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

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1073/pnas.1502971112">http://dx.doi.org/10.1073/pnas.1502971112</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>National Academy of Sciences (U.S.)</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Sat Jun 30 19:29:44 EDT 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/98386">http://hdl.handle.net/1721.1/98386</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions

Xiong Ji,a,b,1 Daniel B. Dadon,a,b,1 Brian J. Abraham,a Tong Ihn Lee,b Rudolf Jaenisch,a,b,2 James E. Bradner,a,b,2 and Richard A. Younga,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 patterns of genome occupancy 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.

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.
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 heterochromatin histone ChIP-MS preparations were not found in the euchromatin 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 initial 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, 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 heterochromatin histone ChIP-MS preparations were not found in the euchromatin 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 initial 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,
as well as many additional transcription factors, were found to be enriched mostly in the H3K27ac and H3K4me3 preparations (Fig. 3A), as is consistent with their established presence at enhancers and promoter-proximal regions (19, 26, 27). Many transcription factors that previously have not been described as playing roles in ES cell gene expression also were found in these regions, e.g., ZFP131, which previously was reported to regulate Oct4 gene expression in an shRNA screen in mESCs (28). ZFP296 and ZFP532 also were found in our preparations. These transcription factors are specifically expressed in ES cells, suggesting potential roles in ES cell-specific transcriptional regulation (Fig. 3A) (29).

Many known components of the transcription initiation apparatus were found in the ChIP-MS preparations (Fig. 3B), including RNA polymerase II, transcriptional cofactors (such as the ATP-dependent chromatin remodeling complexes with SMARC subunits), and the general transcription factor TFIIID (containing TBP and TAF subunits) (1, 30, 31). TFIIID was found enriched only in the promoter-associated H3K27ac and H3K4me3 preparations, as is consistent with its known promoter-binding properties (Fig. 3B). Some proteins associated with the transcription-initiation apparatus (RNA polymerase II, cohesin) are thought to traverse actively transcribed genes and were found in all the euchromatin regions (19, 26, 27). Many transcription factors that previously have not been described as playing roles in ES cell gene expression also were found in these regions, e.g., ZFP131, which previously was reported to regulate Oct4 gene expression in an shRNA screen in mESCs (28). ZFP296 and ZFP532 also were found in our preparations. These transcription factors are specifically expressed in ES cells, suggesting potential roles in ES cell-specific transcriptional regulation (Fig. 3A) (29).
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 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.

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.
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 histones and to map that type of histone 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 nucleosome 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.

**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 (NB91818409) antibody was purchased from Novus. The PUS7 (HPA024116) antibody was purchased from Sigma.

**ChIP-MS and ChIP-Seq.** We used 6–8 × 10⁶ mESCs for each ChIP-MS and 6–8 × 10⁷ 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 Sunset 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


