Acute myeloid leukemia requires Hhex to enable PRC2-mediated epigenetic repression of Cdkn2a

Benjamin J. Shields[1,3], Jacob T. Jackson[1], Donald Metcalf[1,3], Wei Shi[2,4], Qiutong Huang[1], Alexandra. L Garnham[2], Stefan P. Glaser[1,3], Dominik Beck[6], John E. Pimanda[6], Clifford W. Bogue[7], Gordon K. Smyth[2,5], Warren S. Alexander[1,3] and Matthew P. McCormack[1,3]

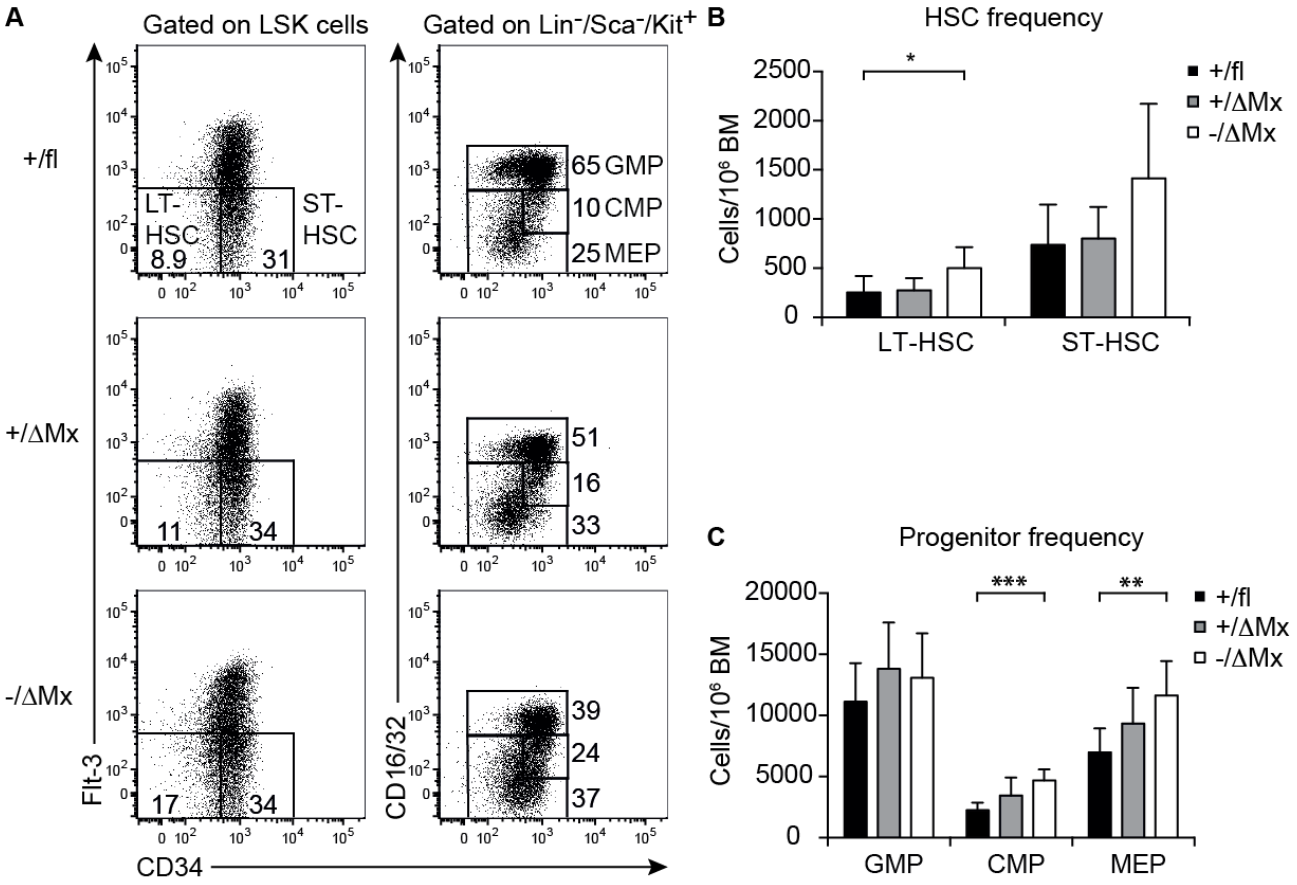

Figure S1. Related to Figure 2. Hhex is dispensable for the development of myeloid progenitors. A. Cellular phenotyping of steady-state Hhex$^{+/fl}$, Hhex$^{+/\Delta Mx}$ and Hhex$^{-/\Delta Mx}$ BM by flow cytometry. **B.** The frequency of Long-term Hematopoietic Stem Cells (LT-HSCs, Lin$^-$/Kit$^+$/Sca-1$^+$/Flt-3$^-$/CD34$^-$) and Short-Term (ST-HSCs, Lin$^-$/Kit$^+$/Sca-1$^+$/Flt-3$^-$/CD34$^+$) and **C.** Granulocyte-Macrophage Progenitors (GMP, Lin$^-$/Kit$^+$/CD34^{hi}/CD16/32^{hi}), Common Myeloid Progenitors (CMP, Lin$^-$/Kit$^+$/CD34^{hi}/CD16/32^{Int}) and Megakaryocyte-Erythroid Progenitors (MEP, Lin$^-$/Kit$^+$/CD34$^-$/CD16/32$^-$) cell subsets are shown (n=6-7 mice/group). *P<0.05, **P<0.01, ***P<0.001, Student's t-test.

Table S1. Colony formation by Hhex-deleted BM cells

Stimulus	Hhex Genotype	Colonies per 25,000 BM cells						
		Blast	G	GM	M	Eo	Meg	TOTAL
GM-CSF	+/fl	-	23 ± 12	5 ± 3	43 ± 21	4 ± 2	-	75 ± 30
	+/ΔMx	-	22 ± 10	4 ± 3	38 ± 2	3 ± 2	-	67 ± 11
	-/ΔMx	-	43 ± 12**	3 ± 2	6 ± 3***	5 ± 3	-	57 ± 16
G-CSF	+/fl	-	14 ± 8	-	-	-	-	14 ± 8
	+/ΔMx	-	11 ± 3	-	-	-	-	11 ± 3
	-/ΔMx	-	17 ± 7	-	-	-	-	17 ± 7
M-CSF	+/fl	-	6 ± 5	7 ± 5	48 ± 22	-	-	61 ± 31
	+/ΔMx	-	5 ± 3	4 ± 2	58 ± 23	-	-	67 ± 26
	-/ΔMx	-	2 ± 2*	2 ± 2*	18 ± 6**	-	-	22 ± 9**
IL-3	+/fl	5 ± 3	28 ± 10	9 ± 6	17 ± 9	3 ± 3	5 ± 3	68 ± 27
	+/ΔMx	5 ± 4	33 ± 9	7 ± 3	20 ± 7	3 ± 3	7 ± 1	75 ± 17
	-/ΔMx	8 ± 4	49 ± 19*	15 ± 7	15 ± 5	3 ± 2	7 ± 5	98 ± 31
IL3,SCF, EPO	+/fl	6 ± 3	37 ± 12	15 ± 4	18 ± 7	4 ± 3	20 ± 9	100 ± 25
	+/ΔMx	8 ± 6	37 ± 9	15 ± 3	26 ± 6	2 ± 2	22 ± 10	110 ± 19
	-/ΔMx	6 ± 4	56 ± 12**	18 ± 5	28 ± 10*	3 ± 2	24 ± 11	135 ± 27*

Data are combined from triplicate cultures of n=6 (+/fl), 3 (+/ΔMx) and 9 (-/ΔMx) mice per group.
*p<0.05 **p<0.01 ***p<0.001 vs +/fl (Student's T-test). Significant differences are shaded.

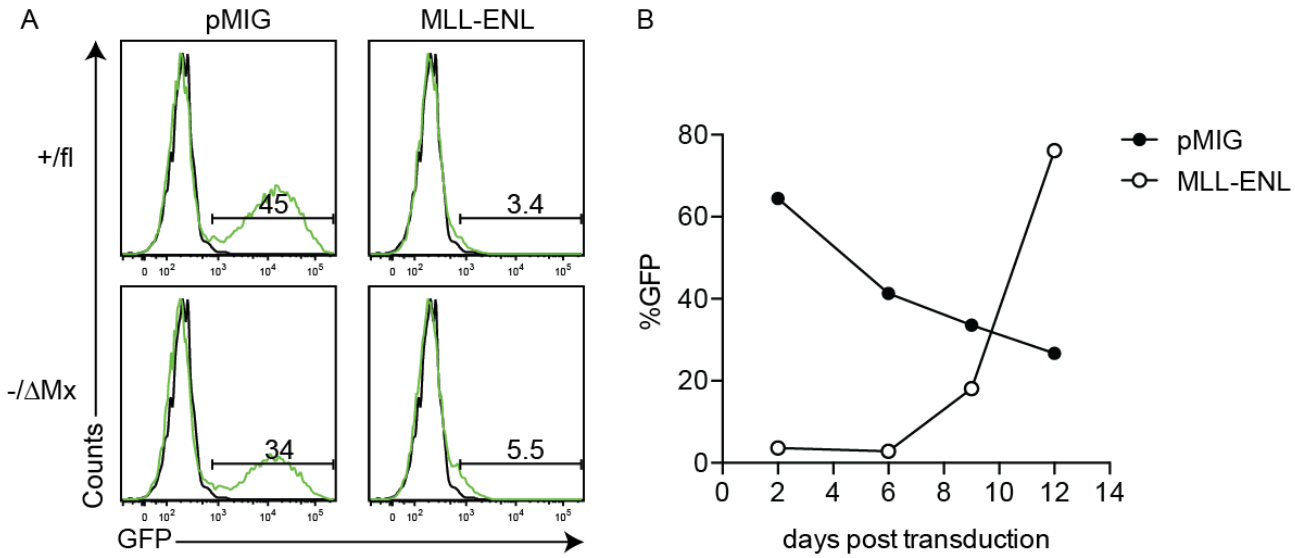

Figure S2. Related to Figure 2. Efficient transduction of bone marrow progenitors with MSCV-GFP-MLL-ENL is Hhex independent. A. GFP expression in control (pMIG) and MLL-ENL transduced lineage-depleted BM progenitors of the indicated Hhex genotypes, 2 days post-infection. The black line represents untransduced cells and green line transduced cells. **B**. To track the selection for MLL-ENL expressing cells, control and MLL-ENL transduced $Hhex^{+/fl}$ cells were maintained in liquid culture and GFP expression determined by flow cytometry.

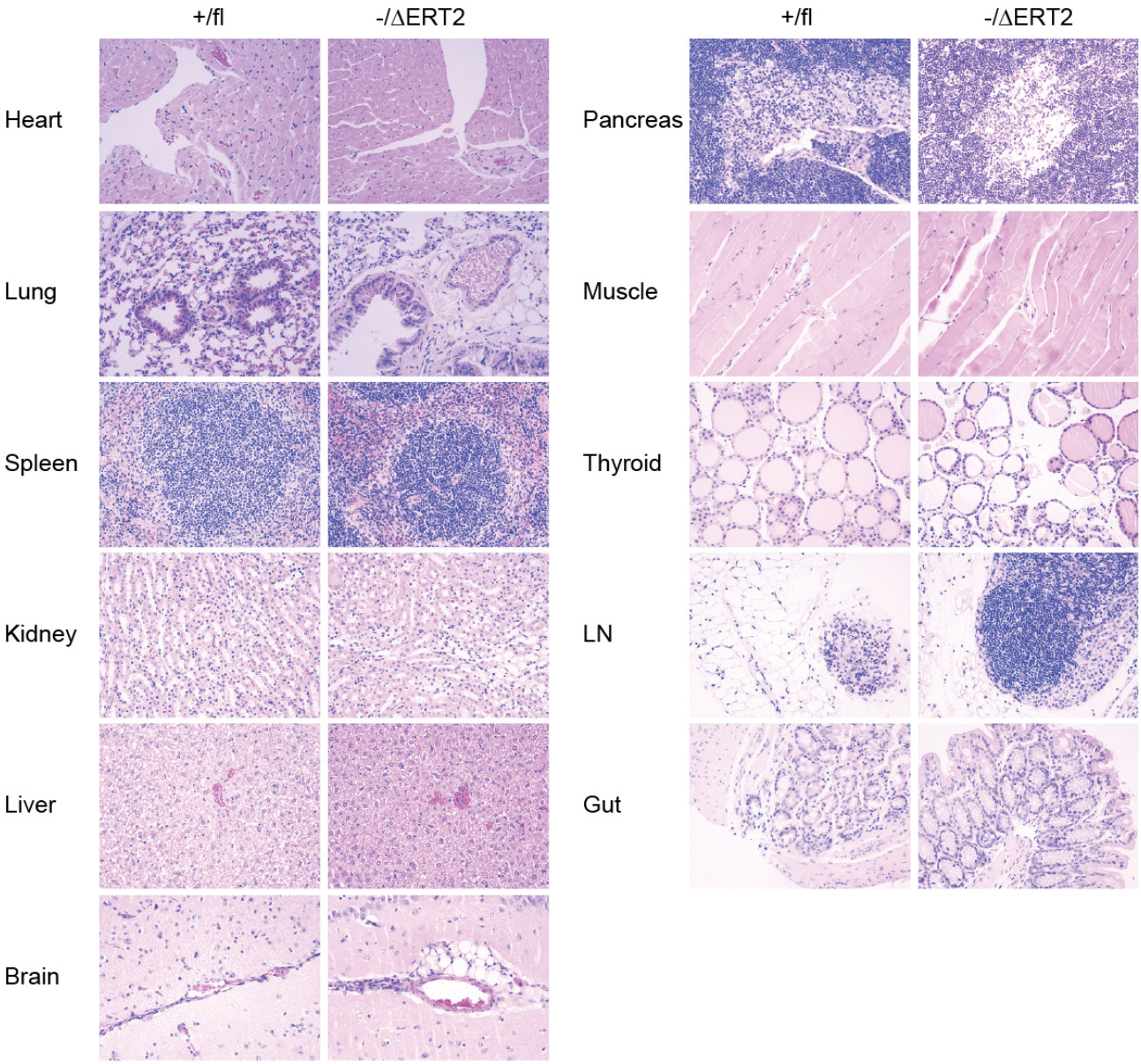

Figure S3 Related to Figure 3. Systemic loss of Hhex does not induce adverse pathology in tamoxifen treated Cre^{ERT2};Hhex$^{-/fl}$ mice. Tissues were harvested from mice of the indicated genotypes then fixed, sectioned and stained with haematoxylin and eosin. LN=Lymph Node.

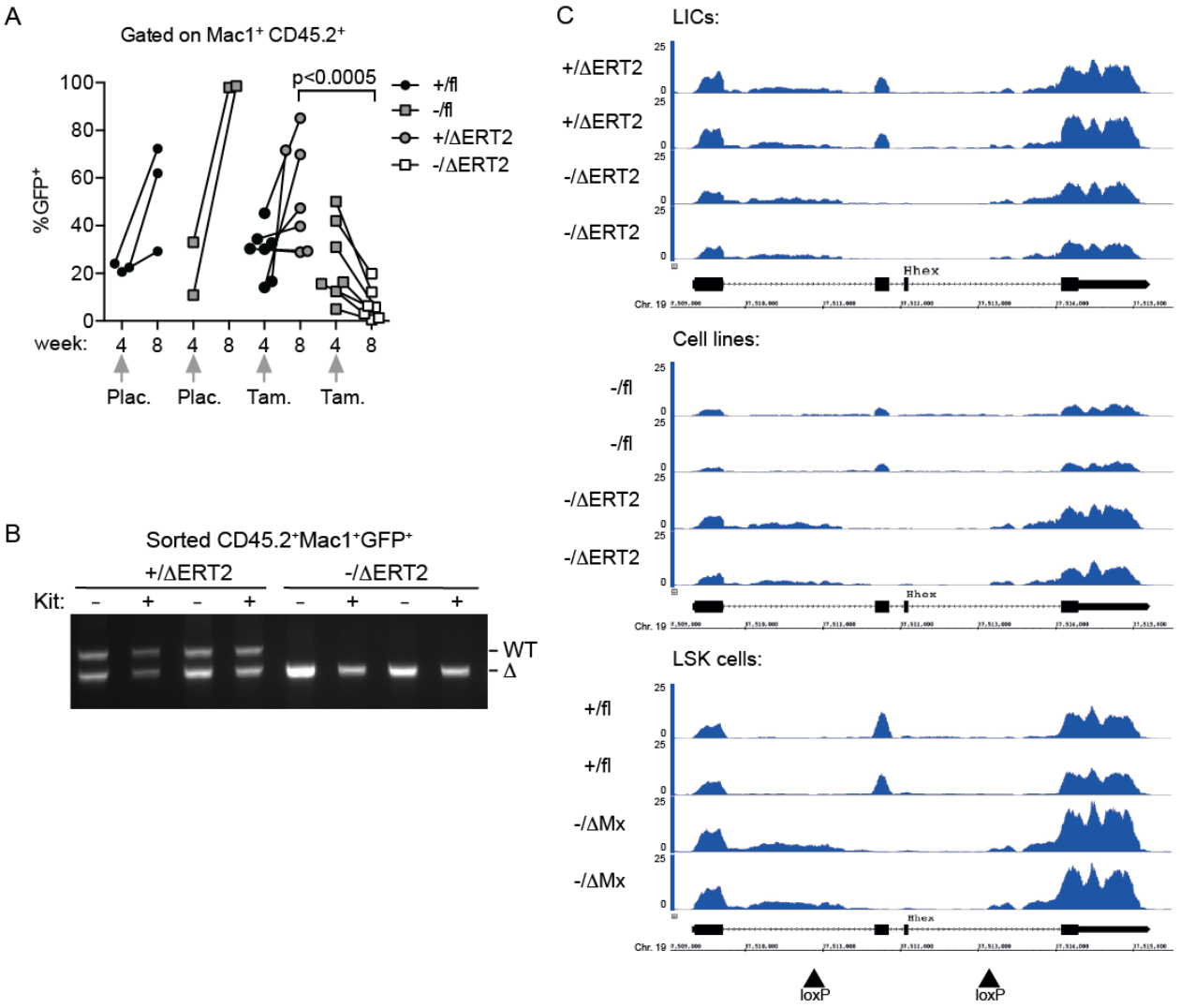

Figure S4. Related to Figure 5. Generation of Hhex-deleted MLL-ENL LICs for RNA sequencing. **A.** Peripheral blood analysis of Ly5.1 recipient mice injected with MLL-ENL transduced lineage-depleted Cre^{ERT2};$Hhex^{+/fl}$ and Cre^{ERT2};$Hhex^{-/fl}$ BM. Four weeks post injection, mice were administered placebo (Plac.) or tamoxifen (Tam.). Lines connect sequential samples taken from individual mice. *P* value was calculated using Student's T test. **B.** Hhex deletion PCR analysis of sorted Kit^- and Kit^+ virally transduced donor myeloid cells at 4-weeks post tamoxifen administration. Kit^+ cells are LICs used for RNA sequencing. **C**. Read depth analysis is shown for RNA-sequencing data obtained from control (top two panels) and Hhex-deleted (bottom two panels) LICs, MLL-ENL cell lines and LSK cells. This shows a complete loss of reads at exons 2-3 in Hhex-deleted cells. Arrowheads at the bottom show the positions of loxP sites in the $Hhex^{fl}$ allele. Units are reads per million mapped reads (RPM). Chr., chromosome.

Table S2: Excel file: RNA sequencing - differentially expressed genes in LICs

Table S3: Excel file: RNA sequencing - differentially expressed genes in MLL-ENL cell lines

Table S4. Gene Set Enrichment Analysis of Hhex-deleted cell lines

MSigDB Gene Set	NES	*p* value	FDR *q* value
Downregulated			
SCHUHMACHER_MYC_TARGETS_UP	1.69	0.002	0.151
WONG_EMBRYONIC_STEM_CELL_CORE	1.66	0.000	0.170
DANG_MYC_TARGETS_UP	1.60	0.000	0.232
DANG_REGULATED_BY_MYC_UP	1.53	0.015	0.275
Upregulated			
ONGUSAHA_TP53_TARGETS	-2.09	0.000	0.001
BROWN_MYELOID_CELL_DEVELOPMENT_UP	-1.79	0.000	0.129

GSEA was used to compare the indicated MSigDB gene sets to normalized RNA expression data from Hhex-knockout versus Hhex-heterozygous MLL-ENL AML cell lines. NES, normalized enrichment score. FDR, false discovery rate.

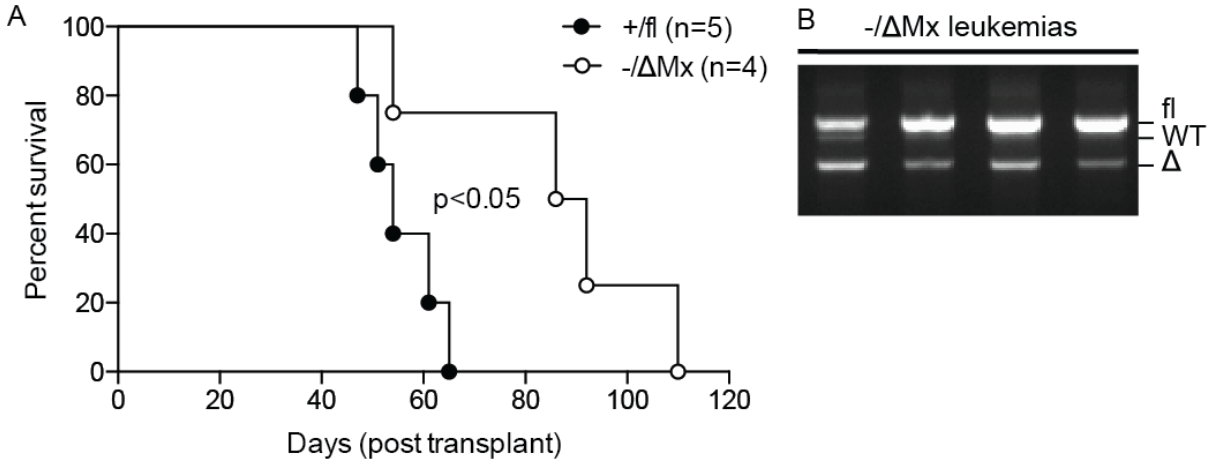

Figure S5 Related to Figure 5. Hhex is required for HoxA9;Meis1 driven leukemia A. Kaplan-Meier survival curve of recipients of MCSV-HoxA9;Meis1 transduced BM of the indicated Hhex genotypes. **B**. $Hhex^{fl}$ is selected for in $Hhex^{-/\Delta ERT2}$ HoxA9;Meis1 primary leukemias. Genomic DNA was extracted from whole BM of leukemic mice (>95% $CD45.2^+$, $Mac1^+$ cells) of the indicated Hhex genotypes and analyzed by PCR to reveal floxed (fl), wild-type (WT) and null (Δ) Hhex alleles.

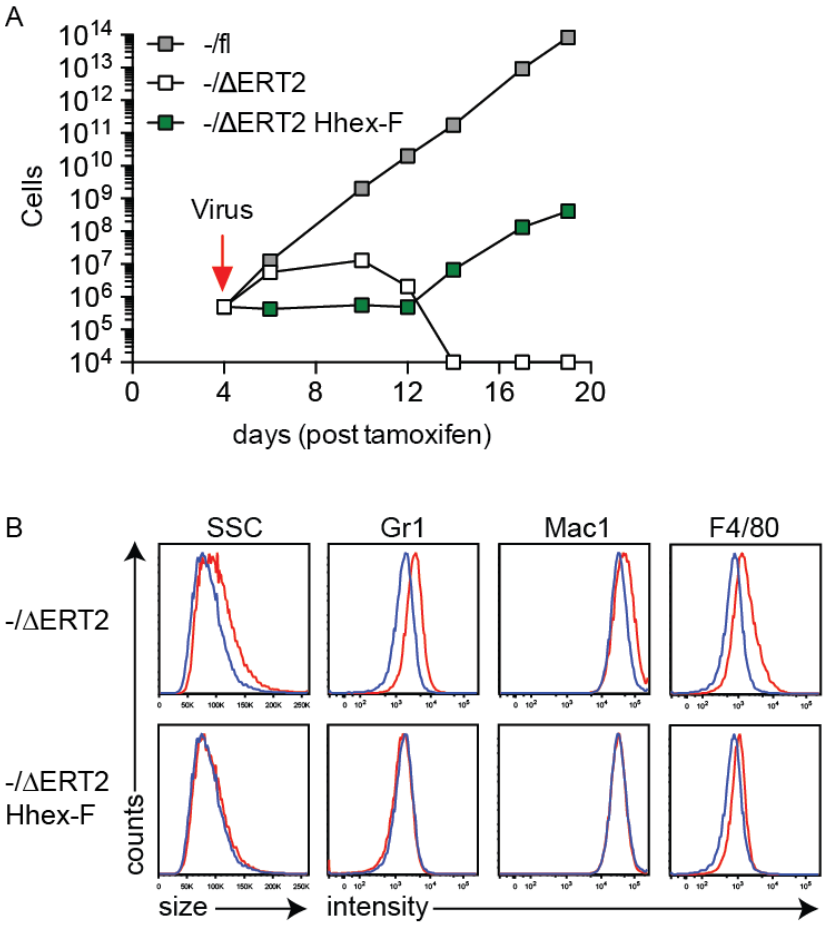

Figure S6. Related to Figure 5. Retroviral re-expression of Flag tagged Hhex rescues growth and differentiation block in Hhex$^{-/\Delta ERT2}$ MLL-ENL cells. A. Growth rates of parental and Hhex-F rescued cell lines. Cre^{ERT2};Hhex$^{-/fl}$ MLL-ENL cell lines were either left untreated (-/fl) or treated with tamoxifen for 4 days (-/ΔERT2), then infected with MSCV-GFP-Hhex-F retroviruses. **B**. Analysis of Hhex$^{-/fl}$,Hhex$^{-/\Delta ERT2}$ and stable Hhex-F rescued cell lines for granularity (side scatter, SSC) and differentiation markers by flow cytometry (Blue line is untreated parental Hhex$^{-/fl}$, Red line is Hhex$^{-/\Delta ERT2}$ or Hhex-F rescued cells).

Table S5. Oligonucleotides sequences for CRISPR-Cas9-mediated gene disruption					
Gene	**Target**	**Reverse complement**	**Exon**	**Quality score***	**# off-target sites (in genes) [mismatches]****
mTp53					
C1	**CACCG**AGGAGCTC CTGACACTCGGA	**AAAC**TCCGAGTGT CAGGAGCTCCT**C**	3	79	187 (18) [2]
C2	**CACCG**GACACTCG GAGGGCTTCACT	**AAAC**AGTGAAGCC CTCCGAGTGTC**C**	3	76	101 (10) [2]
p16					
C1	**CACCG**CGGGGCGT TGGGCGAAACCC	**AAAC**GGGTTTCGC CCAACGCCCCG**C**	1	95	26 (11) [3]
C2	**CACCG**GGTACGAC CGAAAGAGTTCG	**AAAC**CGAACTCTT TCGGTCGTACC**C**	1	96	12 (4) [3]
p19					
C1	**CACCG**TGGTGAAG TTCGTGCGATCC	**AAAC**GGATCGCAC GAACTTCACCA**C**	1	94	46 (10) [3]
C2	**CACCG**CGGGCCGC CCACTCCAAGAG	**AAAC**CTCTTGGAG TGGGCGGCCCG**C**	1	85	49 (12) [3]
p16+p19					
C1	**CACCG**CGGTGCAG ATTCGAACTGCG	**AAAC**CGCAGTTCG AATCTGCACCG**C**	2	94	33 (8) [3]
C2	**CACCG**CGCTGCGT CGTGCACCGGGC	**AAAC**GCCCGGTGC ACGACGCAGCG**C**	2	94	32 (11) [3]
p15					
C1	**CACCG**TTGGGCGG CAGCAGTGACGC	**AAAC**GCGTCACTG CTGCCGCCCAA**C**	1	84	97 (22) [2]
C2	**CACCG**CACTTGCC CCCGCGCCGCGG	**AAAC**CCGCGGCGC GGGGGCAAGTG**C**	1	90	52 (21) [3]
ffLuc					
C1	**CACCG**CTTCGAAA TGTCCGTTCGGT	**AAAC**ACCGAACGG ACATTTCGAAG**C**	2	96	31 (2) [3]

Boldface indicates non-targeting sequence added to generate *BsmB*1 restriction recognition sites and the PAM (NGG) site.

* Quality score is the inverse likelihood of off target binding in the mouse genome (Shalem et al. 2014).

** Number of potential off-target sites in the mouse genome. Numbers in parenthesis indicate the number of sites within genes. Numbers in square brackets indicate the minimum number of mismatches between the guide sequence and the potential off-target sites.

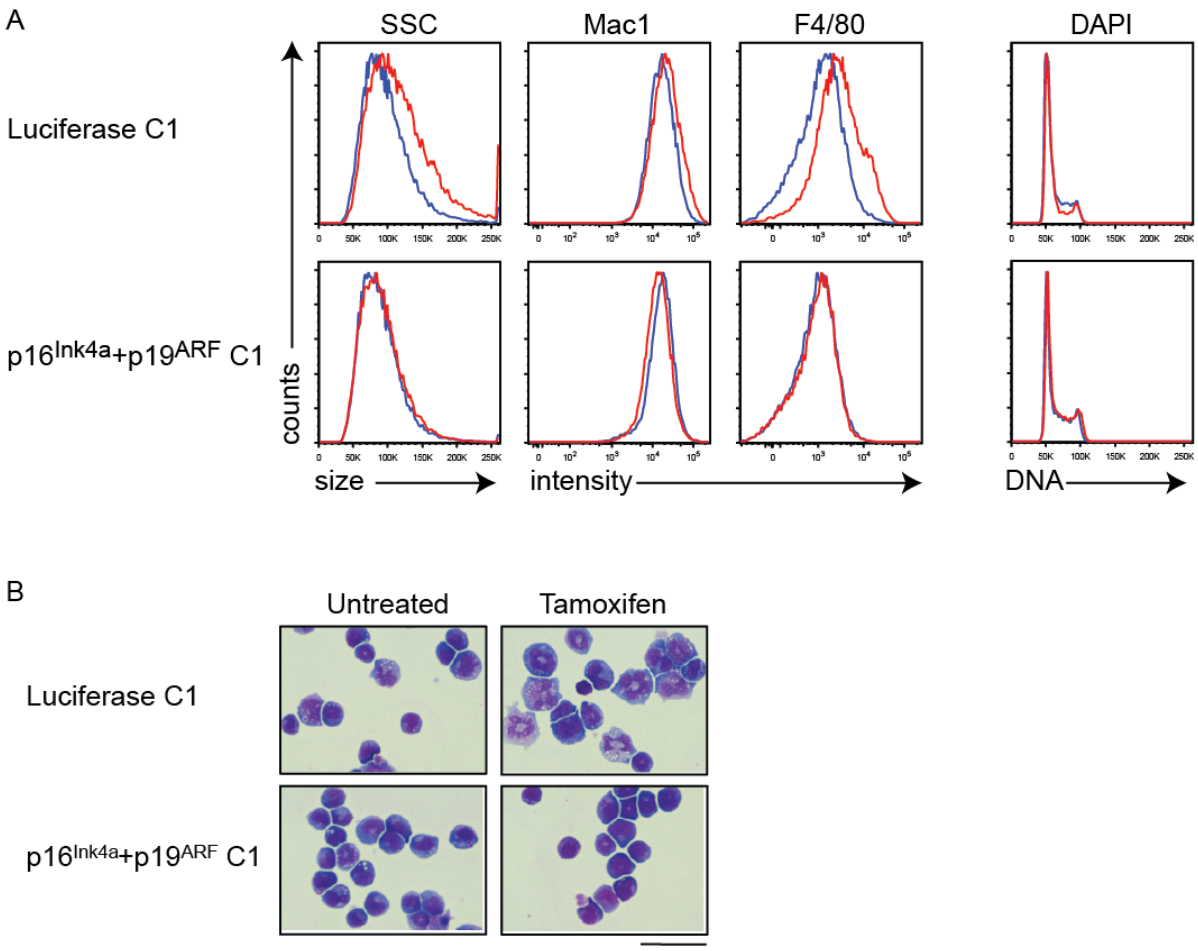

Figure S7 Related to Figure 6. Loss of Hhex does not induce differentiation of MLL-ENL lines lacking p16^{Ink4a} and p19^{ARF}. Cre^{ERT2};Hhex$^{-/fl}$ MLL-ENL cell lines expressing the indicated CRISPR constructs were grown for 7 days +/- tamoxifen, then either **A.** analyzed for granularity (side scatter, SSC), differentiation markers and DNA content by flow cytometry (Blue line is untreated, Red line is tamoxifen treated) or **B.** cytocentrifuged onto microscope slides and stained with May-Gruenwald-Giemsa stain. Scale bar= 10μM.

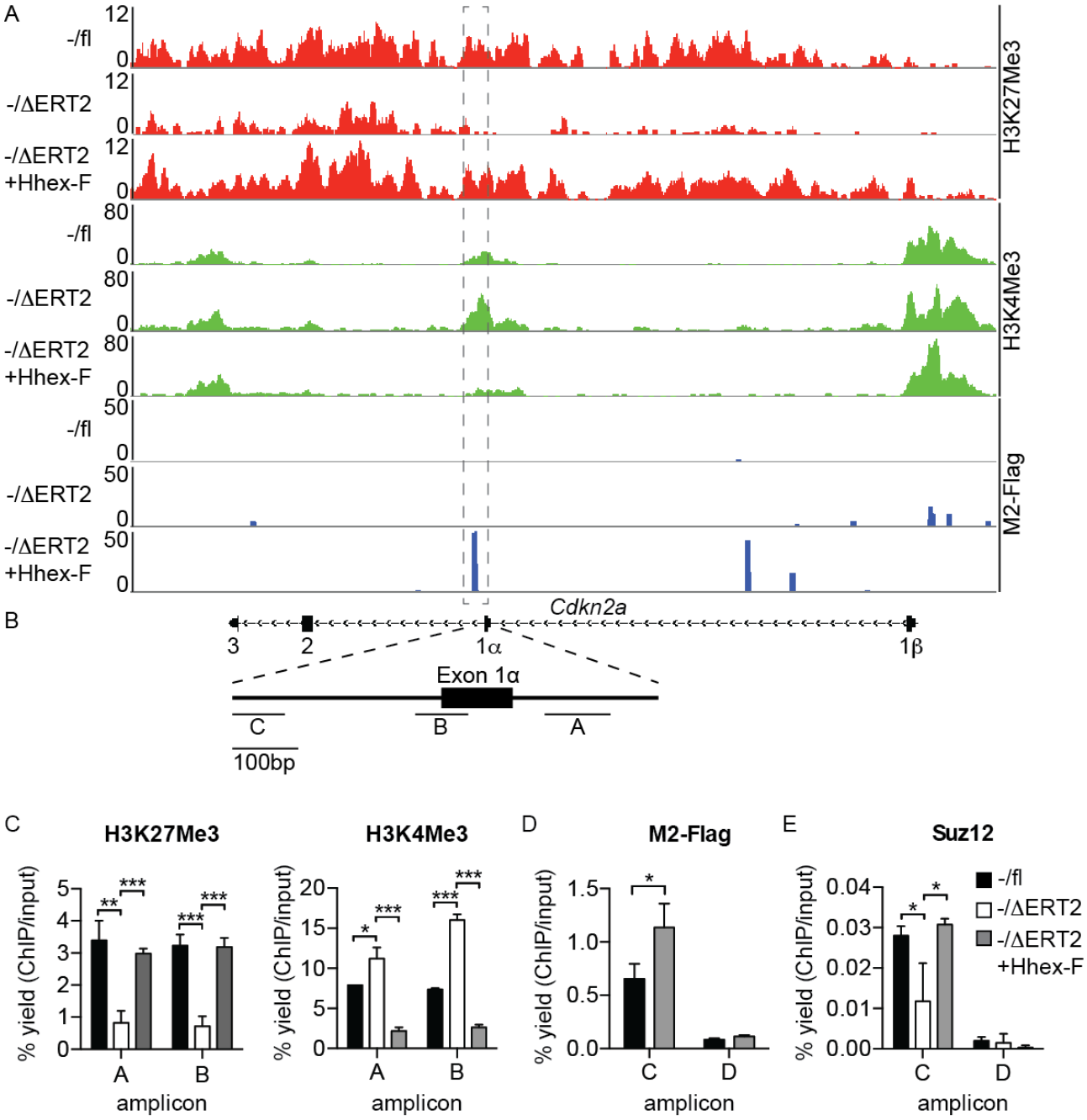

Figure S8. Related to figure 7. Hhex binds directly to *Cdkn2a* at a site ~130bp downstream of Exon 1α A. ChIP sequencing analysis, showing enrichment of H3K27Me3 (red), H3K4Me3 (green) and M2-Flag (blue) at the *Cdkn2a* locus in the indicated cell lines. Units are reads per million mapped reads (RPM). **B.** Schematic diagram of $p16^{Ink4a}$ Exon 1α showing the regions targeted for qPCR amplification, where A=p16 upstream promotor, B=p16 Exon 1α, C=Hhex binding motif. D= β actin Intron 1 (not shown). **C.** ChIP qPCR analysis of H3K27Me3 and H3K4Me3 enrichment at $p16^{Ink4a}$ Exon 1α in the indicated cell lines. **D-E.** ChIP qPCR analysis of Hhex-F binding (**D.** M2-Flag ChIP) and Suz12 binding (**E.**) enrichment at Hhex binding site identified in **A**. in the indicated cell lines. Data are mean +/- SD of triplicate determinations and each panel is representative of 2 independent experiments. *P* was calculated using the Student's T test, where *= *P*>0.05, **=*P*>0.01 and ***=*P*>0.001.

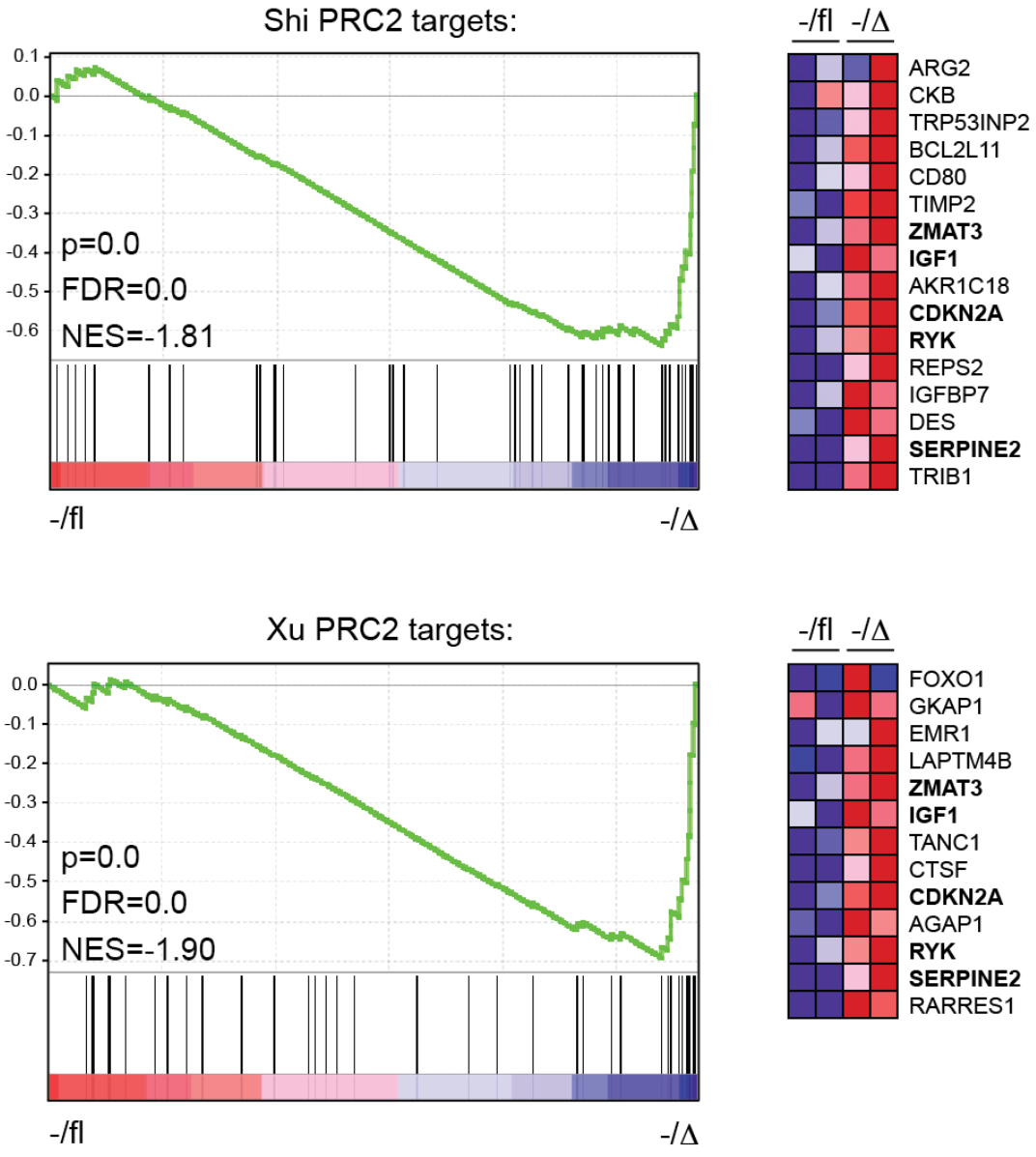

Figure S9. Related to Figure 7. Hhex regulates the expression of a discreet set of PRC2 target genes in MLL-ENL cell lines. GSEA analysis of RNA sequencing data from MLL-ENL-induced control (-/fl) and Hhex-deleted (-/Δ) cell lines demonstrates significant positive association between loss of Hhex and reactivation of PRC2 regulated genes. **Bold**= genes in common to both datasets.

Table S6: Excel file: Gene sets used in GSEA

Supplementary Materials and Methods

Soft agar colony assays

Soft agar colony assays were performed as previously described (Alexander et al. 1995). BM cells were seeded at 25,000 cells/plate in 1 ml of 0.3% agar in DMEM supplemented with 20% fetal calf serum in the presence of mIL-3 (10ng/mL), mSCF (100ng/mL), mouse GM-CSF (mGM-CSF; 10ng/mL), mouse G-CSF (mG-CSF; 10ng/mL), mouse M-CSF (mM-CSF; 10ng/mL), and human EPO (hEPO; 4U/mL) as indicated. Cultures were incubated for 7 days at 37°C in a fully humidified atmosphere of 10% CO_2 in air. Cultures were then fixed and sequentially stained for acetylcholinesterase and with Luxol fast blue and hematoxylin, and the cellular composition of each colony was determined by microscopic examination.

RNA sequencing

RNA was extracted from sorted LSKs, GFP^+ LICs and GFP^+ leukemia cell lines using the RNeasy Kit (Qiagen, Valencia CA). RNA quality and quantity was evaluated with a 2100 Bioanalyser (Agilent, Santa Clara CA). 16-46 million single-end 100bp reads were generated for Illumina libraries by an Illumina HiSeq sequencer. Reads were mapped to the GRCm37/mm9 build of the *Mus musculus* genome using the *Subread* aligner (Liao et al. 2013). Only uniquely mapped reads were retained. 83-91% of reads were successfully mapped for each library. Mapped reads were assigned to mouse RefSeq genes using the *featureCounts* program (Liao et al. 2014). NCBI RefSeq mouse annotation build 37.2 was used. Genes were removed from analysis if they failed to achieve a CPM (counts per million assigned reads) value of 1 or greater in at least one library. Read counts were converted to log2 counts per million, quantile normalized and precision weighted with the voom function of the limma package (Smyth 2005; Law et al. 2014). A linear model was fitted to each gene, and empirical Bayes moderated t-statistics were used to assess differences in expression (Smyth 2004). P-values were adjusted to control the global false discovery rate across all comparisons with the 'global' option of the limma package. Genes were called differentially

expressed if they achieved a false discovery rate of 0.1 or less. Read depth analysis was performed using the Integrated Genome Browser (www.bioviz.org) and heat maps were generated using the HeatMapViewer module of the GenePattern suite (genepattern.broadinsitute.org). Gene Set Enrichment Analysis (GSEA) was performed essentially as described (Subramanian et al. 2005) with GSEA v2.0 software, using gene set permutation and 1000 permutations per analysis. RNA sequencing data have been deposited in NCBI Gene Expression Omnibus under accession number GSE74019.

ChIP sequencing (ChIP-Seq)

ChIP-Seq DNA libraries were constructed using TruSeq Nano DNA sample preparation according to the manufacturer's instructions and sequenced on an Illumina NextSeq 500 at WEHI. Each sample achieved 8.5-114 million 150bp single-end reads. The reads were aligned to the mouse genome (mm10) using the Rsubread aligner (Liao et al. 2014). The BAM files of read alignments were then sorted and indexed using SAMtools (Li et al. 2009) and used to generate depth graph plots using IGB software.

To determine differential binding gene body of the H3K27Me3 ChIP sequencing samples the number of fragments overlapping each region were counted using featureCounts (Liao et al. 2014) and NCBI RefSeq annotation. Differential binding analyses were then performed using the Bioconductor package edgeR (Robinson et al. 2010). All genes that did not achieve a count per million of 2 in at least 1 sample were deemed to be unexpressed and subsequently filtered from the analysis. Additionally, genes with no official symbol in the NCBI gene information file were removed. Following filtering, 14,247 genes were left in the H3K27Me3 gene body data for downstream analysis. Compositional differences between samples were normalized using the trimmed mean of M-values (TMM) method (Robinson and Oshlack 2010). Differential binding for the two comparisons within each sample group, -/ΔERT2 versus -/fl and -/ΔERT2-Hhex-F versus -/fl, was assessed using t-statistics generated by generalized linear model (glm) methods developed

by McCarthy *et al* (McCarthy et al. 2012), with a pre-set dispersion of 0.05. Genes were considered to be differentially bound if they attained a false discovery rate of 0.05. The mean-difference plots were drawn using edgeR's plotSmear function, and the strength of the correlation shown in the log-fold change plot was evaluated using a linear model between the log-fold change RNA-seq and H3K27Me3 ChIP data. ChIP-Seq data have been deposited in NCBI Gene Expression Omnibus under accession number GSE74019.

ChIP quantitative PCR analysis

Purified ChIP DNA was used in quantitative PCR reactions using the following oligonucleotide primers pairs to amplify $p16^{Ink4a}$ Upstream promotor (A): 5'-GGCTGTGGAGCCAGGTCAGG-3', 5'-GCCCAATCGCCCAGTCGTGT-3', $p16^{Ink4a}$ Exon 1α (B): 5'-TCGCCCAACGCCCCGAAC-3', 5'-TCCTGAACAAAAGTTACCCGACTGC-3' (B), Hhex binding site (C) 5'- TTGAGAAGTCTTGTTTCTCCCA-3', 5'-GCCAGGACTCCTTTTAGGCT-3' and β actin Intron 1 (D): 5'-CGTATTAGGTCCATCTTGAGAGTACACAGTATT-3', 5'-GCCATTGAGGCGTGATCGTAGC-3' using SYBR Green (Roche). qPCR experiments were performed on a LightCycler 480 Real- Time PCR System (Roche).

Supplemental References

Alexander WS, Metcalf D, Dunn AR. 1995. Point mutations within a dimer interface homology domain of c-Mpl induce constitutive receptor activity and tumorigenicity. EMBO J 14: 5569-5578.

Law CW, Chen Y, Shi W, Smyth GK. 2014. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome biology 15: R29.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079.

Liao Y, Smyth GK, Shi W. 2013. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic acids research 41: e108.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30: 923-930.

McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic acids research 40: 4288-4297.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139-140.

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome biology 11: R25.

Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE, Doench JG et al. 2014. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 343: 84-87.

Smyth GK. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 3: Article3.

Smyth GK. 2005. 2005. Limma: linear models for microarray data. in Bioinformatics and computational biology solutions using R and Bioconductor (ed. R Gentleman), pp. xix, 473 p. Springer, New York ; London.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102: 15545-15550.