Supplemental Data

H2AZ Is Enriched at Polycomb Complex

Target Genes in ES Cells and

Is Necessary for Lineage Commitment

Menno P. Creyghton, Styliani Markoulaki, Stuart S. Levine, Jacob Hanna, Michael A. Lodato, Ky Sha, Richard A. Young, Rudolf Jaenisch, and Laurie A. Boyer

Supplemental Experimental Procedures

Growth of murine embryonic stem (ES) cells

V6.5 (129SvJae and C57BL/6; male) ES cells were plated with or without irradiated murine embryonic fibroblasts (MEFs) and grown under typical ES cell conditions on gelatinized tissue culture plates. Briefly, cells were grown in DMEM supplemented with 15% fetal calf serum (Hyclone), leukemia inhibitory factor (LIF), non-essential amino acids (Invitrogen), L-glutamine (Invitrogen), and penicillin/streptomycin (Invitrogen) as previously described (Boyer et al., 2006). H2AZ-depleted ES cells were similarly cultured in ES cell media with the addition of 1ug/ml puromycin (Sigma). The suz12 null ES cell lines used in this study were a gift from K. Helin and were cultured in ES cell media (Pasini et al., 2007).

Derivation of neural precursor (NP) cells

V6.5 ES cells were differentiated along the neural lineage after plating 4 day-old embryoid bodies followed by selection for 5-7 days in chemically defined ITSFn media (Okabe et al., 1996). Pan-neural precursor cells were isolated, and cultured on polyornithine-coated tissue culture plastic in N3 media containing FGF and EGF (R&D Systems, Minneapolis, USA) as described previously (Conti et al., 2005). In the presence of growth factors these cells can be labeled homogenously with antibodies against Nestin, Sox2, and A2B5. Upon growth factor withdrawal, the cells differentiate into TUJ1-positive neurons, 04-positive oligodendrocytes and GFAP-positive astrocytes, the three major cell types of the nervous system (data not shown) indicating that the precursor cells sustain differentiation potential.

Chromatin immunoprecipitation

The ChIP protocol has been adapted from previous studies (Boyer et al., 2005; Boyer et al., 2006).

Murine embryonic stem cells were grown to a final count of $5x10^7$ – $1x10^8$ cells for each location analysis reaction. Cells were chemically cross-linked by the addition of one-tenth volume of fresh 11% formaldehyde solution for 15 minutes at room temperature. Cells were rinsed twice with 1xPBS and harvested using a silicon scraper and flash frozen in liquid nitrogen and stored at –80°C prior to use. Cells were resuspended, lysed, and sonicated to solubilize and shear crosslinked DNA. We used a Misonix Sonicator

3000 and sonicated at power 4 for 10 x 30 second pulses (90 second pause between pulses) at 4°C while samples were immersed in an ice bath. The resulting whole cell extract was incubated overnight at 4°C with 100 μl of Dynal Protein G magnetic beads that had been preincubated with 10 ug of the appropriate antibody for at least 3 hrs. Beads were washed 5 times with RIPA buffer and 1 time with TE containing 50 mM NaCl. Bound complexes were eluted from the beads by heating at 65°C with occasional vortexing and crosslinking was reversed by overnight incubation at 65°C. Whole cell extract DNA (reserved from the sonication step) was also treated for crosslink reversal. Immunoprecipitated DNA and whole cell extract DNA were then purified by treatment with RNaseA, proteinase K and multiple phenol:chloroform extractions. Purified DNA was blunted and ligated to linker and amplified using a two-stage PCR protocol. Amplified DNA was labeled and purified using Invitrogen Bioprime random primer labeling kits (immunoenriched DNA was labeled with Cy5 fluorophore, whole cell extract DNA was labeled with Cy3 fluorophore). Labeled DNA was combined (5 – 6 μg each of immunoenriched and whole cell extract DNA) and hybridized to arrays in Agilent hybridization chambers for 40 hours at 65°C according to the manufacturer's instructions. Arrays were then washed and scanned.

Antibodies

The H2AZ antibody used in this study (Abcam ab4174) has been shown to specifically recognize its target protein in ChIP assays (Barski et al., 2007). The Suz12 antibody was obtained from Abcam (ab12201) and was previously validated for ChIP assays (Boyer et al., 2006; Lee et al., 2006; Squazzo et al., 2006). Chromatin immunoprecipitations against H2AZ were compared to reference DNA obtained by chromatin immunoprecipitation of total histone H3 (Abcam ab1791; epitope derived from C-terminal 100 amino acids of histone H3) to normalize for nucleosome density. Additional antibodies used for ChIP, co-immunoprecipitation and immunoblot analysis were as follows: Histone H3 (Abcam ab1791), Histone H2A (Abcam ab18255), Oct4 (C-10 Santa Cruz sc-5279), H2AZ (Upstate 07-594), Rnf2 (Abcam ab3832), Ezh2 (Upstate 07-400), Nap1l1 (Abcam ab33076) and beta-actin (Abcam ab8226). Secondary antibodies used for immunoblot analysis were anti-Rb IgG HRP conjugate (Calbiochem 401393) and anti-mouse IgG HRP conjugate (Calbiochem 401253).

Control chromatin immunoprecipitations

Location analysis was performed with Histone H3 specific antibodies (Abcam ab1791) as a control for antibody specificity. We did not observe any similar enrichment with ChIP data generated with H2AZ specific antibodies. Control ChIP experiments were also performed with rabbit and mouse IgG pre-immune serum in murine ES cells as a negative control (data not shown). We did not find any enrichment for sequences occupied by H2AZ in the IgG control experiments.

Oligo microarray design

244K Promoter Array

The mouse promoter array set used in this study is based on a promoter array design described in previous studies (Boyer et al., 2006). The arrays were custom manufactured at Agilent Technologies (www.agilent.com (G4496A/B)). This platform consists of 2

slides each containing ~244,000 60mer oligos (slide ID 14928, 14929) (NCBI build 34 of the mouse genome (mm6) designed to cover regions between approximately -4 kb and +4 kb relative to the transcription start sites of all RefSeq transcripts and surrounding all mature miRNA sequences in RFAM. Additionally the entire mouse *HoxA* cluster is also tiled on these arrays. All regions are tiled at a density of approximately 1 oligo per 250 bp.

244k Chromosome 19 tiling microarray

The mouse whole chromosome array was purchased from Agilent Technology (www.agilent.com). The array consists of 1 slide containing ~244,000 60-mer oligos (slide ID 14926) based on sequences from NCBI build 34 of the mouse genome (mm6) designed to cover the entire non-repeat region of chromosome 19 at a density of approximately 1 oligo per 250bp. In addition, the array also contains oligos tiling homologous chromosomal regions to those manually selected for the human ENCODE project (ENm001-ENm014) including the HOXA gene cluster. Homologous regions were identified using the Liftover function at the UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgLiftOver).

Selection of regions and design of subsequences for promoter microarrays

To define transcription start sites, we first selected transcripts from three of the most commonly used databases for sequence information (Refseq, Ensembl, MGC). Transcription start sites within 500 bp of each other were considered single start sites. To restrict our array to the most likely transcription start sites, we selected only those start sites that were found in at least two of the three databases.

We used repeat-masked sequence from the mm6 build (March 2005) of the mouse genome. 25 kb of sequence around each transcription start site was extracted for analysis (20 kb upstream, 5 kb downstream). Each transcription start site was considered independent, even if the 25 kb region would overlap with another transcript's 25 kb region.

We used the program ArrayOligoSelector (AOS) to score 60-mers for every unmasked subsequence greater than 62 bp across all promoter regions. The scores for each oligo were retained but not put through the built-in AOS selection process.

The collection of scored 60-mers was divided by promoter and sorted by genomic position. Each set of 60-mers was then filtered based on the oligo scoring criteria. AOS uses a scoring system for four criteria: GC content, self-binding, complexity and uniqueness. For our most stringent filter, we selected the following ranges for each parameter: GC content between 30 percent and 100 percent, self-binding score less than 100, complexity score less than or equal to 24, uniqueness greater than or equal to –40.

From this subset of 60-mers, we selected oligos designed to cover the promoter region with an estimated density of one probe every 280 bp. At this point, we restricted oligo selection to those oligos found within the region 4 kb upstream to 4 kb downstream of the transcription start site. To achieve more uniform tiling, we instituted a simple method to find probes within a particular distance from each other. Starting at the upstream end of

the region, we selected the first qualified probe and then selected the next qualified probe that was between 150 bp and 280 bp away. If there were multiple, eligible probes, we chose the most distal probe within the 280 bp limit. If there were no probes within this limit, we continued scanning until we found the next acceptable probe. The process was then repeated with the most recently selected probe until we reached the end of the promoter region.

For regions that were not covered by high quality probes, we returned to the full set of scored 60-mers and filtered using less stringent criteria. This gave us an additional set of 60-mers that we then used to fill gaps in our coverage. After this second pass, we identified gaps in our coverage and added oligos that were properly spaced and best fit our criteria regardless of whether they passed the filter cutoffs. This iterative process gave us a compromise between optimal probe quality and optimal probe spacing.

Array Scanning and Data Extraction

Slides were scanned using an Agilent DNA microarray scanner BA. PMT settings were set manually to normalize bulk signal in the Cy3 and Cy5 channel. For efficient batch processing of scans, we used Agilent Feature Extractor Software. Scans were automatically aligned and then manually examined for abnormal features. Intensity data were then extracted in batch.

Data Normalization and Analysis

We used Feature extractor (Agilent) to obtain background-subtracted intensity values for each fluorophore for every feature on the whole genome and sub genomic arrays respectively. Among the Agilent controls is a set of negative control spots that contain 60-mer sequences that do not cross-hybridize to murine genomic DNA. We calculated the median intensity of these negative control spots in each channel and then subtracted this number from the intensities of all other features.

To correct for different amounts of each sample of DNA hybridized to the chip, the negative control-subtracted median intensity value of control oligonucleotides from the Cy3-enriched DNA channel was then divided by the median of the control oligonucleotides from the Cy5-enriched DNA channel. This yielded a normalization factor that was applied to each intensity value in the Cy5 DNA channel.

We calculated the log of the ratio of intensity in the Cy3-enriched channel to intensity in the Cy5 channel for each probe and used a whole chip error model (Hughes et al., 2001) to calculate confidence values for each spot on each array (single probe p-value). This error model functions by converting the intensity information in both channels to an X score which is dependent on both the absolute value of intensities and background noise in each channel using an f-score calculated as described (Boyer et al., 2005) for promoter regions or using a score of 0.3 for tiled arrays. When available, replicate data were combined, using the X scores and ratios of individual replicates to weight each replicate's contribution to a combined X score and ratio. The X scores for the combined replicate are assumed to be normally distributed which allows for calculation of a p-value for the enrichment ratio seen at each feature. P-values were also calculated based on a second

model assuming that, for any range of signal intensities, IP:control ratios below 1 represent noise (as the immunoprecipitation should only result in enrichment of specific signals) and the distribution of noise among ratios above 1 is the reflection of the distribution of noise among ratios below 1.

Identification of Bound Regions

To automatically determine bound regions in the datasets, we developed an algorithm to incorporate information from neighboring probes. For each 60-mer, we calculated the average X score of the 60-mer and its two immediate neighbors. If a feature was flagged as abnormal during scanning, we assumed it gave a neutral contribution to the average X score. Similarly, if an adjacent feature was beyond a reasonable distance from the probe (1000 bp), we assumed it gave a neutral contribution to the average X score. The distance threshold of 1000 bp was determined based on the maximum size of labeled DNA fragments hybridized to the array. Since the maximum fragment size was approximately 550 bp, we reasoned that probes separated by 1000 bp or more would not contribute reliable information about a binding event located half the distance between them. This set of averaged values gave us a new distribution that was subsequently used to calculate p-values of average X (probe set p-values). If the probe set p-value was less than 0.001, the three probes were marked as potentially bound.

As most probes were spaced within the resolution limit of chromatin immunoprecipitation, we next required that multiple probes in the probe set provide evidence of a binding event. Candidate bound probe sets were required to pass one of two additional filters: two of the three probes in a probe set must each have single probe p-values < 0.005 or the center probe in the probe set has a single probe p-value < 0.001 and one of the flanking probes has a single point p-value < 0.1. These two filters cover situations where a binding event occurs midway between two probes and each weakly detects the event or where a binding event occurs very close to one probe and is very weakly detected by a neighboring probe. Individual probe sets that passed these criteria and were spaced closely together were collapsed into bound regions if the center probes of the probe sets were within 1000 bp of each other.

Comparisons to Known Genes

The coordinates for the complete list of bound regions for H2AZ and Suz12 in mES cells and NPs can be found in Tables S1, S3, S5, S7, and S9 (see Index of Tables).

The location of all bound regions were compared to a composite set of transcripts compiled from three databases: RefSeq, Ensembl, and UCSC annotated known genes (http://genome.ucsc.edu/cgi-bin/hgTrackUi?g=knownGene) that were associated with EntrezGene identifiers. All coordinate information was downloaded from the UCSC Genome Browser (NCBI build 34 mm6; March 2005).

Thresholds and the fraction of bound gene calls

For simplicity in interpreting ChIP/chip data, genes have been assigned as either bound or unbound based on a single threshold model. However, binding data is rarely bimodal and instead represents a continuum that reflects the frequency of occupancy of a particular

binding site in a large population of cells and is affected by the availability of the epitope, the quality of the oligos, shear distribution, and the density of target sites. In previous studies we have identified thresholds that minimize false positives on genome wide analyses while accepting fairly high false negative rates. This threshold is termed "high confidence" and gene assignments based on this cutoff are listed in Tables S2, S6, and S10.

Gene ontology classification of bound genes

Gene ontology analysis was performed using DAVID (http://david.abcc.ncifcrf.gov/) (Dennis et al., 2003). EntrezGene IDs from the bound gene list were used to generate enrichment statistics for the biological process category based on a background list of all represented genes on our promoter microarray (Table S4 and S11).

K-means cluster analysis of binding data

The spatial pattern of H2AZ and Suz12 enrichment was compared at the high-confidence set of 1655 H2AZ-enriched genes using K-means clustering with three nodes (see Table S8 for list of genes in each cluster). For each start site, the raw enrichment ratio for the probe closest to that start site were selected in 250 bp increments to cover the region –3.5 to +3.5 kb relative to the TSS. Promoters where >25% of probes were unavailable (typically due to repeat masking) were removed from the analysis. The remaining probe ratios from the two experiments were used to create a single vector for each start site. The data were processed and the cluster diagram was generated using Cluster software (http://rana.lbl.gov/EisenSoftware.htm).

Generation and characterization of control and H2AZ-depleted clonal ES cell lines

19-mer hairpin oligonucleotides for H2AZ #1 (5'-TGGAGATGAAGAATTGGAT-3') and #2 (5'-GGTAAGGCTGGAAAGGACT-3') and for GFP (5'-GCTGACCCTGAAGTTCATC-3') were cloned into pRETRO-SUPER as described previously (Brummelkamp et al., 2002). Ecotropic retroviral supernatants were generated by transfection of phoenix packaging cells using FuGENE 6 (Roche) according to the manufacturers protocol. 48 hours post-transfection, the tissue culture medium was filtered through a 0.45 um filter, and diluted 1:1 with ES cell media. This viral supernatant was used for infection of ES cells after addition of 4 ug/ml polybrene (Sigma). Cells were infected over night and allowed to recover for 48 hr with fresh ES cell medium. Stable viral integrants were selected with puromycin 1 ug/ml for 72 hr. Single colonies were picked and clonally expanded under puromycin selection.

Immunoblot analysis

Cells were lysed in NETN lysis buffer (100 mM NaCl, 1 mM EDTA, 20 mM Tris at pH 8.0, 0.5% NP-40) containing complete protease inhibitor cocktail (Roche Diagnostics). Samples were normalized for protein concentration using Bio-Rad protein assay detection reagent and boiled in SDS sample buffer (62.5 mM Tris PH 6.8, 10% Glycerol, 5% β-mercapto-ethanol, 2% SDS, 0.05% bromophenolblue). Samples were resolved by SDS-PAGE and transferred to polyvinylidine difluoride (PVDF) membrane (Millipore) using the NuPAGE Novex System (Invitrogen). Immunoblots were blocked and probed in 1X PBS-0.1% TWEEN-1% BSA (Sigma) with the indicated antibodies at the recommended

concentrations. Immunoblots were then incubated with an HRP-conjugated secondary antibody (Calbiochem) and visualized using the ECL western blotting detection System (Amersham).

Cell cycle analysis

ES cells were trypsinized and subsequently MEF-depleted by pre-plating for 45 minutes. 1 x 10^6 cells were resuspended in 0.5 ml PBS followed by the addition of 0.5ml of 100% ice-cold ethanol to the cells in a drop-wise while vortexing. After incubation for 20 minutes on ice, cells were harvested by centrifugation (1000 rpm for 5-7 minutes) and ethanol was decanted. Finally, 0.5ml of 7AAD-RNAse solution [(final concentrations 50ug/ml 7AAD (BD pharmingen 7-AAD cat. 51-68981E + 100ug/ml RNAse Type I-A in PBS)] was added to the cells. After 30 minutes of incubation, the samples were analyzed by flow-cytometry by using BD- FACS Calibur.

Proliferation Assay

Oct4-GFP knock-in V6.5 ES cells were mixed with either control wild-type or H2AZ deficient V6.5 ES cells. Mixed samples were passaged on MEF feeder cells, and every 3-4 days they were harvested and MEF depleted by pre-plating for 60 minutes. A fraction of the sample was analyzed to quantify the ratio between GFP positive (wild-type) and GFP negative (H2AZ-depleted) ES cells, while the remaining cellswere cultured on feeder cells for continued analysis. FACS analysis at time points 0, 4d, 8d and 13d are shown. .

Karyotype analysis of H2AZ-depeleted ES cell line

A T25 flask of H2AZ-depleted ES cell line #2 (passage 10) was sent to Cell Line Genetics (Madison, WI) for cytogenetic analysis. Twenty G-banded metaphase cells were examined from this cell line. While most of the cells had an apparently normal male karyotype (Figure S3), there was a minor clone consisting of two cells with loss of the Y chromosome. Clones with loss of the Y chromosome are not unusual in mouse ES cell lines (Eggan et al., 2002).

RNA isolation, real-time PCR and analysis

Relative transcript levels in ES cells were determined by reverse transcriptase real-time PCR, RNA was isolated from 10^7 ES cells with the TRIzol® Plus RNA Purification System (Invitrogen). cDNA was synthesized using the SuperScript® III First-Strand Synthesis SuperMix (invitrogen) using oligo(dT) primers according to the manufacturer's instructions. qRT-PCR reactions were performed with SYBR Green Master Mix (Invitrogen) and primers specific for H2AZ enriched (E) and control unenriched (U) genes (for sequences see Table S12) on an ABI Prism 7900HT Sequence Detection System (Applied Biosystems). The majority of primers were designed to amplify exon/exon boundaries using the program Primer Express 2.0 (ABI). Relative mRNA levels were determined in triplicate for each gene in both H2AZ-depleted cell lines using Relative Quantitation (RQ) using comparative C_T (Livak and Schmittgen, 2001) and relative GAPDH levels were used for normalization. The transcript levels in H2AZ-depleted ES cells were calculated as fold change relative to control ES cells expressing a non-specific hairpin. Data analysis was performed using the manufacturers software

(Applied Biosystems 7900HT Sequence Detection Systems (SDS) version 2.3). qRT primers used in Figure 7 are Sox1 forward : 5'-GGAAAACCCCAAGATGCACAAC-3', reverse: 5'-CGCAGTCTCTTGGCCTCGTC-3' and Nestin forward 5'- GCCTATAGTTCAACGCCCCC-3', reverse 5'-AGACAGGCAGGGCTAGCAAG-3'.

Chromatin immunoprecipitation and site-specific PCR analysis

Chromatin immunoprecipitation was performed in control (GFP) and H2AZ-depleted ES cell lines with antibodies directed against H2AZ, Suz12, or Rnf2/Ring1b using our standard ChIP protocol as described above. For relative quantification by real-time PCR, dilutions of the whole cell extract (WCE) DNA were used to construct standard curves. 1 nanogram (ng) ChIP DNA or WCE control DNA was used per reaction to determine the relative enrichment of target sequences. Triplicate reactions were performed with the SYBR Green PCR master mix (Invitrogen) on the ABI 7900HT Sequence Detection System using relative quantification (RQ) and comparative C_T. Primers were chosen to amplify a region within 1 kb upstream of the transcriptional start site. Primer sequences are listed in Table S13. Biological replicates were performed for ChIP experiments conducted in control, H2AZ-depleted cell lines, and in *suz12* null ES cells.

Co-immunoprecipitation of PRC2 and H2AZ in ES cells

Cells were lysed in 0.5 ml NETN lysis buffer (100mM NaCl, 1 mM EDTA, 20mM Tris pH8.0, 0.5% NP40) containing complete protease inhibitor cocktail (Roche) and complexes were immunoprecipitated for at least 6 hours with 5 μg of the indicated antibody that was pre-conjugated to protein G sepharose beads at 4°C. After 3 wash steps with cold lysis buffer, bound proteins were eluted by boiling in SDS-sample buffer and resolved by SDS-PAGE.

Analysis of the in vivo developmental potential of H2AZ-depleted ES cells

Because H2AZ is required for proper regulation of developmental target genes, we analyzed the developmental potential of the H2AZ-depleted ES cells by tetraploid complementation. This method is frequently utilized to examine the developmental potential of ES cells because the resulting embryos are composed entirely from the injected ES cells (Eggan et al., 2002). Control or H2AZ-depleted ES cells were injected into tetraploid (4N) blastocycts and embryos were recovered at dpc 8.5 for comparison (Figure S8A, left panel). While dpc (days post-coitum) 8.5 control embryos displayed normal morphology, H2AZ-depleted embryos were much smaller and appeared arrested at an earlier developmental stage. Embryos were also recovered from 4N injected blastocysts at dpc 6.5, fixed, sectioned and stained with hematoxylin and eosin (Figure S8A, right panel). The H2AZ-depleted embryos appeared disorganized and developmentally delayed around the time of implantation as compared to control embryos, phenotypes similar to those observed for *h2az* null embryos (Faast et al., 2001). Notably, *suz12* and *ezh2* (also a core component of PRC2) null embryos also fail to develop beyond this stage (O'Carroll et al., 2001; Pasini et al., 2004).

Because tetraploid complementation is the most stringent assay for pluripotency, where only the injected cells give rise to the embryo, we next determined whether the developmental arrest phenotype is cell autonomous. To this end, control and H2AZ-

depleted ES cell lines were transduced with a GFP expressing lentivirus and injected into diploid blastocysts to generate chimeric mice. GFP labeling allowed for the distinction of the experimental cell population from that of the host inner cell mass. Embryos were recovered at dpc 13.5 and the contribution of the control and H2AZ-depleted ES cells to development was determined by visualization of GFP expression (Figure S8B). While control ES cells could give rise to high contribution chimeras (left), H2AZ-depleted cells were unable to contribute to development (right). Together, these results demonstrate that H2AZ-depleted cells, similar to *h2az* null cells, cannot contribute to embryonic development and that this effect is cell autonomous.

- **2n and tetraploid blastocyst injections**

Tetraploid and diploid blastocyst injections were performed as previously described (Wang and Jaenisch, 2004) with the following modifications. Briefly, pronuclear stage embryos (dpc. 0.5) were obtained from hormone primed B6D2F1 females 18-20 h after mating to B6D2F1 males. Embryos for diploid injections were cultured continuously in KSOM medium in a CO_2 incubator for about 72h prior to injections, whereas embryos for tetraploid injections were subjected to fusion at the 2-cell stage (dpc 1.5) before injection at the blastocyst stage (dpc 3.5). Diploid blastocysts were injected with control or H2AZ-depleted ES cells infected with the lentiviral FUGW vector expressing GFP in order to visualize tissue contribution (Lois et al., 2002). Ten to twelve ES cells were injected into the blastocoel cavity of each embryo and embryos were transferred to pseudo-pregnant females (control 66 injections, 25 implanted, H2AZ#1 55 injections, 22 implanted, and H2AZ#2 75 injections, 25 implanted). Fetuses from tetraploid injections were dissected on dpc 8.5 for morphological assessment, or deciduas were dissected on dpc 6.5 and paraffin embedded for sectioning followed by hematoxylin and eosin staining after fixation in 10% formalin. Fetuses from diploid injections were retrieved on dpc 13.5 and contribution of injected GFP expressing cells to tissues was determined using fluorescence microscopy.

Analysis of the in vitro developmental potential of H2AZ-depleted cells

- **Embryoid body (EB) formation**

Embryoid Body (EB) culture is used to examine the differentiation potential of ES cells as shown in Figure 7. ES cells were diluted to 10,000 cells/ml in complete growth medium (see growth conditions for ES cells) lacking leukemia inhibitory factor (LIF) and 20 μl drops (200 cells/drop) were arranged over the surface of a Petri dish. The plate was gently inverted and incubated at 37°C with 5% CO_2 for three days to allow aggregation and embryoid body (EB) formation. EBs were then transferred to low attachment cell culture plates (Corning) with the addition of media for the remainder of the experiment. EBs were allowed to grow for up to 10-12 days and samples were analyzed at various time points by hematoxylin and eosin staining after fixation in 10% formalin or by real-time PCR following mRNA isolation.

- **Retinoic acid induced differentiation of ES cells**

Embryoid bodies were formed as described above. At day 3, aggregates were transferred to complete growth medium (see growth conditions for ES cells) lacking leukemia inhibitory factor (LIF) and supplemented with 0.1 μM All-*trans* Retinoic acid (Sigma) to

induce neuronal differentiation (Glaser and Brustle, 2005; Li et al., 1998). Aggregates were then cultured on cell culture dishes to allow attachment. Images were taken and RNA was isolated at day 10 for analysis of differentiation markers (Figure 6 in main text).

- **Hematopoietic differentiation, methylcellulose colony formation and FACs analysis**

For hematopoietic differentiation (Kyba et al., 2002) (Figure S5), EBs were obtained as described above (embyoid body formation) at 100 cells per drop in (IMDM (Invitrogen), 200microg/ml Holo-transferrin (Sigma), 1% pen/strep/glutamine, 15% serum (Stem Cell technologies), 50μg/ml ascorbic acid (sigma). On day six, EBs were collected and dissociated by adding 250μl of dissociation enzyme mix (500 mg Collagenase IV (Invitrogen), 1 gm Hyaluronidase (Sigma), 40,000U DNase (Sigma) in 50ml DMEM) and 1 ml PBS at 37°C for 20 minutes while occasionally swirling tube to mix the enzymes and EBs. After adding 8 ml enzyme-free dissociation buffer, mixtures were triturated using 5 ml pipette until EBs were fully dissociated and cells were collected by centrifugation. Cells were re-suspended at 50,000 cells/ml and infected with viral supernatant of a Hoxb4-GFP expressing retrovirus (kindly provided by K. Humphries and G. Sauvageau). Polybrene was added to a final concentration of 8 μg/ml. Cells plus virus was transferred to a 6-well plate at 2ml/well which was pre-plated with OP9 stromal cells (ATCC) at 25,000 cells per well. Plates were centrifuged at 900 rpm at room temperature for 60 minutes and incubated at 37°C 5% CO_2 . After 12-16 hours, suspension cells were collected and resuspended in fresh 10% IMDM+cytokines (100 ng/mL Flt3L, 100 ng/mL SCF, 40 ng/mL TPO, 40 ng/mL VEGF, 10 ng/ml IFN from Peprotech) to get rid of the viral supernatant and subsequently re-plated on the same plates. Cells were cultured for an additional 10 days with a medium change at day 7. Methylcellulose colony forming assay was performed as previously described (Wang et al., 2005). FACS analyses were performed on a Becton-Dickinson cell sorter. Antibodies used were APC labeled anti-IgM, anti-Gr1, and PE labeled anti-CD41 purchased from EBiosciences. APC labeled anti-mSSEA1 antibody was purchased from R&D systems.

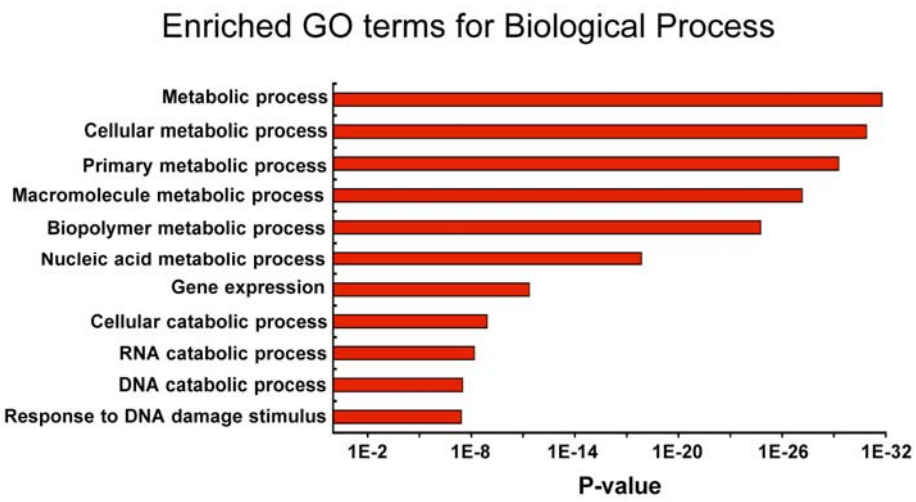

Figure S1. H2AZ is enriched at genes involved in cell metabolism in neural progenitors.
Gene ontology analysis of H2AZ-enriched genes and highly represented categories for biological process are shown. Ontology terms are represented on the y-axis and the p-value for enrichment of a bound gene relative to all genes represented on the microarray is shown for each category on the x-axis.

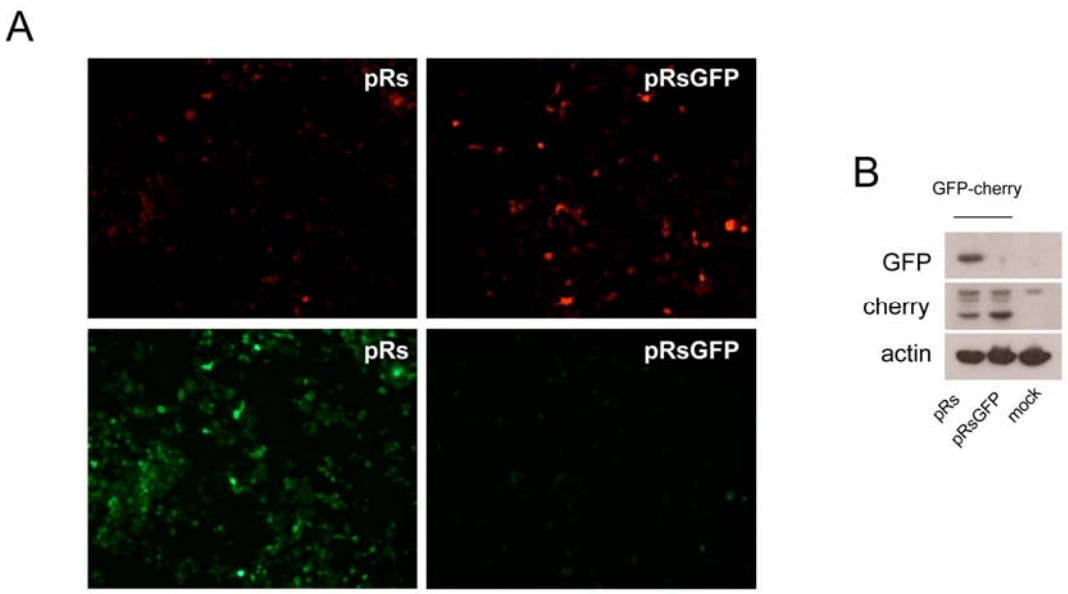

Figure S2. GFP control hairpin is expressed and functional.
(A) A lentiviral plasmid expressing GFP and cherry-Red with a pRetroSuper control vector (left panels) or pRetroSuper control vector expressing a GFP specific hairpin were co-transfected into 293 cells. Upper panels show cherry-red fluorescence and lower panels show GFP fluorescence of the same field. (B) Expression of the GFP hairpin from a pRetroSuper vector specifically reduces GFP levels. Immunoblot analysis of the indicated proteins of empty pRetroSuper-transduced control ES cell line (left lane) and GFP hairpin expressing control ES cell line (middle lane) both infected with lentivirus expressing GFP and cherry-Red from a bidirectional promoter. Right lane shows uninfected cells for background fluorescence.

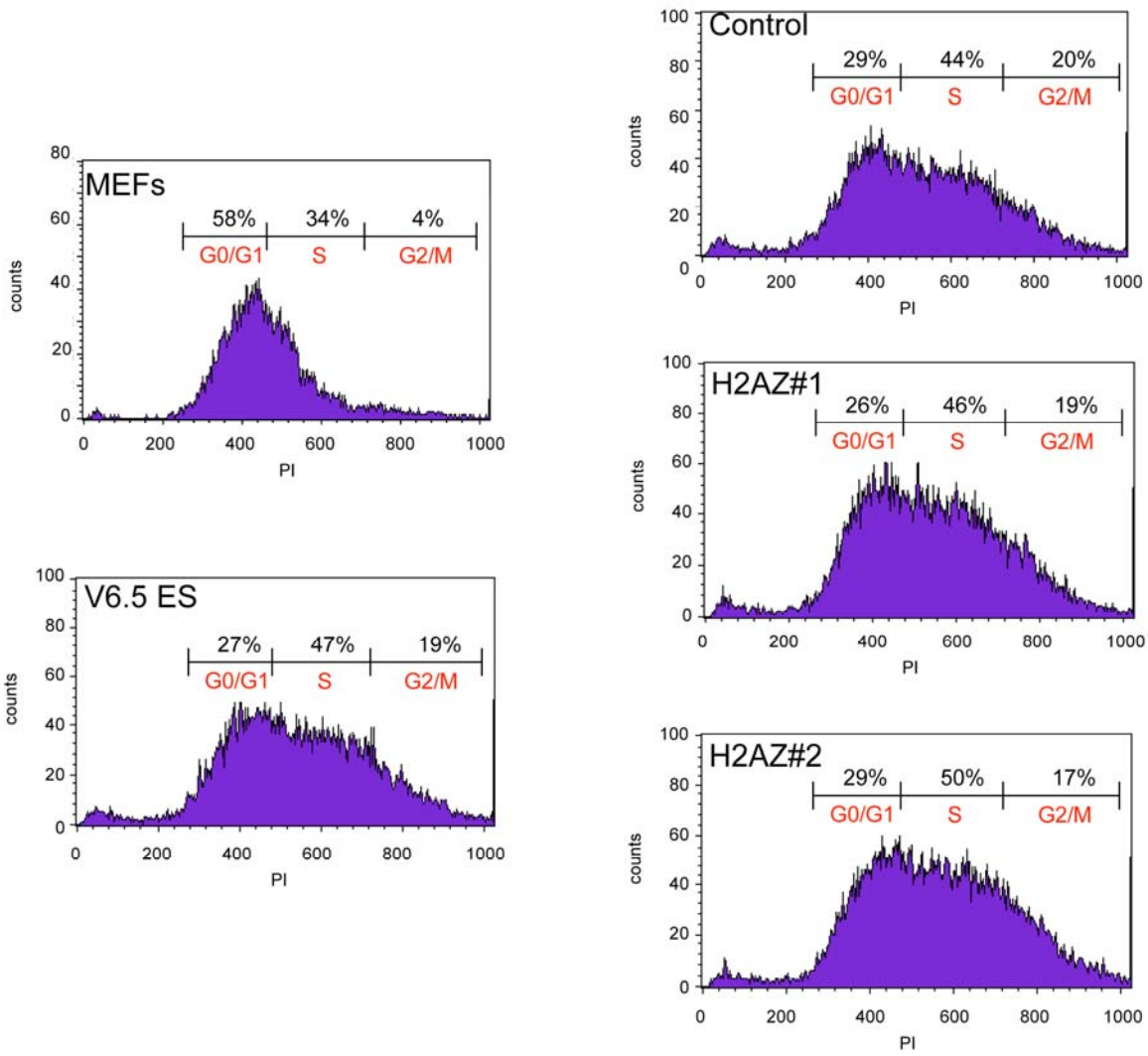

Figure S3. H2AZ-depleted cells display ES cell specific cell cycle kinetics.
Cell cycle analysis by measuring DNA content with FACS analysis. Control and H2AZ-deficient ES cells were permeabilized and stained with 7AAD for cell cycle analysis. MEF cells were used for comparison as ES cells and MEFs display markedly different cell cycle kinetics. The X-axis of the histogram represents the intensity of the 7AAD signal. The far left peak demonstrates cells in (G0/G1) state, the far right peak represents cells in (G2/M) state and the area between these peaks indicates cells in S phase. Percentages of cells found in different stages of cell cycles are indicated on the histogram graphs. One representative data set is shown out of three performed.

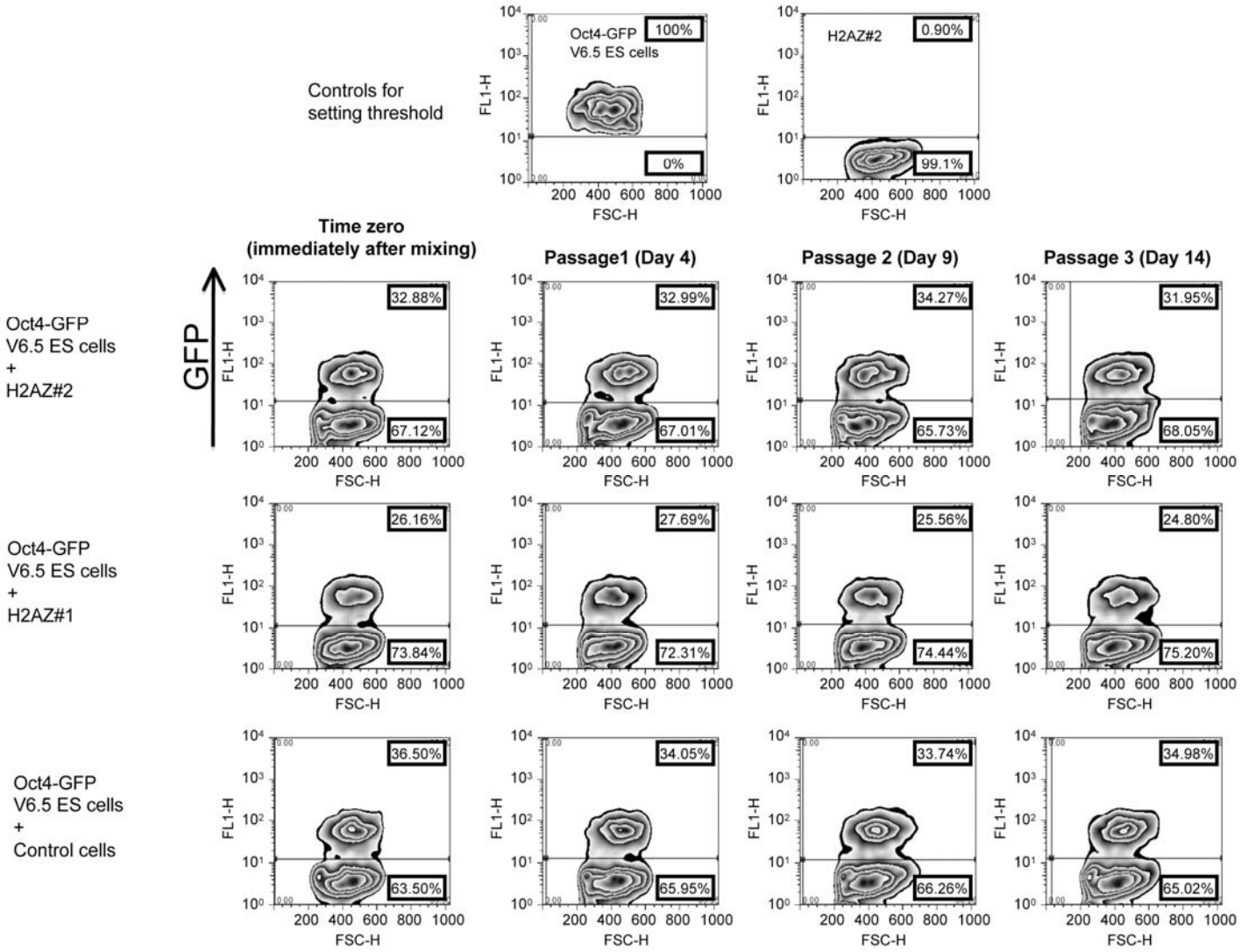

Figure S4. H2AZ-depleted cells display normal proliferation as compared to control ES cells.

FACS analysis for wild type ES cells expressing GFP from the Oct4 locus (upper left panel) mixed with non-GFP expressing cells as indicated on the left for each set of graphs. GFP expression is represented on the y-axis of the graphs (see upper two graphs for control). Cells were followed for 3 passages over 14 days (graphs from left to right passage 0-3) in order to detect small changes in proliferation for the indicated cell lines tested compared to wild type control cells.

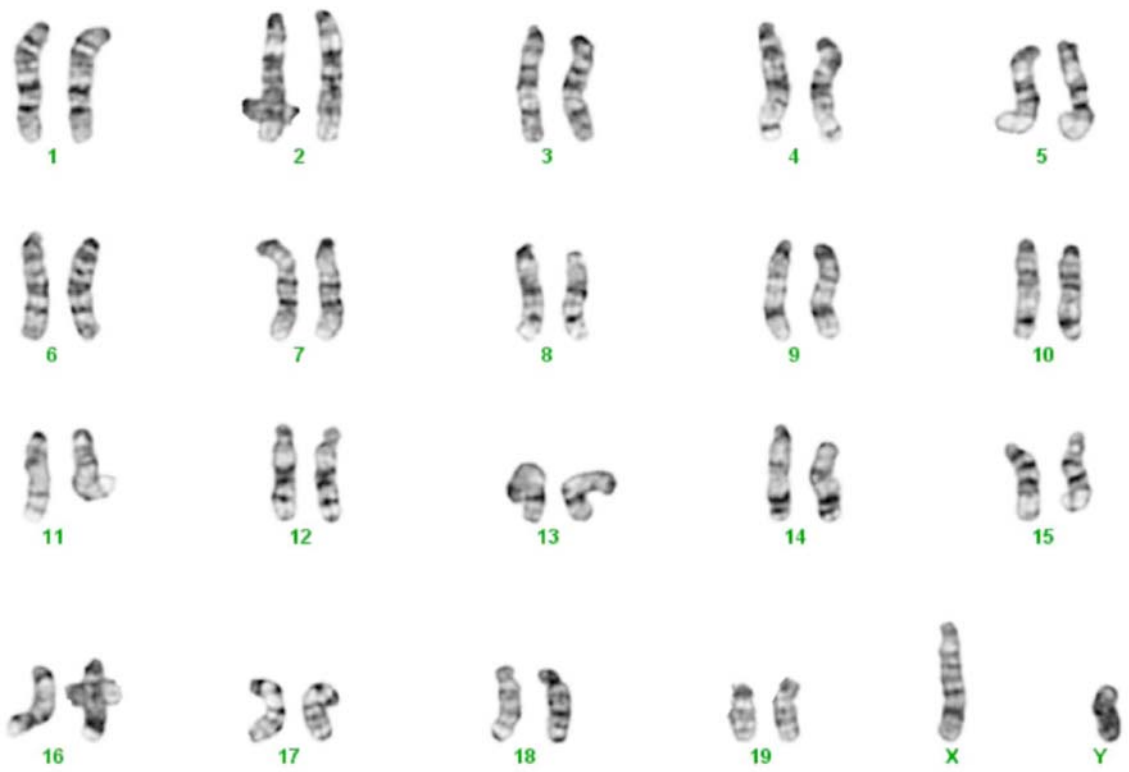

Figure S5. H2AZ-depleted ES cells display normal karyotype.
Representative cytogenetic analysis performed on twenty G-banded metaphase cells from passage 10 H2AZ-depleted ES cells (clone #2) shows normal karyotype (Cell Line Genetics).

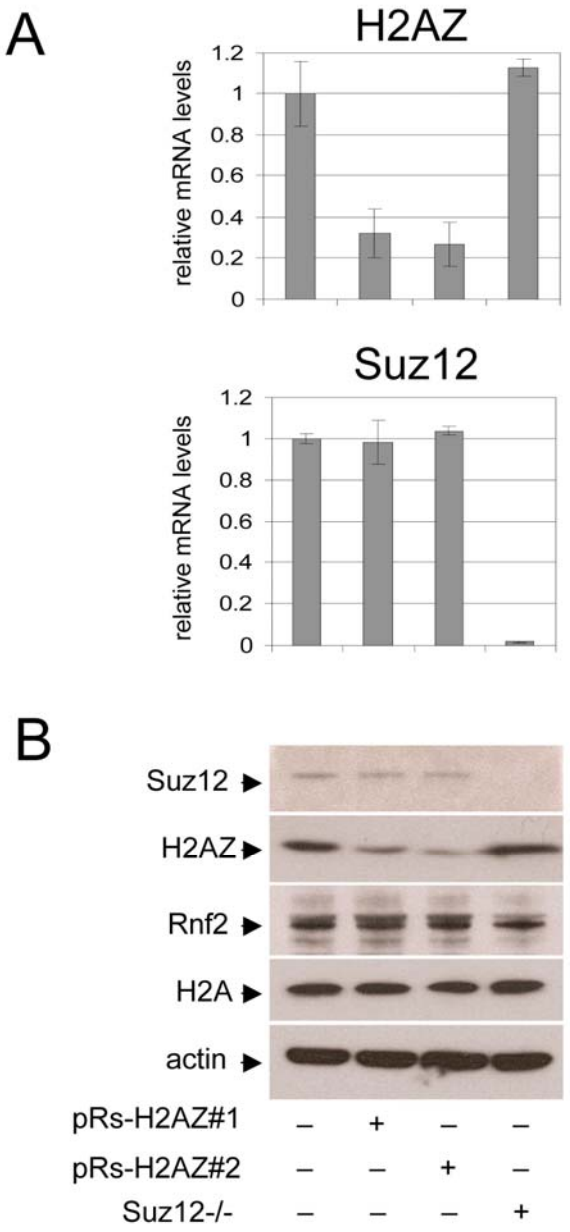

Figure S6. Expression of H2AZ, Rnf2/Ring1b, and Suz12 is not interdependent.
(A) Real Time PCR for H2AZ (upper graph) or Suz12 (lower graph) mRNA levels normalized to Gapdh in H2AZ-depleted (clone #2) and *suz12* null ES cells as compared to a control ES cell line expressing a non-specific hairpin. Reactions were performed in triplicate and error bars represent 2 standard deviations. Similar results were obtained in both H2AZ-depleted cell lines. (B) Immunoblot analysis for the indicated proteins in the control cell line expressing a non-specific hairpin (left lane), H2AZ-depleted (middle lanes) and *suz12* null ES cells (right lane).

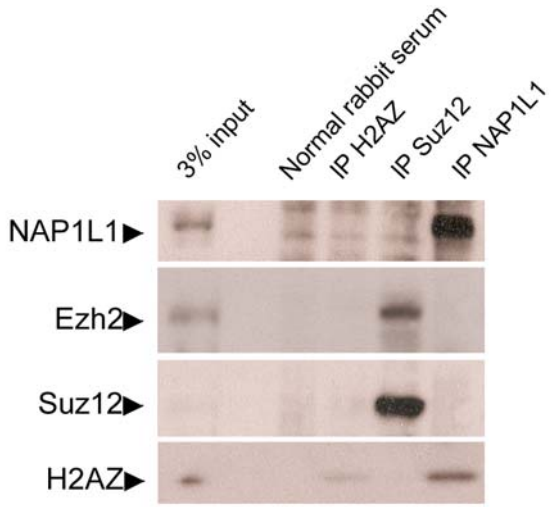

Figure S7. PRC2 does not co-immunoprecipitate H2AZ from ES cells
Immunoprecipitation (IP) was performed in lysates of wild type V6.5 ES cells and immunoprecipitated fractions were resolved by SDS-PAGE and immunoblotting. The left lane shows 3% of total cell extract used for the IP. Immunoblots show absence or presence of proteins in the immunoprecipitated lysate as indicated. Normal rabbit serum was used as a control for antibody specificity. Note that while Suz12 co-precipiates Ezh2, a component of PRC2, an interaction with H2AZ is not detected. Consistent with this, H2AZ does not IP the core PRC2 subunits. As further evidence, Nap1l1, the histone chaperone with roles in histone variant exchange, interacts with H2AZ, but not with PRC2 subunits.

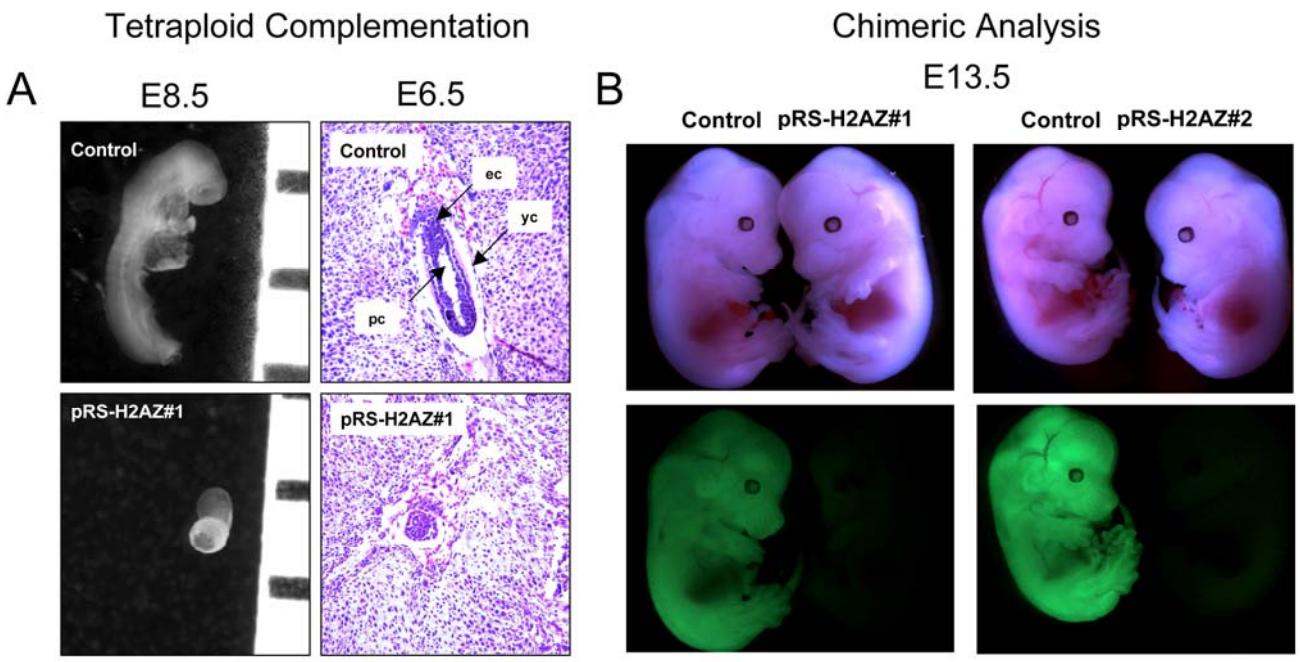

Figure S8. Developmental potential of H2AZ-depleted ES cells
(A) Tetraploid complementation embryos derived from control ES cells (upper panels) and H2AZ-depleted ES cells (lower panels). Left panels show formalin fixed embryos retrieved at dpc 8.5 (1 mm scale). Right panels show hematoxylin and eosin staining of sectioned deciduas containing embryos, obtained on dpc 6.5 from both H2AZ-deficient ES cell lines. Specified in the control section are: pc (pro-amniotic cavity), ec (ecto-placental cone) and yc (yolk cavity). (B) Developmental defects in H2AZ-deficient embryos are cell autonomous. Bright field images of embryos obtained at dpc 13.5 derived from wild-type B6D2F1 blastocysts injected with either control (left) or H2AZ-depleted ES cell lines (right). The bottom panel shows GFP expression by fluorescent microscopy to visualize contribution of the GFP-labeled donor cells to embryo development.

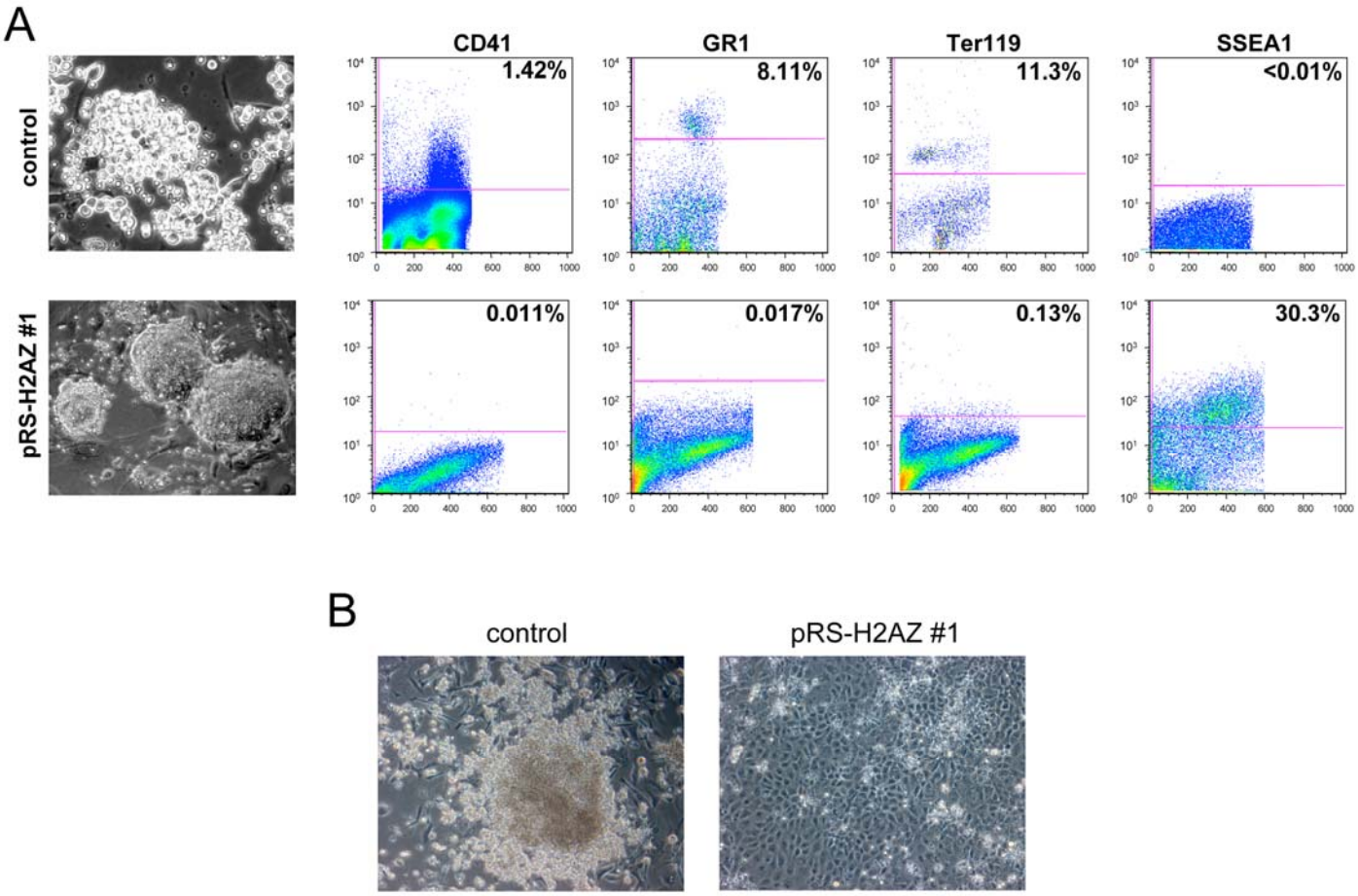

Figure S9. H2AZ is required for hematopoietic differentiation.
(A) Control cells display morphological characteristics typical of hematopoietic differentatiation (upper left) whereas H2AZ-depleted ES cells fail to show any such changes (bottom left). FACS analysis of control cells (upper right) and H2AZ-depleted cells (bottom right) for the indicated markers following hematopoetic differentiation. The y-axis represents the indicated marker tested, the x-axis represents forward scatter. (B) Methylcellulose colony forming assay showing control (left panel) and H2AZ-depleted ES cells (right panel). Similar results were obtained for a second H2AZ-depleted ES cell line (not shown).

Supplemental References

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., *et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., *et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

Brummelkamp, T.R., Bernards, R., and Agami, R. (2002). Stable suppression of tumorigenicity by virus-mediated RNA interference. Cancer Cell *2*, 243-247.

Conti, L., Pollard, S.M., Gorba, T., Reitano, E., Toselli, M., Biella, G., Sun, Y., Sanzone, S., Ying, Q.L., Cattaneo, E., *et al.* (2005). Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. PLoS Biol *3*, e283.

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol *4*, P3.

Eggan, K., Rode, A., Jentsch, I., Samuel, C., Hennek, T., Tintrup, H., Zevnik, B., Erwin, J., Loring, J., Jackson-Grusby, L., *et al.* (2002). Male and female mice derived from the same embryonic stem cell clone by tetraploid embryo complementation. Nat Biotechnol *20*, 455-459.

Faast, R., Thonglairoam, V., Schulz, T.C., Beall, J., Wells, J.R., Taylor, H., Matthaei, K., Rathjen, P.D., Tremethick, D.J., Lyons, I. (2001) Histone variant H2A.Z is required for early mammalian development. Curr Biol. *11*, 1183-1187.

Glaser, T., and Brustle, O. (2005). Retinoic acid induction of ES-cell-derived neurons: the radial glia connection. Trends Neurosci *28*, 397-400.

Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., *et al.* (2001). Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat Biotechnol *19*, 342-347.

Kyba, M., Perlingeiro, R.C., and Daley, G.Q. (2002). HoxB4 confers definitive lymphoid-myeloid engraftment potential on embryonic stem cell and yolk sac hematopoietic progenitors. Cell *109*, 29-37.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., *et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.

Li, M., Pevny, L., Lovell-Badge, R., and Smith, A. (1998). Generation of purified neural precursors from embryonic stem cells by lineage selection. Curr Biol *8*, 971-974.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402-408.

Lois, C., Hong, E.J., Pease, S., Brown, E.J., and Baltimore, D. (2002). Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. Science *295*, 868-872.

O'Carroll, D., Erhardt, S., Pagani, M., Barton, S.C., Surani, M.A., Jenuwein, T. (2001). The polycomb-group gene Ezh2 is required for early mouse development. Mol Cell Biol. *21*, 4330-4336.

Odom, D.T., Zizlsperger, N., Gordon, D.B., Bell, G.W., Rinaldi, N.J., Murray, H.L., Volkert, T.L., Schreiber, J., Rolfe, P.A., Gifford, D.K., *et al.* (2004). Control of pancreas and liver gene expression by HNF transcription factors. Science *303*, 1378-1381.

Okabe, S., Forsberg-Nilsson, K., Spiro, A.C., Segal, M., and McKay, R.D. (1996). Development of neuronal precursor cells and functional postmitotic neurons from embryonic stem cells in vitro. Mech Dev *59*, 89-102.

Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M., and Helin, K. (2007). The polycomb group protein Suz12 is required for embryonic stem cell differentiation. Mol Cell Biol *27*, 3769-3779.

Squazzo, S.L., O'Geen, H., Komashko, V.M., Krig, S.R., Jin, V.X., Jang, S.W., Margueron, R., Reinberg, D., Green, R., and Farnham, P.J. (2006). Suz12 binds to silenced regions of the genome in a cell-type-specific manner. Genome Res *16*, 890-900.

Wang, Y., Yates, F., Naveiras, O., Ernst, P., and Daley, G.Q. (2005). Embryonic stem cell-derived hematopoietic stem cells. Proc Natl Acad Sci U S A *102*, 19081-19086.

Wang, Z., and Jaenisch, R. (2004). At most three ES cells contribute to the somatic lineages of chimeric mice and of mice produced by ES-tetraploid complementation. Dev Biol *275*, 192-201.