Supporting Information

Ji et al. 10.1073/pnas.1502971112

SI Materials and Methods

ChIP-MS and ChIP-Seq. mESCs were grown in 15-cm plates with 20 mL ES medium to full confluence. Two milliliters of 11% (wt/vol) formaldehyde was added to the 20-mL ES medium and was incubated at room temperature for 10 min to fix the cells. One milliliter of 2.5 M glycine was added to each plate and incubated for 5 min to quench the formaldehyde. The cells were washed twice with PBS. The cells were pelleted at 1,300 × *g* for 5 min at 4 °C. Then 4×10^7 cells were collected in each tube. The cells were lysed gently with 1 mL of ice-cold Nonidet P-40 lysis buffer containing protease inhibitor on ice for 5 min (all buffer recipes are given below). The cell lysate was layered on top of 2.5 volumes of sucrose cushion made up of 24% (wt/vol) sucrose in Nonidet P-40 lysis buffer. This sample was centrifuged at 18,000 × *g* for 10 min at 4 °C to isolate the nuclei pellet (the supernatant represented the cytoplasmic fraction). The nuclei pellet was washed once with PBS/1 mM EDTA. The nuclei pellet was resuspended gently with 0.5 mL glycerol buffer followed by the addition of an equal volume of nuclei lysis buffer on ice for 2 min. The sample was centrifuged at 16,000 × *g* for 2 min at 4 °C to isolate the chromatin pellet (the supernatant represented the nuclear soluble fraction). The chromatin pellet was washed twice with PBS/1 mM EDTA. The chromatin pellet was stored at −80 °C (usually 6–8 × 10^8 cells/ChIP-MS).

The chromatin pellet was resuspended in 3.5 mL of sonication buffer containing protease inhibitors. Sonication was performed in cycles of 30 s on followed by 1 min off for a total of 5 min (21–24 watts). The sonicated chromatin was split into Eppendorf tubes and spun down at 18,000 × *g* for 15 min at 4 °C to collect the chromatin. The supernatant fractions were combined, and 50 μL was saved as input. To check the sonication efficiency, the sonicated sample was reverse crosslinked at 65 °C overnight. After treatment with RNase A (R4642; Sigma) followed by treatment with proteinase K (catalog no. 25530-049; Life Technologies) (working concentration, 0.4 mg/mL), the DNA was purified using phenol-chloroform extraction (100 mL; catalog no. P3803; Sigma). The sonicated DNA sizes were analyzed by agarose gel electrophoresis. The majority of the sonicated DNA fragments were sheared to a size of around 200–600 bp.

Protein G Dynabeads (catalog no.10004D; Life Technologies) were washed and incubated at room temperature for 5 min in sonication buffer and then were washed twice with sonication buffer. The washed Dynabeads were added to the soluble chromatin (1:50) with the antibodies (20 μL beads per 1.5 μg antibody) and were incubated in a cold room overnight. The magnetic Dynabeads were pelleted by placing the tubes in a magnetic rack and were washed once with sonication buffer. Using the magnetic rack, the beads then were washed with high-salt wash buffer, then with LiCl wash buffer, and then with Tris-EDTA (TE)-NaCl buffer. The sample was spun (18,000 × *g* for 1 min) to remove any supernatant remaining after the last washing. Then the beads were resuspended in ChIP-MS elution buffer. At this point, the sample was divided into two parts. One-tenth of the sample was set-aside for DNA sequencing, and the remaining 9/10ths of the sample was used for MS. The sequencing sample was incubated at 65 °C overnight. Then the DNA was precipitated (following the DNA purification protocol described above) for sequencing. Input DNA was used as a control for ChIP enrichment. Protein was prepared from the remaining sample by resuspending with SDS loading buffer containing 100 mM DTT followed by incubation at 100 °C for 12 min. The sample was centrifuged, vortexed, and then incubated at 100 °C for another 12 min. The proteins for each sample were separated using SDS/PAGE gel electrophoresis and subsequently were stained with Coomassie blue for downstream MS analysis. For each preparation, distinct bands were visible in the SDS/PAGE gel, suggesting enrichment of specific proteins from the cellular extract. Preparations using antibodies against different histone modifications showed distinct sets of bands, indicating that different proteins were enriched in different ChIPs. Finally, we noted a strong signal where we expect IgG to migrate; we interpret this signal as Ig leaching from the beads used during the ChIP. To compensate for the potential saturation of the mass spectrometer by IgG peptides, the IgG band was cut out specifically and run independently from the other gel bands. Thus the proteins with the same molecular weight as IgG may not be detected as well as the other proteins because of saturation of the mass spectrometer at this size range.

The Nonidet P-40 lysis buffer contained 10 mM Tris·HCl (pH 7.5), 150 mM NaCl, and 0.05% Nonidet P-40. The glycerol buffer contained 20 mM Tris·HCl (pH 7.9), 75 mM NaCl, 0.5 mM EDTA, 0.85 mM DTT, and 50% (vol/vol) glycerol. The nuclei lysis buffer contained 10 mM Hepes (pH 7.6), 1 mM DTT, 7.5 mM $MgCl_2$, 0.2 mM EDTA, 0.3 M NaCl, 1 M urea, and 1% Nonidet P-40. The sonication buffer contained 20 mM Tris·HCl (pH 8.0), 150 mM NaCl, 2 mM EDTA (pH 8.0), 0.1% SDS, and 1% Triton X-100. The high-salt wash buffer contained 20 mM Tris·HCl (pH 8.0), 500 mM NaCl, 2 mM EDTA, 0.1% SDS, and 1% Triton X-100. The LiCl wash buffer contained 10 mM Tris·HCl (pH 8.0), 250 mM LiCl, 1 mM EDTA, and 1% Nonidet P-40. The TE-NaCl wash buffer contained 1 mM EDTA, 10 mM Tris·HCl (pH 8.0), and 50 mM NaCl. The ChIP-seq elution buffer contained 50 mM Tris·HCl (pH 8.0), 10 mM EDTA, and 1% SDS. All buffers were prepared with protease inhibitors.

Validation by ChIP-Western Blot and ChIP-qPCR. For ChIP-Western blot analysis, 2 × 10^8 mESCs were used for each histone mark preparation. The proteins were isolated using the protocol as described above for ChIP-MS but were analyzed by Western blot. For ChIP-qPCR analysis, 1 × 10^8 mESCs were used for each reaction. The sonication was performed as in the ChIP-MS, and 1% of the starting material was kept for input. The precleared sonicated chromatin was incubated with antibodies (3 μg antibodies per 1 mL sonicated chromatin) overnight in the cold room. The next day, protein G Dynabeads were added into the chromatin solution (30 μL of beads per 1 mL sonicated chromatin) and were incubated in the cold room for another 4 h. The beads were washed (for 10 min each washing) once with sonication buffer, once with high-salt wash buffer, and three times with TE buffer. The DNA was eluted two times with 150 μL of elution buffer at 65 °C for 10 min each time. The eluates were pooled together and heated at 65 °C overnight to reverse the formaldehyde crosslinks. For the input sample, 300 μL of elution buffer was added into the 1% aliquot of the starting material and then was heated at 65 °C overnight to reverse the formaldehyde crosslinks. DNA fragments were purified with the QIAquick PCR Purification Kit to yield 100 μL of DNA solution. Primers targeting gene enhancers, promoters, and gene bodies of *Zfml* and *Agfg1* were used for ChIP-qPCR analysis (Fig. S2 and Table S2). Primers targeting a gene desert region were used as a negative control for ChIP-qPCR.

Sample Preparation for MS. Squares (~2 mm) were cut from the polyacrylamide gel and were washed overnight in 50% (vol/vol) methanol. These samples were washed for 2 h with 47.5/47.5/5%

(vol/vol/vol) methanol/water/acetic acid and then were dehydrated in acetonitrile and dried using a SpeedVac (Thermo Scientific). Disulfide bonds were reduced and alkylated by the addition of 30 μL of 10 mM DTT in 100 mM ammonium bicarbonate for 30 min. The DTT solution was removed, and 100 mM iodoacetamide in 100 mM ammonium bicarbonate was added for 30 min to alkylate the free cysteine residues to form carbamidomethyl cysteine. The samples then were washed with acetonitrile, 100 mM ammonium bicarbonate, and acetonitrile and subsequently were dried in a SpeedVac. The dried gel bands were treated with 300 ng of trypsin enzyme in 50-mM ammonium bicarbonate for 10 min on ice. Ammonium bicarbonate volumes were adjusted according to the relative volumes of acrylamide to rehydrate the gel pieces. The samples were digested overnight at 37 °C. Peptides were extracted by adding 50 μL (or more, if needed to generate supernatant) of 50-mM ammonium bicarbonate and gentle shaking for 10 min. The supernatant was collected into a 0.5-mL conical autosampler vial. Acetonitrile/water/formic acid 47.5/47.5/5% (vol/vol/vol) was added twice with gentle shaking for 10 min, and the supernatant was added to the 0.5-mL autosampler vial. The organic solvent was removed from the samples, and the volumes were reduced to 15 μL using a SpeedVacfor the downstream analyses.

Chromatographic Separations. The digested extracts were analyzed by running the Waters NanoAcquity HPLC in reversed phase, an autosampler, and a Thermo Fisher Orbitrap Elite mass spectrometer using a nano flow configuration. Peptides were trapped and washed using a 20 mm × 180 μm column packed with 5 μm Symmetry C18 material (Waters) using a flow rate of 15 μm/min for 2 min. The peptides then were eluted onto a self-packed analytical column packed with 3 μm Jupiter C18 material (Phenomenex) in fritted 10 cm × 75 μm fused silica tubing pulled to a 5-μm tip. The flow was an isocratic gradient starting with 1% buffer A (1% formic acid in water) for 1 min at 250 nL/min with increasing concentrations of buffer B (1% formic acid in acetonitrile) up to 15% buffer B at 14 min, 27% buffer B at 21 min, 40% buffer B at 24.5 min, and 85% buffer B at 25.5 min. The column was washed with high percent buffer B and then was reequilibrated between analytical runs for a total time of ∼37 min per cycle.

MS. The dependent data acquisition mode was used on the Thermo Fisher Orbitrap Elite mass spectrometer in which the five most abundant peptides detected in full-scan mode were subjected to daughter ion fragmentation. A list of parent ions was collected as an exclusion list to increase the number of peptides analyzed during the chromatographic run.

MS Data Analysis. SEQUEST algorithms were used to search a species-specific database generated from a subset of the Reference Sequence (RefSeq) database from the National Center for Biotechnology Information to identify peptides in the MS data. The results from SEQUEST were input into the Scaffold Proteome Software and were validated using the Protein Prophet algorithm applying a minimum 95% peptide probability and a 99% protein probability. A cutoff minimum of two total peptides was used to consider an identification positive. The protein list and corresponding total spectrum counts were visualized by the Scaffold Proteome software. We used label-free quantitation (LFQ) counts normalized to the total spectrum counts generated in Scaffold for downstream analyses.

Selection Criteria for Chromatin-Associated Protein Candidates. First, the peptide counts from all isoforms of a protein were summed to generate a single value for each protein. Second, proteins were identified as chromatin-associated protein candidates if they had two or more total peptide counts in a preparation and at least twice as many LFQ counts in that histone mark preparation as in the control IgG and input samples. Third, documented ribosomal, mitochondrial, secreted, plasma membrane, translational machinery, and tRNA-related proteins (58 proteins in total) were removed from the detected protein candidate lists (1, 2).

Analysis of Protein Candidates. Protein peptide ranking analysis was performed in Excel (Fig. S1*B*). In all samples, the proteins were ranked according to their counts in the input sample. The presence of peptide counts for a protein in each preparation was marked blue in the corresponding protein position.

Three categories of chromatin-associated proteins were defined. The known category contained proteins reported to associate with chromatin in the UniProt, GeneCards, or PubMed databases. The implicated category contained proteins that were not reported to associate with chromatin but that contain DNA-binding domains or chromatin-binding domains or belong to a protein complex known to associate with chromatin. The lists of proteins that contain DNA-binding domains or chromatin-binding domains were compiled from the transcription factor prediction database and other studies (3–5). The novel category contained proteins not previously known or implicated to associate with chromatin. Disease-associated proteins were identified in the Online Mendelian Inheritance in Man database, Catalogue of Somatic Mutations in Cancer database, and in additional references (6–8).

ChIP-Seq Data Analysis. Reads from mESC ChIP-seq were aligned to the mm9 reference genome using Bowtie 0.12.8 or 1.0.1 with parameters –k 1 –m 1 –S –best –n 2. WIG files for visualization of ChIP-seq data at individual loci were created using MACS with parameters –w –S –space = 50 –nomodel –shiftsize = 200. Metagenes for multiple types of RefSeq regions using bamToGFF (https://github.com/BradnerLab/pipeline) with parameters –e 200 –f 1 –r, and displayed lines represent the mean of all genes interrogated. Metagene analysis for gene bodies were created in three parts: 50 bins split 2 kb upstream of the transcription start site, 150 bins split a normalized gene body, and 50 bins split 2 kb downstream from the transcription termination site. Enhancers were downloaded from a previous study (9). Briefly, peaks for OCT4, SOX2, and NANOG (OSN) were identified using MACS, and OSN enhancers were defined as the intersection of these regions. Metagene analyses at enhancers were created by dividing a 4-kb region centered on the center of OSN enhancers into 150 bins.

1. Wang J, et al. (2006) A protein interaction network for pluripotency of embryonic stem cells. *Nature* 444(7117):364–368.
2. Antão JM, Mason JM, Déjardin J, Kingston RE (2012) Protein landscape at Drosophila melanogaster telomere-associated sequence repeats. *Mol Cell Biol* 32(12): 2170–2182.
3. Wilson D, Charoensawan V, Kummerfeld SK, Teichmann SA (2008) DBD—taxonomically broad transcription factor predictions: New content and functionality. *Nucleic Acids Res* 36(Database issue):D88–D92.
4. Heinäniemi M, et al. (2013) Gene-pair expression signatures reveal lineage control. *Nat Methods* 10(6):577–583.
5. Letunic I, Doerks T, Bork P (2015) SMART: Recent updates, new developments and status in 2015. *Nucleic Acids Res* 43(Database issue):D257–D260.
6. Lee TI, Young RA (2013) Transcriptional regulation and its misregulation in disease. *Cell* 152(6):1237–1251.
7. Forbes SA, et al. (2011) COSMIC: Mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res* 39(Database issue):D945–D950.
8. Lawrence MS, et al. (2014) Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505(7484):495–501.
9. Whyte WA, et al. (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153(2):307–319.

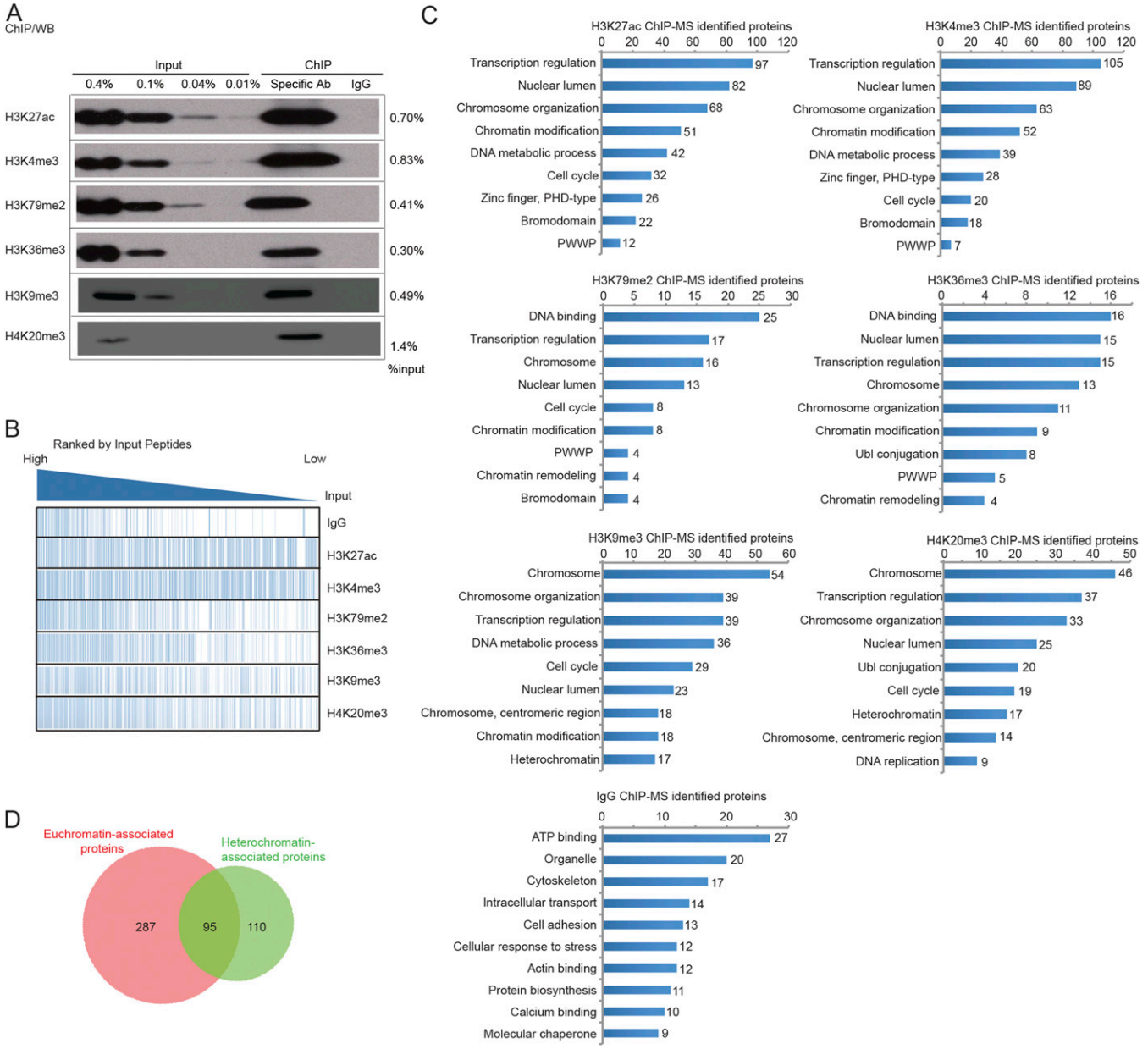

Fig. S1. Histone mark ChIP yields measured by Western blot, relative comparison of protein detection across preparations, and Gene Ontology analysis for the proteins identified by chromatin proteomic profiling. (*A*) Western blot shows the yield of the histone mark detection for each histone mark ChIP. Between 0.3–1.4% of the starting material (input) was captured for each of the histone mark ChIPs. IgG ChIP serves as the negative control. (*B*) The distribution of detected proteins in the IgG, H3K27ac, H3K4me3, H3K79me2, H3K36me3, H3K9me3, and H4K20me3 ChIP-MS preparations. In all samples, the proteins were ranked according to their peptide counts in the input sample. The presence of peptide counts for a protein in each preparation was marked with a blue line in the corresponding protein position. Proteins with lower abundance in the input were detected more often in the histone mark preparations than in the IgG preparation. (*C*) Gene Ontology analyses of proteins identified in the histone mark preparations and the IgG preparation. Gene Ontology categories enriched among the proteins identified by chromatin proteomic profiling differed from the categories enriched among the proteins in the IgG ChIP-MS preparation. Numbers on the right ends of bars indicate the number of identified proteins in each GO category. Among the biological process terms associated with each preparation, all terms displayed have a *P* value of $P \leq 5.4 \times 10^{-15}$ for H3K27ac, $P \leq 8.8 \times 10^{-6}$ for H3K4me3, $P \leq 4.7 \times 10^{-3}$ for H3K79me2, $P \leq 2.0 \times 10^{-4}$ for H3K36me3, $P \leq 2.4 \times 10^{-7}$ for H3K9me3, $P \leq 3.0 \times 10^{-7}$ for H4K20me3, and $P \leq 1.1 \times 10^{-2}$ for IgG. (*D*) Venn diagram analysis of euchromatin-associated proteins identified in the H3K27ac, H3K4me3, H3K79me2, and H3K36me3 preparations and heterochromatin-associated proteins identified in the H3K9me3 and H4K20me3 preparations.

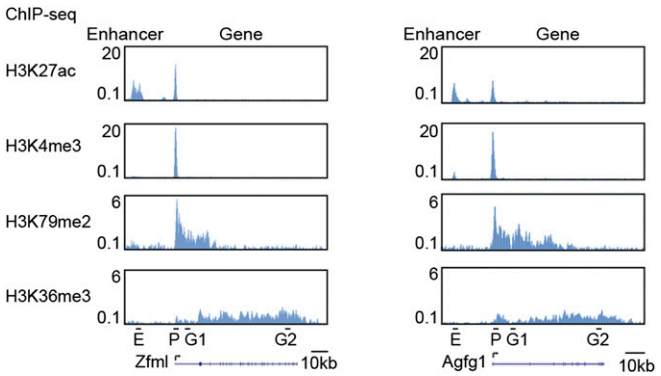

Fig. S2. Positions of ChIP-qPCR primers relative to the ChIP-seq signals for histone modifications around the *Zfml* and *Agfg1* gene loci. Shown are gene track illustrations of H3K27ac, H3K4me3, H3K79me2, and H3K36me3 ChIP-seq signals around the *Zfml* (*Left*) and *Agfg1* (*Right*) gene loci. The relative positions of the enhancer and gene are shown above the panel. The histone modification used for ChIP is identified at the left of each track. The scale for the ChIP-seq signal in reads per million is shown on the *y* axis of each track. The locations of primers targeting enhancers (E), promoters (P), and gene body regions (G1 and G2) are labeled immediately below the panels. The gene model and the scale bar are shown along the *x* axis at the bottom of the panels.

Other Supporting Information Files

Table S1 (XLS)
Table S2 (XLS)
Table S3 (XLS)