

Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions

Xiong Ji^{a,1}, Daniel B. Dadon^{a,b,1}, Brian J. Abraham^a, Tong Ihn Lee^a, Rudolf Jaenisch^{a,b}, James E. Bradner^{c,d}, and Richard A. Young^{a,b,2}

^aWhitehead Institute for Biomedical Research, Cambridge, MA 02142; ^bDepartment of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139; ^cDepartment of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02115; and ^dDepartment of Medicine, Harvard Medical School, Boston, MA 02115

Contributed by Richard A. Young, February 13, 2015 (sent for review January 7, 2015; reviewed by Jerry L. Workman)

More than a thousand proteins are thought to contribute to mammalian chromatin and its regulation, but our understanding of the genomic occupancy and function of most of these proteins is limited. Here we describe an approach, which we call “chromatin proteomic profiling,” to identify proteins associated with genomic regions marked by specifically modified histones. We used ChIP-MS to identify proteins associated with genomic regions marked by histones modified at specific lysine residues, including H3K27ac, H3K4me3, H3K79me2, H3K36me3, H3K9me3, and H4K20me3, in ES cells. We identified 332 known and 114 novel proteins associated with these histone-marked genomic segments. Many of the novel candidates have been implicated in various diseases, and their chromatin association may provide clues to disease mechanisms. More than 100 histone modifications have been described, so similar chromatin proteomic profiling studies should prove to be valuable for identifying many additional chromatin-associated proteins in a broad spectrum of cell types.

chromatin | genomics | proteomics

There are more than 1,000 transcription factors, cofactors, and chromatin regulators encoded in the mammalian genome, but we have limited understanding of the genomic occupancy and function of most of these (1–4). Understanding how these proteins interact with specific active and repressed portions of the genome would provide clues to their functions in global gene control, but limitations inherent in widely used genomic and proteomic technologies make acquiring this information laborious and expensive. Chromatin immunoprecipitation coupled to sequencing (ChIP-seq) can reveal the sites that a specific protein occupies in the genome (5–7) but is laborious and is limited by the availability of antibodies specific to candidate genome-binding proteins. Mass spectrometry (MS) can identify large populations of proteins present in specific preparations (8–10) but does not reveal how these proteins occupy specific portions of the genome. A recently developed approach combined ChIP with MS (ChIP-MS) to identify protein complexes that are associated with other proteins known to occupy sites in the genome (11–13). Here we adapt this approach to profile the proteins associated with chromatin containing specific histone modifications across the genome of mouse embryonic stem cells (mESCs). These chromatin modifications mark regions of the genome where specific transcriptional activities occur, thus implicating the associated proteins in these activities.

Results

Chromatin Proteomic Profiling Identifies Proteins Associated with Histone-Marked Regions of Chromatin. We used ChIP-MS to identify proteins associated with specific euchromatic histone modifications, including H3K27ac, H3K4me3, H3K79me2, and H3K36me3, and heterochromatic histone modifications, including H3K9me3 and H4K20me3, in mESCs (Fig. 1). Active enhancers are occupied by nucleosomes with H3K27ac; active promoters are occupied by nucleosomes with H3K27ac and H3K4me3; and portions of transcribed gene bodies are occupied by nucleosomes

with H3K79me2 and H3K36me3 (14, 15). In contrast, relatively silent genes are occupied by nucleosomes with H3K9me3 and H4K20me3 (16–18). We hypothesized that ChIP-MS, when used with antibodies specific to these chromatin modifications, would create a catalog of candidate proteins associated with specifically marked chromatin in vivo and that ChIP-seq analysis of the same material could confirm that the correct portions of the genome had been enriched by the procedure (Fig. 1A).

We first carried out ChIP of crosslinked euchromatic fractions using antibodies that recognize histone H3K27ac, H3K4me3, H3K79me2, and H3K36me3 and subjected part of this material to DNA sequence analysis to confirm that previously established patterns of genome occupancy were obtained. The results, when analyzed for individual genes (Fig. 1B) or the population of genes (Fig. 1C), were consistent with the well-established locations of nucleosomes with these modifications (14, 15). We then subjected the rest of this material to MS to identify proteins that were enriched in these ChIP preparations relative to proteins found in the initial chromatin preparation (input) and in a control preparation (IgG). Four lines of evidence suggest that selective enrichment of specific chromatin-associated proteins was achieved: (i) ChIP-Western blot analysis indicated that modified histones were immunoprecipitated with specific antibodies (Fig. S1A); (ii) ChIP-MS preparations of modified histones contained proteins that were present at low abundance in the initial chromatin preparation (input), indicating that the identified proteins were not simply reflective of cellular abundance (Fig. S1B); (iii) modified histone ChIP-MS preparations contained proteins

Significance

More than a thousand proteins are thought to contribute to mammalian chromatin and its regulation, but our understanding of the genomic occupancy and function of most of these proteins is limited. We have used a chromatin proteomic profiling approach to produce a catalogue of proteins associated with genomic regions whose chromatin is marked by specific modified histones. A substantial number of the newly identified proteins are associated with human disease. Future chromatin proteomic profiling studies should prove valuable for identifying additional chromatin-associated proteins in a broad spectrum of cell types.

Author contributions: X.J., D.B.D., J.E.B., and R.A.Y. designed research; X.J. and D.B.D. performed research; X.J. and D.B.D. contributed new reagents/analytic tools; X.J., D.B.D., and B.J.A. analyzed data; and X.J., D.B.D., T.I.L., R.J., J.E.B., and R.A.Y. wrote the paper.

Reviewers included: J.L.W., Stowers Institute for Medical Research.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE62380).

¹X.J. and D.B.D. contributed equally to this work.

²To whom correspondence should be addressed. Email: young@wi.mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1502971112/-DCSupplemental.

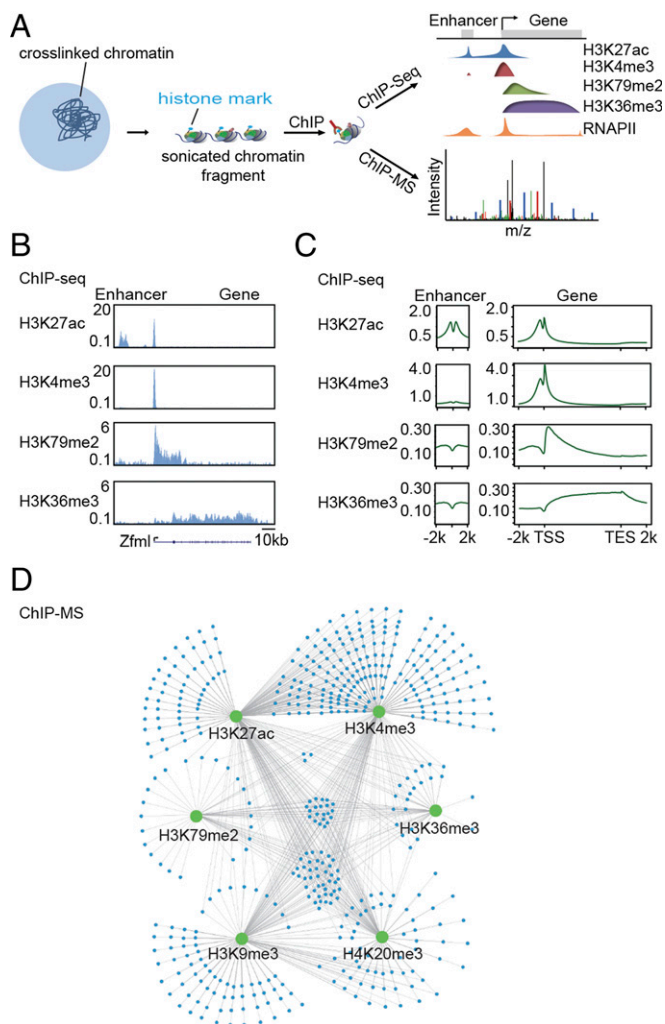


Fig. 1. Chromatin proteomic profiling identifies proteins associated with histone-marked regions of chromatin. (A) Schematic overview of the chromatin proteomic profiling approach. (B) Gene track illustration of H3K27ac, H3K4me3, H3K79me2, and H3K36me3 ChIP-seq signals (in reads per million) at the *Zfml* gene locus. (C) Metagene analysis of H3K27ac, H3K4me3, H3K79me2, and H3K36me3 ChIP-seq showing the average signal (in reads per million) for these modifications at enhancers and genes. (D) Distribution of proteins that were enriched in the H3K27ac, H3K4me3, H3K79me2, H3K36me3, H3K9me3, and H4K20me3 ChIP-MS preparations relative to the IgG control and input samples. Green nodes represent different histone mark preparations. Blue nodes represent identified proteins.

whose functions were enriched in relevant Gene Ontology categories, including transcription regulation and chromatin organization, which differed considerably from the categories enriched among the proteins in the IgG ChIP-MS preparation (Fig. S1C); and (iv) many of the proteins identified in these preparations have been shown previously to associate with chromatin containing nucleosomes with these modified histones, as described below.

We then carried out ChIP of crosslinked heterochromatic fractions using antibodies that recognize histone H3K9me3 and H4K20me3 and subjected this material to MS to identify proteins that were enriched in these ChIP preparations, which were expected to be substantially different from the euchromatic preparations. Indeed, the majority of the proteins identified in the heterochromatic histone ChIP-MS preparations were not found in the euchromatic histone H3K27ac, H3K4me3, H3K79me2, and H3K36me3 ChIP-MS preparations (Fig. S1D). The extent of

overlap of proteins identified in all the preparations is shown in Fig. 1D and is summarized in Table S1, which largely reflects the overlap of histone modifications across the genome (Fig. 1A).

The proteins identified by chromatin proteomic profiling were grouped into three categories, “known,” “implicated,” and “novel,” based on prior evidence for their association with regions of the genome marked by these histone modifications (Fig. 2A). The known category contained proteins reported to associate with chromatin in the UniProt, GeneCards, or PubMed databases. The implicated group contained proteins that have not been reported to associate with chromatin but contain DNA-binding domains, chromatin-binding domains, or belong to a protein complex known to associate with chromatin. The novel category contained proteins not previously reported or implicated in chromatin association. We found 332 proteins known to be associated with chromatin, 46 proteins that have been implicated, and 114 novel candidates (Fig. 2A).

Chromatin Regulators Enriched in ChIP-MS. We focused our initial analysis on proteins that are known or are likely to be histone modification readers, writers, and erasers (Fig. 2B and Table S1). A large number of proteins that have well-established roles in these activities were enriched in the appropriate preparations (Fig. 2C and Table S1). For example, the bromodomain reader BRD4, the histone acetyltransferase writer CBP/P300, and the histone deacetylase eraser HDAC2 were especially enriched in the H3K27ac preparation, consistent with their functions at enhancers and promoters (19). The methyltransferase writer ASH2L and the demethylase erasers PHF8, KDM5A, and KDM5B were especially enriched in the H3K4me3 preparation, consistent with their functions in promoter regions (20–22). The PWWP domain reader N-PAC, an LSD2/KDM1B cofactor that stimulates H3K4 demethylation, was especially enriched in the H3K36me3 preparation (23). The chromobox readers CBX1, CBX3, and CBX5 and the histone methyltransferase writer SETDB1 were especially enriched in the H3K9me3 and H4K20me3 preparations, consistent with their functions in heterochromatic regions (24). There were many additional proteins identified in these preparations that have established enzymatic activities (25) but that have not been reported to be associated with chromatin (Fig. 2B and Table S1); these candidate histone readers, writers, and erasers thus may contribute to chromatin regulation.

To confirm that a subset of the candidate proteins was in fact associated with specific histone-marked genomic regions suggested by the ChIP-MS data, we further studied PHIP, UBR7, and USP48, proteins for which specific and ChIP-capable antibodies were available. We anticipate that PHIP, UBR7, and USP48 may function as a chromatin reader, writer, and eraser, respectively; this possibility warrants future investigation. ChIP followed by Western blot analysis confirmed that PHIP, UBR7, and USP48 were associated with the corresponding histone mark ChIP-MS preparations (Fig. 2D). ChIP-quantitative PCR (qPCR) then was performed for PHIP, UBR7, and USP48 with probes targeting enhancers, promoters, and gene bodies. The ChIP-qPCR signals for PHIP, UBR7, and USP48 were relatively enriched in enhancers and promoters, consistent with the enrichment of PHIP, UBR7, and USP48 in the H3K27ac and H3K4me3 preparations (Fig. 2E, Fig. S2, and Table S2). Small differences in the relative enrichments at different regions are seen with ChIP-qPCR and ChIP-MS, likely reflecting the difference between a gene-specific view using ChIP-qPCR and the global average view provided by ChIP-MS, as well as technical differences in the two approaches.

Transcription Factors and Cofactors Enriched in ChIP-MS. We extended our analysis of proteins in the ChIP-MS preparations to transcription factors (Fig. 3A), proteins associated with transcription initiation (Fig. 3B), and proteins associated with the elongation apparatus (Fig. 3C). The well-studied ES cell pluripotency transcription factors OCT4, SOX2, ESRRB, and SALL4,

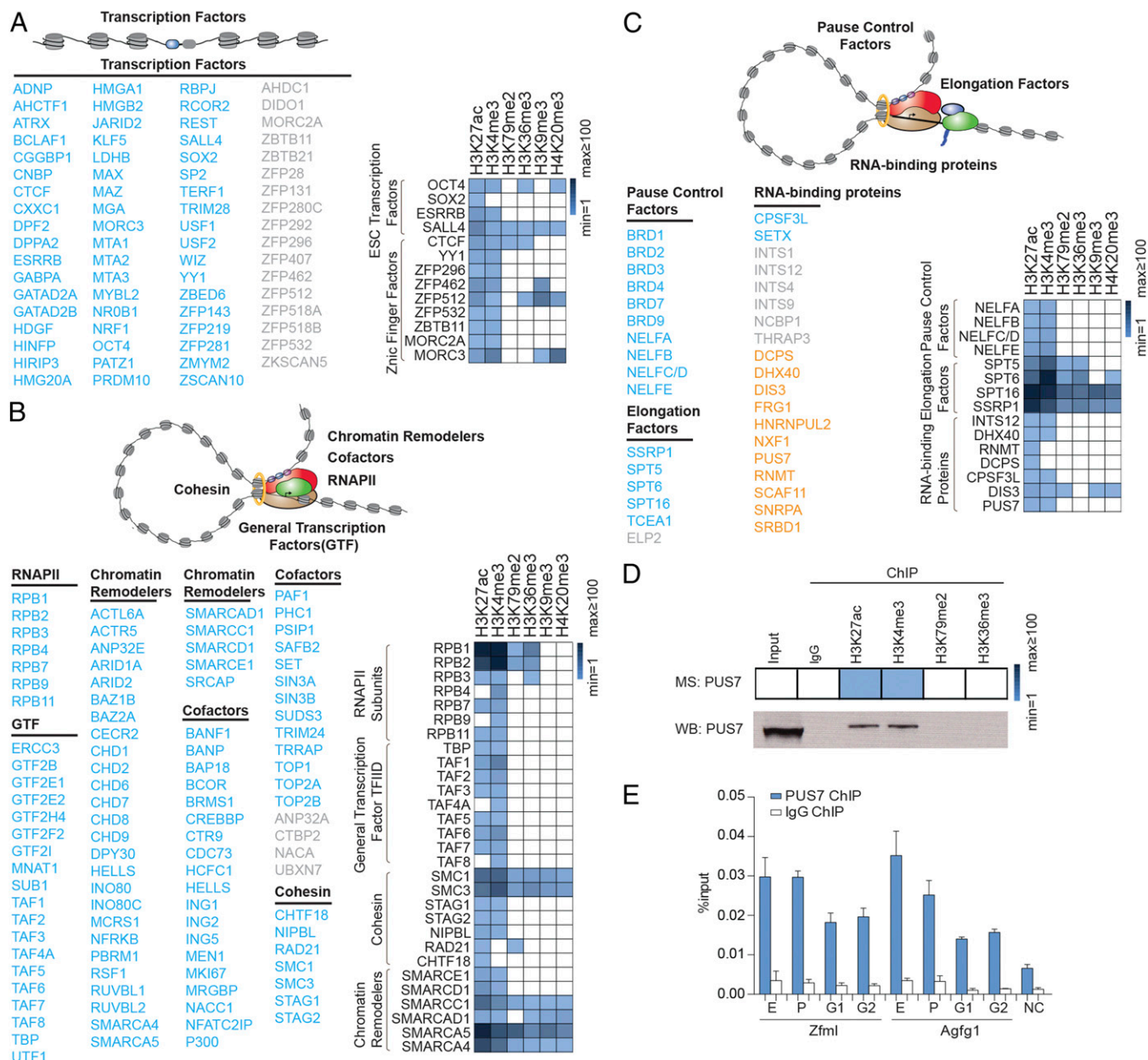


Fig. 3. Transcriptional regulators enriched in the active histone mark ChIP-MS preparations in mESCs. (A–C) Transcription factors (A), proteins associated with the transcription-initiation apparatus (B), and proteins associated with the elongation apparatus (C) were enriched in the histone mark preparations. (D, Upper) Relative enrichment of PUS7 in various ChIP preparations. (Lower) Histone mark ChIP followed by Western blot analysis of PUS7. Of the input, 0.25% was loaded for ChIP-Western blotting. (E) ChIP-qPCR analysis of PUS7 with primers targeting enhancers (E), promoters (P), and gene body regions (G1 and G2) of *Zfml* and *Agfg1*. NC, negative control region for PCR. Error bars indicate SD based on two independent experiments.

preparations (Fig. 3B). Proteins associated with the elongation apparatus included pause-control factors, elongation factors, and several RNA-binding proteins (Fig. 3C) (32, 33). Recent studies indicate that RNA-binding proteins can be associated with specific portions of chromatin, in part because of their interaction with RNA species associated with the transcriptional machinery (34, 35). For example, the integrator complex, an snRNA-processing machinery, can be detected at the gene promoters, where it contributes to the regulation of transcriptional initiation and pause release (36). The RNA-binding proteins observed in these chromatin preparations included the RNA helicase DHX40, the RNA-capping enzyme RNMT, the RNA-decapping enzyme DCPS, the RNA cleavage factor CPSF3L, and many others.

To confirm that an RNA-binding protein was associated with specific histone-marked genomic regions, we further studied a pseudouridine synthase, PUS7, for which a specific and ChIP-capable antibody was available. PUS7 recently was found to be required for mRNA pseudouridylation in yeast and human cells (37, 38); however, its chromatin association is unknown. We used ChIP followed by Western blot analysis to confirm that PUS7 was associated with the corresponding histone mark ChIP-MS preparations (Fig. 3D). We then performed ChIP-qPCR for PUS7 with probes targeting enhancers, promoters, and gene bodies. The ChIP-qPCR signals for PUS7 were relatively enriched in enhancers and promoters, consistent with the enrichment of PUS7 in the H3K27ac and H3K4me3 preparations

(Fig. 3E). This evidence suggests that PUS7 may play a role at the cotranscriptional level.

Discussion

The chromatin proteomic profiling approach described here combines ChIP-MS and ChIP-seq both to identify proteins associated with specific types of chromatin and to map that type of chromatin genome-wide. The mESC chromatin proteomic profiling data produced a catalog of chromatin-associated proteins, which includes a large number of known and candidate chromatin regulators and transcriptional regulators. Our results overlap and confirm chromatin-associated proteins identified in other MS studies (8, 12) but also provide a larger catalog of chromatin proteins associated with a larger variety of chromatin types. We confirmed that several of the candidates (PHIP, UBR7, USP48, and PUS7) were indeed associated with appropriate histone-marked segments of the genome as detected by chromatin proteomic profiling. These results indicate that chromatin proteomic profiling is a powerful means to identify proteins associated with chromatin containing specific histone modifications.

Although many well-studied proteins were found in the expected ChIP-MS preparations, we noted two limitations of the ChIP-MS chromatin proteomic profiling data. Proteins that are known to be present in these regions of chromatin were not detected, showing that the method did not produce a complete catalog. For example, five of the smaller subunits of the 12-subunit RNA polymerase II enzyme were not detected (39). The transcription factors NANOG and KLF4 co-occupy many promoters together with OCT4, SOX2, ESRRB, and SALL4 (19), but these were not detected. The failure to detect these and other proteins may result from their relatively small size, low abundance, loss during preparation, or poor signal detection by MS. Another potential limitation of the ChIP-MS chromatin proteomic profiling data is that some proteins are found both in the expected preparations and in others. For example, the H3K4me3 writer ASH2L was found most abundantly in the H3K4me3 preparation but also was found in the H3K27ac and H3K36me3 preparations. The presence of ASH2L in the H3K27ac preparation was expected, because this modification occurs in the same region as H3K4me3 (Fig. 1A), but it was not expected to be present in the H3K36me3 preparation. In the future, ChIP-MS of single nucleosomes prepared through nuclease digestion or ChIP-MS using sequential chromatin immunoprecipitations of multiple histone marks might help increase the specificity of chromatin proteomic profiling.

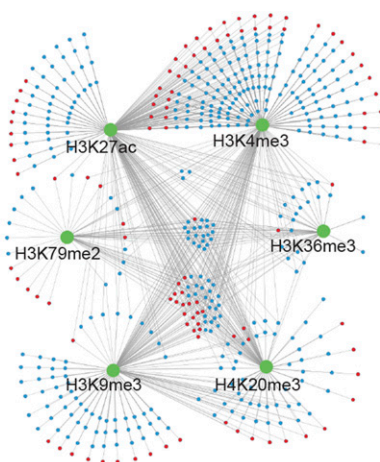
We note that many proteins identified by this chromatin proteomic profiling study have been implicated in various diseases (40–43) but have poorly understood functions (Fig. 4 and Table S3). Nearly 20% of the proteins in these chromatin preparations (110 proteins) have been implicated in cancer and/or Mendelian diseases. Future studies of these proteins, nearly one-fourth of which fell into the novel or implicated categories, may lead to additional insights into disease mechanisms.

The chromatin proteomic profiling study described here produced a catalog of known, implicated, and novel proteins associated with genomic regions whose chromatin is marked by specific modified histones. More than 100 additional histone modifications have been reported (44), so further chromatin proteomic profiling studies should prove valuable for identifying additional chromatin-associated proteins in ES cells and other cell types. Chromatin proteomic profiling also should prove useful in investigating changes in protein composition in dynamic settings such as drug exposure, metabolism, cell growth/cell cycle, stress response, and development.

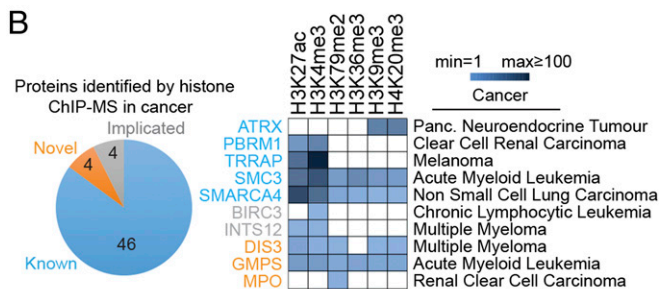
Materials and Methods

Cell Culture. V6.5 (C57BL/6–129) mESCs were grown on irradiated murine embryonic fibroblasts (MEFs) under standard ES cell culture conditions as described previously (45). mESCs were collected for specific experiments after being grown for two passages without MEFs.

A



B



C

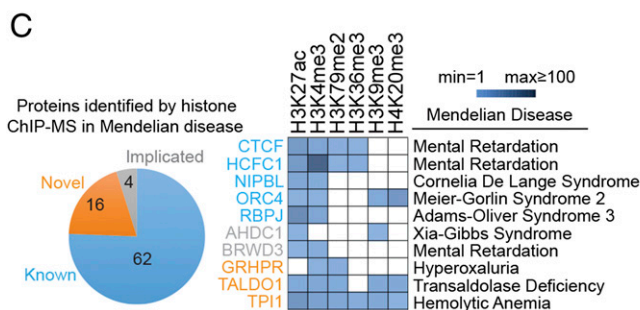


Fig. 4. Proteins identified by chromatin proteomic profiling that are associated with human diseases. (A) Distribution of proteins (blue or red nodes) that were enriched in each histone mark ChIP-MS preparation (green nodes) relative to the IgG control and input samples. Red nodes represent proteins associated with diseases. (B and C) The number and relative enrichment of chromatin-associated proteins identified in this study that are mutated in cancers (B) and Mendelian diseases (C).

Antibodies. The following antibodies were used for Western blot analysis, ChIP-MS, and ChIP-seq/qPCR: H3K27ac (ab4729), H3K4me3 (ab8580), H3K79me2 (ab3594), H3K36me3 (ab9050), H3K9me3 (ab8898), and H4K20me3 (ab9053) (all from Abcam). The USP48 (A301-190A) and PHIP (A302-055A) antibodies were purchased from Bethyl Laboratories. The UBR7 (NBP1-88409) antibody was purchased from NOVUS. The PUS7 (HPA024116) antibody was purchased from Sigma.

ChIP-MS and ChIP-Seq. We used $6\text{--}8 \times 10^8$ mESCs for each ChIP-MS and $6\text{--}8 \times 10^7$ mESCs for each ChIP-seq. See *SI Materials and Methods* for further details regarding the protocols and data processing.

Accession Codes. The ChIP-seq data for H3K27ac, H3K4me3, H3K79me2, and H3K36me3 are available at the Gene Expression Omnibus under the accession number GSE62380.

ACKNOWLEDGMENTS. We thank Eric Spooner at the Whitehead Proteomics Core for mass spectrometry; Tom Volkert and Sumeet Gupta at the Whitehead Genome Technology Core for sequencing; Zi Peng Fan and Violaine Saint-André for protein domain analysis; Jessica Reddy and Lyndon Zhang

for helpful bioinformatics discussions; and Jill Downen, Denes Hnisz, and Jurian Schuijers for comments on the manuscript. This work was supported

by National Institutes of Health Grants HG002668 (to R.A.Y.), HG006046 (to R.A.Y.), and HD045022 (to R.J.).

1. Levine M, Cattoglio C, Tjian R (2014) Looping back to leap forward: Transcription enters a new era. *Cell* 157(1):13–25.
2. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM (2009) A census of human transcription factors: Function, expression and evolution. *Nat Rev Genet* 10(4): 252–263.
3. Dawson MA, Kouzarides T (2012) Cancer epigenetics: From mechanism to therapy. *Cell* 150(1):12–27.
4. Maston GA, Landt SG, Snyder M, Green MR (2012) Characterization of enhancer function from genome-wide analyses. *Annu Rev Genomics Hum Genet* 13(13):29–57.
5. Zhou VW, Goren A, Bernstein BE (2011) Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* 12(1):7–18.
6. Dunham I, et al.; ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74.
7. Rivera CM, Ren B (2013) Mapping human epigenomes. *Cell* 155(1):39–55.
8. Vermeulen M, et al. (2010) Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* 142(6):967–980.
9. Wang J, et al. (2006) A protein interaction network for pluripotency of embryonic stem cells. *Nature* 444(7117):364–368.
10. Bensimon A, Heck AJR, Aebersold R (2012) Mass spectrometry-based proteomics and network biology. *Annu Rev Biochem* 81(81):379–405.
11. Wang CL, et al. (2013) Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in *Drosophila*. *Nat Struct Mol Biol* 20(2):202–209.
12. Soldi M, Bonaldi T (2013) The proteomic investigation of chromatin functional domains reveals novel synergisms among distinct heterochromatin components. *Mol Cell Proteomics* 12(3):764–780.
13. Soldi M, Bonaldi T (2014) The ChroP approach combines ChIP and mass spectrometry to dissect locus-specific proteomic landscapes of chromatin. *J Vis Exp* 86.
14. Mikkelsen TS, et al. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448(7153):553–560.
15. Creighton MP, et al. (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA* 107(50):21931–21936.
16. Li B, Carey M, Workman JL (2007) The role of chromatin during transcription. *Cell* 128(4):707–719.
17. Berger SL (2007) The complex language of chromatin regulation during transcription. *Nature* 447(7143):407–412.
18. Campos EI, Reinberg D (2009) Histones: Annotating chromatin. *Annu Rev Genet* 43: 559–599.
19. Hnisz D, et al. (2013) Super-enhancers in the control of cell identity and disease. *Cell* 155(4):934–947.
20. Christensen J, et al. (2007) RBP2 belongs to a family of demethylases, specific for tri- and dimethylated lysine 4 on histone 3. *Cell* 128(6):1063–1076.
21. Liu W, et al. (2010) PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. *Nature* 466(7305):508–512.
22. Seward DJ, et al. (2007) Demethylation of trimethylated histone H3 Lys4 in vivo by JARID1 JmjC proteins. *Nat Struct Mol Biol* 14(3):240–242.
23. Fang R, et al. (2013) LSD2/KDM1B and its cofactor NPAC/GLYR1 endow a structural and molecular model for regulation of H3K4 demethylation. *Mol Cell* 49(3):558–570.
24. Kouzarides T (2007) Chromatin modifications and their function. *Cell* 128(4):693–705.
25. Safran M, et al. (2010) GeneCards version 3: The human gene integrator. *Database*, baq020.
26. Ng H-H, Surani MA (2011) The transcriptional and signalling networks of pluripotency. *Nat Cell Biol* 13(5):490–496.
27. Young RA (2011) Control of the embryonic stem cell state. *Cell* 144(6):940–954.
28. Kagey MH, et al. (2010) Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467(7314):430–435.
29. Wu C, Macleod I, Su AI (2013) BioGPS and MyGene.info: Organizing online, gene-centric information. *Nucleic Acids Res* 41(Database issue, D1):D561–D565.
30. Roeder RG (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* 21(9):327–335.
31. Lessard JA, Crabtree GR (2010) Chromatin regulatory mechanisms in pluripotency. *Ann Rev Cell ad Dev Biol* 26:503–532.
32. Zhou Q, Li T, Price DH (2012) RNA polymerase II elongation control. *Annu Rev Biochem* 81(81):119–143.
33. Kwak H, Lis JT (2013) Control of transcriptional elongation. *Annu Rev Genet* 47(47): 483–508.
34. Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T (2011) Epigenetics in alternative pre-mRNA splicing. *Cell* 144(1):16–26.
35. Ji X, et al. (2013) SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. *Cell* 153(4):855–868.
36. Gardini A, et al. (2014) Integrator Regulates Transcriptional Initiation and Pause Release following Activation. *Mol Cell* 56:1–12.
37. Schwartz S, et al. (2014) Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* 159(1):148–162.
38. Carlile TM, et al. (2014) Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* 515(7525):143–146.
39. Cramer P (2002) Multisubunit RNA polymerases. *Curr Opin Struct Biol* 12(1):89–97.
40. 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.
41. Lawrence MS, et al. (2014) Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505(7484):495–501.
42. Lee TI, Young RA (2013) Transcriptional regulation and its misregulation in disease. *Cell* 152(6):1237–1251.
43. Online Mendelian Inheritance in Man OMIM (2014) (Johns Hopkins University, Baltimore). Available at omim.org/. Accessed September 1, 2014.
44. Tan M, et al. (2011) Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* 146(6):1016–1028.
45. Whyte WA, et al. (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153(2):307–319.