SUPPLEMENTARY INFORMATION

Chromatin-immunoprecipitation, endonuclease and sequencing (ChIP-Endo-Seq).

Cells were resuspended in 500 ul of FA-lysis (50 mM HEPES, 150 mM NaCI, 2 mM EDTA (pH 8.0), 1% Triton X-100, 0.1% NaDeoxycholate) with 0.2% added SDS and sonicated twice at power of 4 for 10 seconds using a micro-tip on a Branson Sonifer 250. The sonicate was then spun down at 13,000 rpm for 5 min and then the supernatant is transferred to a new tube and 1.5 mls of FA-lysis without SDS was added to the sonicated lysate and then divided evenly for each 1-3 ul antibody used (H3K4me3, Abcam ab8580; H3K27ac, Abcam ab4729; H3K27me3, Abcam ab6002; H3K4me1, Abcam ab8895; Pol2, Santa Cruz SC-899X). 20 ul of protein A/G magnetic beads (Pierce Protein A/G Magnetic Beads Prod #88803) was added to microtubes (QIAGEN Collection Microtubes Cat No. 19560) and then the mix was incubated overnight at 4C using a rocker.

Before each reaction beads were precipitated with a magnetic block and then washed and resuspended with 200 ul of the following solutions: FA lysis, High Salt (50 mM HEPES, 1 M NaCI, 2 mM EDTA (pH 8.0), 1% Triton X-100, 0.1% NaDeoxycholate), Wash 2 (50 mM Hepes/KOH, 0.5 M NaCI, 2 mM EDTA (pH 8.0), 1% TritonX-100, 0.1% NaDeoxycholate), Wash 3 (25 mM LiCI, 1% NP4O-Nonidet (IPEGAL), 1% NaDeoxycholate, 10 mM Tris-CI (pH 8.0)) and Tris buffer (10 mM Tris-HCI (pH depends on following reaction)) (adapted from Rhee and Pugh Reference). After the first set of washes, the chromatin bound beads are digested with either micrococcal nuclease (MNase) digestion buffer (previous Tris wash pH 8.5; 18.84 ul of 10 mM Tris pH 8.5, 0.16 ul of 500 mM CaCl2 and 1 ul of MNase (0.00002 U/ul, MN info)) for H3K4me3, H3K27ac, H3K27me3 and H3K4me1 or DNaseI digestion buffer (previous Tris wash pH 7.5; 18 ul of 10 mM Tris pH 7.5, 0.2 of 500 mM MgCl2, 0.8 ul of 500 mM CaCl2 and 1 ul of DNaseI (0.02 U/ul for PolII and 0.006 U/ul for transcription factors, DNAseI info). Beads were again washed with the last wash at pH 7.5 then incubated in end repair mix (2 ul of H2O, 4 ul of 10x T4 PNK buffer (NEB), 2 ul of DTT 100 mM, 8 ul of 2 mM dNTP, 2 ul of 20 mM rATP, 1 ul of T4 DNA polymerase (3 U/ul NEB M0203S) and 1 ul T4 polynucleotide kinase (10 U/ul NEB M0201L)) for 30 minutes at 20C at 1400 rpm. Beads were again washed in which the last wash was at pH 7.5 then incubated in adaptor ligation mix (5.3 ul of H2O, 4 ul of 10x T4 PNK buffer (NEB), 8 ul of 30% PEG, 2 ul of 20 mM rATP, 0.4 ul of 50 uM P5 and P7 (Illumina TruSeq adaptors, P5 is IS1: 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3' and IS3 5'-AGATCGGAAGAGC-3'; P7 is IS2 5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3' and IS3) and 0.3 ul T4 DNA ligase (30 Weiss U/ul Thermo Scientific EL0013) for 15 minutes at 20C at 1400 rpm. Beads were again washed in which the last wash was at pH 8.8 then incubated in the fill in mix (10 ul of H2O, 4 ul 10x ThermoPol buffer (NEB B9004S), 5 ul of 2 mM dNTP and 1 ul of Bst Polymerase, Large Fragment (8 U/ul NEB M0275S)) for 20 minutes at 37C at 1400 rpm. Beads were again washed in which the last wash was at pH 7.5 then the chromatin was eluted with 200 ul of elution buffer (25 mM

Trizma, 2 mM EDTA (pH 8.0), 200 mM NaCI and 0.5% SDS) and incubated for 15 min at 65C at 1400 rpm followed by magnetic precipitation and the supernatant was transferred to a new microtubes. 0.5 ul of Protease K (Protease K, recombinant PCR Grade Roche Diagnostics 11733400) was added and then incubated overnight at 65C.

1 ul of glycogen (Glycogen for molecular biology Roche Diagnostics 14397127) and 400 ul 100% ethanol was added to the eluted chromatin and then incubated 1 hour at -80C. DNA was then precipitated for 30 minutes at 13000 rpm and then rinsed with 70% ethanol and resuspended in 30 ul H2O. DNA concentration for each sample was initially measured by QPCR using oligos IS4 (5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT-3') and Indexing Oligo (5'-CAAGCAGAAGACGGCATACGAGATcctgcgaGTGACTGGAGTTCAGACGTGT-3') 2 ul of DNA was added to 10.5 ul H2O and then mixed with 12.5 ul of 2x QPCR mix (QPCR 2x reaction buffer: 20 mM Tris-HCl (pH 8.3), 13 mM $MgCl_2$, 100 mM KCl, 400 uM dNTPs, 4% DMSO, 2x SYBR Green I (Molecular Probes), 0.01% Tween 20, 0.01% NP40, 10 ng/ul of each oligo primer, and 0.025–0.1 U/ul of Taq polymerase (Roche)). This mix was then measured in four separate wells with 5 ul each in a 384 well plate in a Roche Light Cycler 480 QPCR machine (15 sec at 95C, 15 sec at 62C and 60 sec at 72C for 40 cycles)(Bryant et al 2008). The measurement was then analyzed using Absolute Quantification/2^{nd} Derivative Max and then the average Cp for the quadruplicate measurement noted. A second PCR reaction (regular non-quantitative 10 mM Tris HCl pH 8.3, 1.5 mM MgCl2, 50 mM KCl, 0.2 mM dNTP, 10 ng/ul of each oligo primer, and 0.025–0.1 U/ul of Taq polymerase (Roche)) with 2 ul of DNA in a total of 50 ul reaction mix using oligos IS4 and Indexing Oligo for 5 cycles more then the Cp value measured previously by QPCR. 20 ul from this PCR reaction was run on a 3% agarose gel (1.5% Agarose Genetic Analysis Grade Fisher Scientific BP1356-500 and 1.5% NuSieve GTG Agarose Lonza) and imaged using ChemiDoc MP Imaging System from Bio Rad and the lanes were quantitated using Image Lab 5.0 software. DNA abundance was measured for each lane from 150 – 550 bp range. To make the sequencing library, the previously determined DNA measurements were then used to calculate the least number of PCR rounds needed to amplify a sufficient amount of DNA to visualize on an agarose gel (assuming a PCR growth rate of 1.8 fold per round and 20 ul of DNA added to the PCR mix instead of 2 ul). For each ChIP a different Indexing Oligo, each with a unique index was used. After amplification 10 ul of the 50 ul was again run on a 3% agarose gel and DNA measured as described above. This measurement was then used to calculate the amount of DNA needed to mix together in a pool at its desired ratio for sequencing. 10 million hits per ChIP is a good target. Pooled DNA was then ethanol precipitated by adding 1 ul glycogen and 2 volumes of 100% ethanol. This was then resuspended in 20 ul H2O and all of it runs on a 10% acrylamide. The lane from 150 – 550 bp in size was cut out, the gel was mechanically crushed and the DNA was eluted overnight in 600 ul of oligo elution buffer (10 mM TrisCl pH 7.5, 50 mM NaCl, 1 mM EDTA pH 8.0). The eluted DNA was then ethanol precipitated as described

above. This pool was then resuspended in 30 ul H2O and submitted for paired end sequencing with 50 bp read on each side.

Genome Mapping

All data from ChIP-Endo-Seq experiments were mapped to mm10 using the following steps. Data from the sequencing facility was received as fastq files in which the bar codes were already removed and sequence data from each bar code was put into separate pairs of fastq files. Each of the paired fastq files was first modified to eliminate the adaptor sequence (since many sequences are less 50 bp in size) using the program cutadapt-1.9.1 with the command 'cutadapt -a AGATCGGAAGAGC'. The pair of fastq files was then examined and the paired sequence info was eliminated if either sequence in the pair had all of its sequence removed by cutadapt. The sequence was then mapped to mm10 using the program bwa-0.7.13 and the commands 'bwa aln' followed by 'bwa sampe', which generated a sam file for each pair of fastq files. All sam files from the same immunoprecipitation were then examined to eliminate any duplicates (i.e. sequences that have the same chromosome, start and end position). Finally, a new file was generated that contained chromosome number, start and end position for each immunoprecipitation.

Data Analysis

The above position data was then used to calculate multiple different sets of features. Features could be of many different categories, for example, every 2000 bp of the entire genome, all of the genes transcribed regions (GRCm38), or CpG Islands. Feature values were calculated from position data by counting the total number of midpoints that are within a given feature (count). This count is then divided by the size, in bp, of the feature and this quotient is then divided by total number of positions mapped divided by the genome size. Thus, a value of 1 for a feature would indicate that the ChIP-Endo-Seq neither enriched nor depleted immunoprecipitated sequences for that feature. This value will be referred to as the unnormalized value.

To reliably compare feature values between experiments, quantile normalization was performed for each feature category and immunoprecipitation antibody. A standard was chosen from an immunoprecipitation that had a large number unique positions mapped for each antibody. The unnormalized feature values were then quantile normalized to the standard for each antibody and feature category. The above quantile normalized values will be referred to as normalized values.

Genes and CpG Islands Features

Gene locations where taken from GRCm38 where the largest transcript for each gene is used to calculate the transcribed region. Individual CpG Islands where clustered together if they are within 500 bp of each other and the highest individual measurement is used to represent the entire cluster. CpG Island clusters where

associated with the closest gene with a transcribed region greater then 5,000 bp and is within 2000 bp of the transcribed region.

Distribution Plots

Distribution plots were made by first performing a selection (e.g. CpG Islands with a H3K4me3 signal above a threshold) and then measuring another signal associated with the selection (e.g. the Pol II signal in the gene associated with the CpG Island) to generate a distribution function. First, Log2 is calculated for each of the values (note: a small number, e.g. 0.6, is added to the value prior to taking the Log2 to prevent the undefined value of Log2 0). Second, the number of Log2 values that lies within successive non-overlapping ranges is counted (e.g. how many Log2 values are between 1.0 and 1.5, how many between 1.5 and 2.0 etc.). The density value is then calculated by first dividing this count by the total number of values, giving the fraction of values within each range. This fraction is then divided by its range size to give the p value. For example, if there are 1000 values being examined and 100 have a Log2 value between 1.0 and 1.5 then the density value will be (100 / 1000) / (1.5 – 1.0) = 0.2. Each of the density values will then be plotted at the midpoint of its range (i.e. (1.5 + 1.0) / 2 = 1.25).

Statistical Methods

Error for the genomic ChIP-Endo-Seq experiments was estimated using the negative binomial distribution equation:

$$\mathrm{NB}[\mu, \phi, x] = \mathrm{Binomial}\left[\frac{1}{\phi} + x - 1, x\right] \left(\frac{\mu\phi}{\mu\phi + 1}\right)^{x} \left(\frac{1}{\mu\phi + 1}\right)^{\frac{1}{\phi}} \quad (1)$$

where mu is the normalized count and phi is the dispersion for each given feature. Mu is calculated by first adjusting the count value at each experiment, antibody and feature by multiplying the count by the ratio of the average number of genome wide counts for all experiments, for a given antibody, by the total counts genome wide for the given antibody and experiment. These adjusted counts are then quantile normalized as described in Data Analysis. Phi is calculated for each feature, antibody and experiment by comparing replicate experiments and adjusting phi to minimize the difference of eCDF (experimental cumulative distribution function) curve of the p-values between the replicates and the uniform distribution. The phi value calculated does vary dependent on both the mu value and the feature size, so phi is adjusted separately for limited ranges of mu and feature size. A function for calculating phi dependent on mu and feature size: Phi(mu, size) is created by doing a linear interpolation of the phi values calculated at the midpoint of each range of mu and feature size. Phi is then calculated of for all features for each antibody and experiment using the Phi(mu, size) function. P-values were then calculated for all features and antibodies between all pairs of experiments. P-values were directly calculated by comparing the negative binomial distribution equations as follows:

$$P = \frac{S_a + S_b}{S_t} \quad \text{where} \qquad (2)$$

$$S_a = \sum_{x=0}^{a} \mathbf{avgNB}[\mu_1, \phi_1, \mu_2, \phi_2, x] * \mathbf{avgNB}[\mu_1, \phi_1, \mu_2, \phi_2, (\mu_1 + \mu_2 - x)]$$

$$S_b = \sum_{x=b}^{\mu_1+\mu_2} \mathbf{avgNB}[\mu_1, \phi_1, \mu_2, \phi_2, x] * \mathbf{avgNB}[\mu_1, \phi_1, \mu_2, \phi_2, (\mu_1 + \mu_2 - x)]$$

$$S_t = \sum_{x=0}^{\mu_1+\mu_2} \mathbf{avgNB}[\mu_1, \phi_1, \mu_2, \phi_2, x] * \mathbf{avgNB}[\mu_1, \phi_1, \mu_2, \phi_2, (\mu_1 + \mu_2 - x)]$$

$$a = \mathbf{Min}[\mu_1, \mu_2] \qquad b = \mathbf{Max}[\mu_1, \mu_2]$$

$$\mathbf{avgNB}[\mu_1, \phi_1, \mu_2, \phi_2, x] = \mathbf{NB}[\mu, \phi, x]$$

$$v_1 = \frac{\mu_1}{2} + \left(\frac{\mu_1}{2}\right)^2 \phi_1 \qquad v_2 = \frac{\mu_2}{2} + \left(\frac{\mu_2}{2}\right)^2 \phi_2 \qquad \mu = \frac{\mu_1 + \mu_2}{2} \qquad \phi = \frac{v_1 + v_2 - \mu}{\mu^2}$$

Where μ_1 and ϕ_1 is the mean and dispersion for one distribution and μ_2 and ϕ_2 is the mean and dispersion for the second distribution being compared. When multiple different features are examined to look for significant changes the False Discovery Rate of the method of Benjamini and Hochberg (1995) is used. P-values for significant differences comparing distribution functions are calculated using the Kolmogorov-Smirnov Test. The P-values for the RT-PCR experiments was calculated using a students T test on the log of the values.

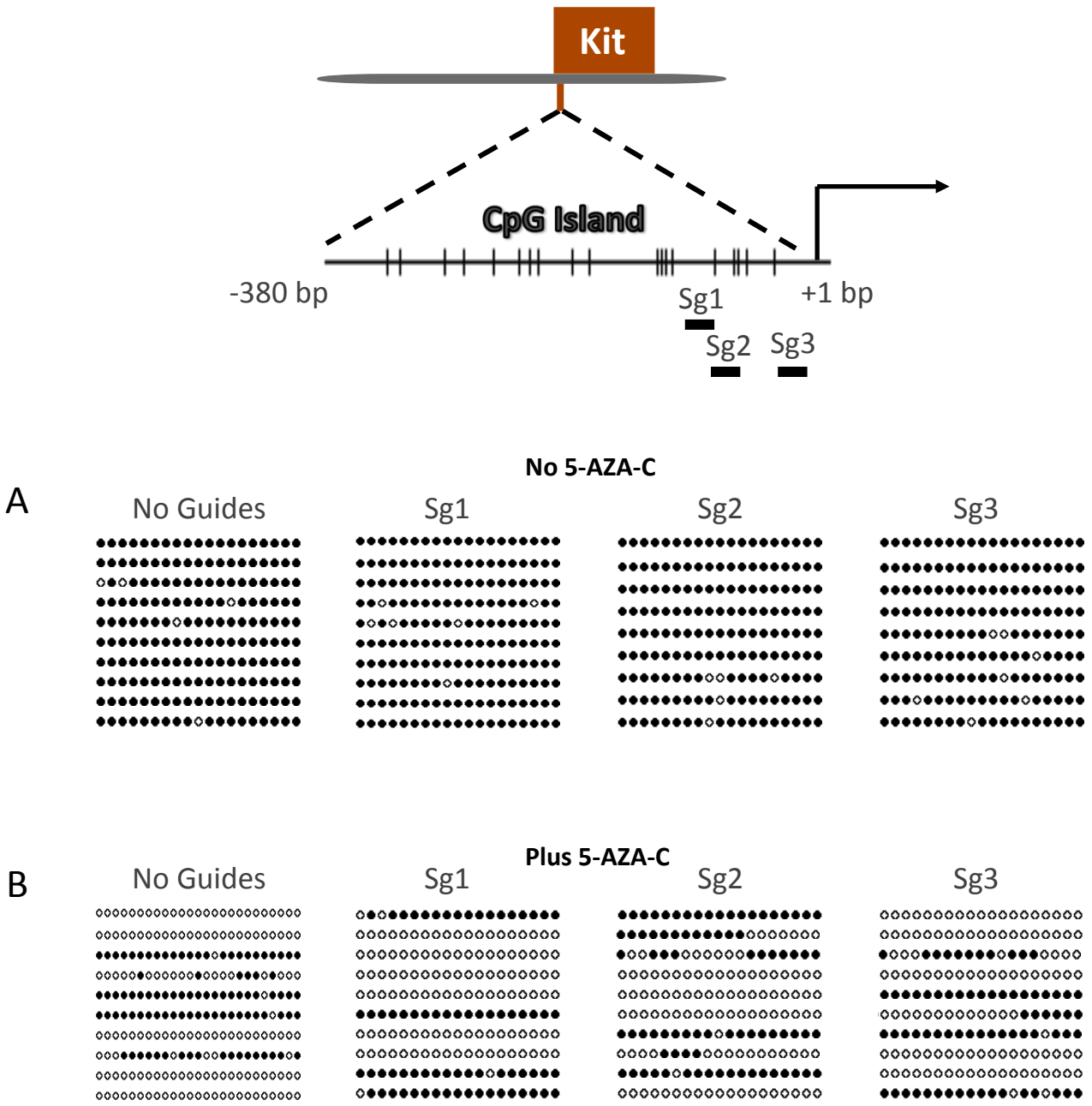

Figure S1. Methylation at the *Kit* promoter, Related with Figure 4. Methylation assay showing presence (black dots) or absence (white dots) of methylation at CpG sites in the *Kit* promoter**.** (**A**) WT myeloid cells, no 5-AZA-C, no added guide RNAs, no CRISPR activators (left-most panel); remaining panels, top line, no 5-AZA-C plus one or another of the guide RNAs, as indicated, plus CRISPR activators. (**B**) (Left most panel) WT myeloid cells plus 5-AZA-C, no CRISPR activators, no guide RNAs; remaining panels the CRISPR activators plus guide RNAs labeled Sg1, Sg2, and Sg3 as indicated.

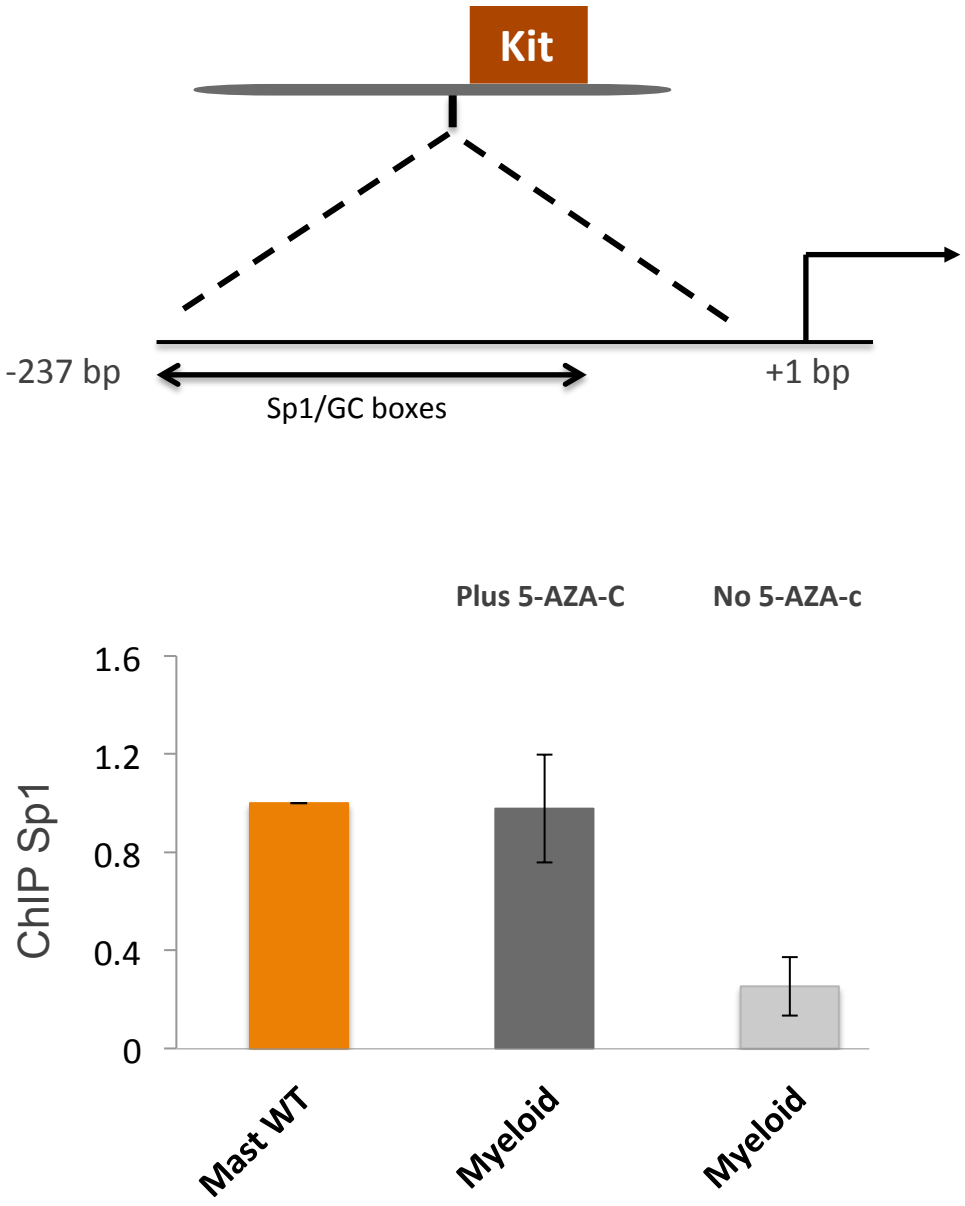

Figure S2. Sp1 binding upstream of the *Kit* promoter, Related to Figure 4. The ChIP-qPCR experiments detect Sp1 bound in the first 200 bp upstream of the *Kit* transcriptional start site. Orange bar: WT mast cells; dark, and light grey bars: myeloid cells expressing the guide-3 and, CRISPR activators with (dark grey) and without (light grey) 5-AZA-C. The significance of the decrease in Sp1 levels of mast WT over myeloid cells expressing the guide-3, No 5-AZA-C: $p < 0.025$.

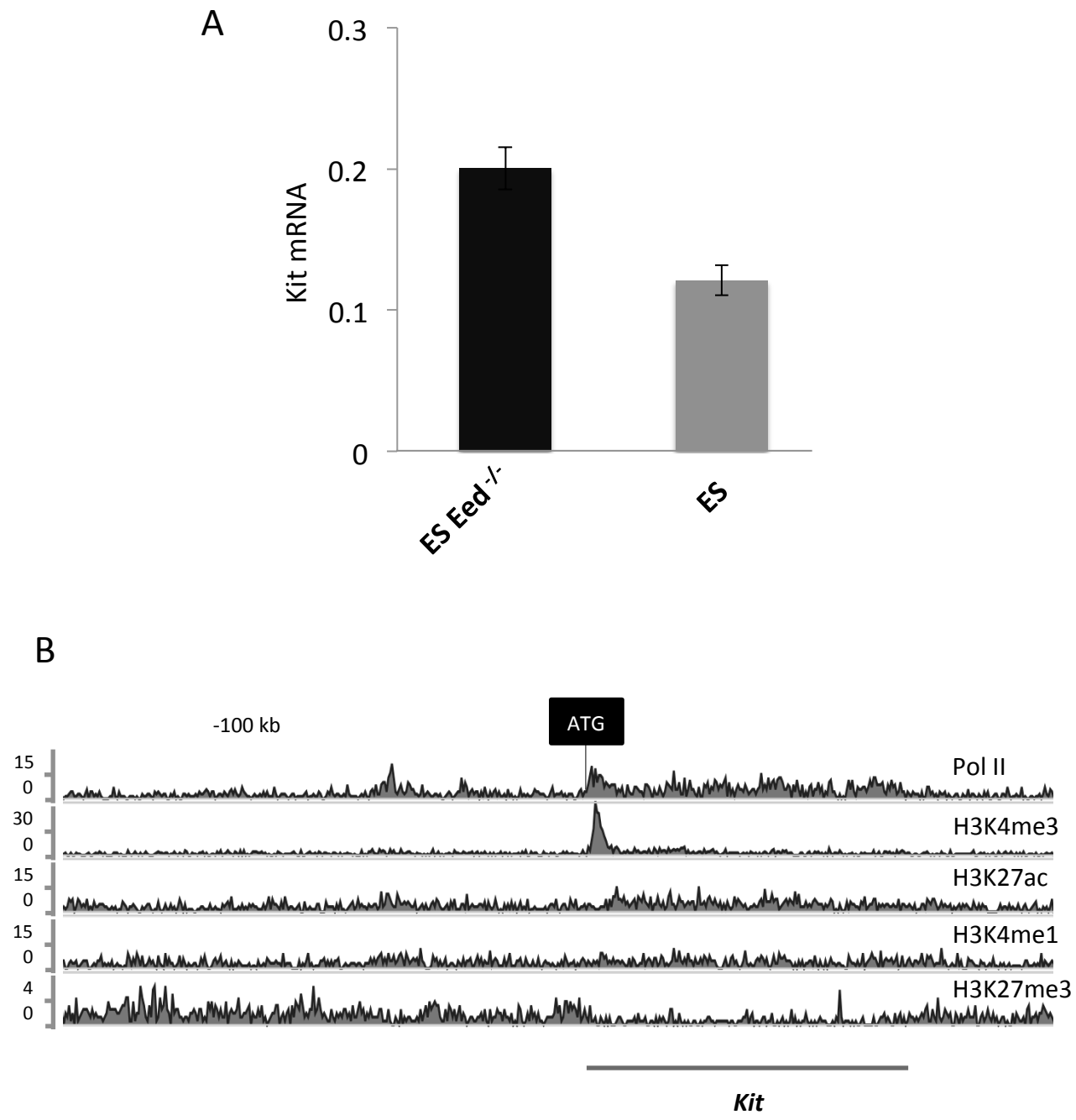

Figure S3. The effect of Pc on transcription of *Kit* in murine ES cells, Related to Figure 1. (**A**) The solid bars show that approximately twice as much *Kit* transcript is produced in Eed$^{-/-}$ ES cells compared to WT ES cells. (**B**) The distribution of several histone marks, including the H3K27me3 mark, at and around the CpG island *Kit* promoter in WT ES cells. The significance of the increase in *Kit* mRNA levels in ES Eed$^{-/-}$ over ES cells: $p < 0.012$.