MLL2 binding events in the *S100A* gene cluster. For simplicity, each binding site is identified by the gene with which it is most closely associated. Results for the *S100A2* locus are from three independent ChIP experiments. Results for other loci are from three independent quantitative PCR analyses of an independent ChIP experiment. The binding event associated with *S100A16* was confirmed by amplifications using two primer pairs.

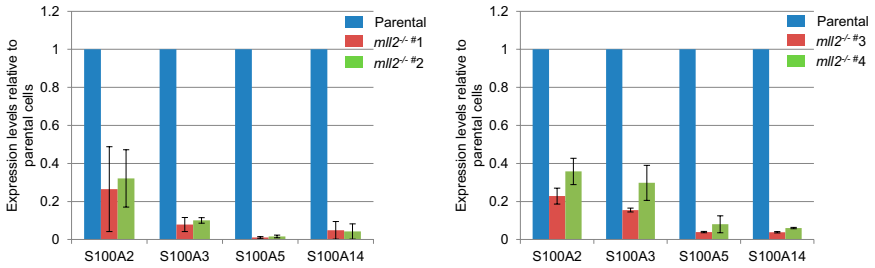

Fig. S6. Down-regulation of *S100A* cluster gene expression. Reverse transcription coupled with quantitative PCR confirmed reduced gene expression in *MLL2*$^{-/-}$ cells. *S100A2*, *S100A3*, *S100A5*, and *S100A14* are shown as examples to confirm the microarray findings. The reduced gene expression was further verified in two additional *MLL2*$^{-/-}$ lines that were derived from an independent *MLL2*$^{+/-}$ progenitor line.

Other Supporting Information Files

Dataset S1 (XLS)
Dataset S2 (XLS)
Dataset S3 (XLS)