



Universiteit
Leiden
The Netherlands

Comprehensive chromatin proteomics resolves functional phases of pluripotency and identifies changes in regulatory components

Ugur, E.; Porte, A. de la; Qin, W.H.; Bultmann, S.; Ivanova, A.; Drukker, M.E.; ... ; Leonhardt, H.

Citation

Ugur, E., Porte, A. de la, Qin, W. H., Bultmann, S., Ivanova, A., Drukker, M. E., ... Leonhardt, H. (2023). Comprehensive chromatin proteomics resolves functional phases of pluripotency and identifies changes in regulatory components. *Nucleic Acids Research*.
doi:10.1093/nar/gkad058

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](#)

Downloaded from: <https://hdl.handle.net/1887/3590031>

Note: To cite this publication please use the final published version (if applicable).

Comprehensive chromatin proteomics resolves functional phases of pluripotency and identifies changes in regulatory components

Enes Ugur^{1,2}, Alexandra de la Porte³, Weihua Qin¹, Sebastian Bultmann¹, Alina Ivanova^{1,7}, Micha Drukker^{3,4}, Matthias Mann^{2,6,*}, Michael Wierer^{2,5,*} and Heinrich Leonhardt^{1,*}

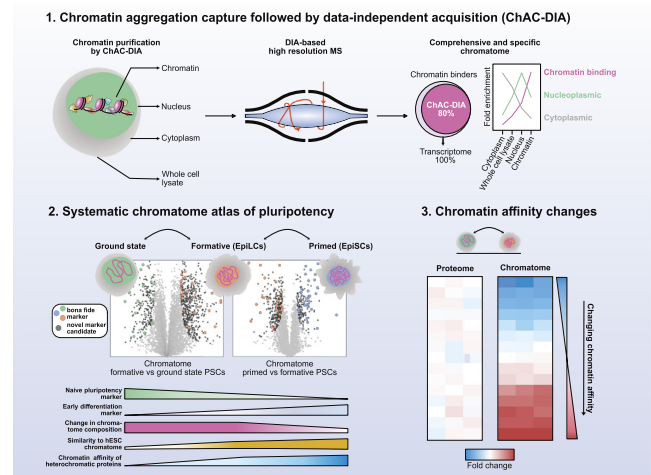
¹Faculty of Biology and Center for Molecular Biosystems (BioSysM), Human Biology and BioImaging, Ludwig-Maximilians-Universität München, Munich 81377, Germany, ²Department of Proteomics and Signal Transduction, Max-Planck Institute of Biochemistry, Martinsried 82152, Germany, ³Institute of Stem Cell Research, Helmholtz Center Munich, Neuherberg 85764, Germany, ⁴Division of Drug Discovery and Safety, Leiden Academic Centre for Drug Research (LACDR), Leiden University, Gorlaeus Building, 2333 CC RA Leiden, The Netherlands, ⁵Proteomics Research Infrastructure, University of Copenhagen, DK-2200 Copenhagen, Denmark, ⁶Novo Nordisk Foundation Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, DK-2200 Copenhagen, Denmark and ⁷Present address: Department of Totipotency, Max-Planck Institute of Biochemistry, Martinsried 82152, Germany

Received August 27, 2022; Revised January 11, 2023; Editorial Decision January 16, 2023; Accepted January 19, 2023

ABSTRACT

The establishment of cellular identity is driven by transcriptional and epigenetic regulators of the chromatin proteome - the chromatome. Comprehensive analyses of the chromatome composition and dynamics can therefore greatly improve our understanding of gene regulatory mechanisms. Here, we developed an accurate mass spectrometry (MS)-based proteomic method called Chromatin Aggregation Capture (ChAC) followed by Data-Independent Acquisition (DIA) and analyzed chromatome reorganizations during major phases of pluripotency. This enabled us to generate a comprehensive atlas of proteomes, chromatomes, and chromatin affinities for the ground, formative and primed pluripotency states, and to pinpoint the specific binding and rearrangement of regulatory components. These comprehensive datasets combined with extensive analyses identified phase-specific factors like QSER1 and JADE1/2/3 and provide a detailed foundation for an in-depth understanding of mechanisms that govern the phased progression of pluripotency. The technical advances reported here can be readily applied to other models in development and disease.

GRAPHICAL ABSTRACT



INTRODUCTION

DNA- and chromatin-binding proteins regulate gene expression and thereby govern cellular identity. During early embryonic development, the chromatin of pluripotent stem cells (PSCs) undergoes dynamic changes that are conserved among mammals (1–5). Pluripotency progresses in separate phases controlled by distinct signaling pathways and downstream transcription factors (3,6,7). Three major interme-

*To whom correspondence should be addressed. Tel: +49 89 2180 74232; Email: h.leonhardt@lmu.de
Correspondence may also be addressed to Michael Wierer. Tel: +45 35 3369 53; Email: michael.wierer@sund.ku.dk
Correspondence may also be addressed to Matthias Mann. Tel: +49 89 8578-2557; Email: mmann@biochem.mpg.de

diate phases of pluripotency have been described: naive (also referred to as ground state), formative, and primed (3). Ground state PSCs harbor a homogeneously organized and transcriptionally permissive chromatin with high plasticity and low levels of repressive epigenetic marks (8,9). In transition to the formative phase, PSCs gain trimethylation of lysines 4 and 27 of histone H3 at promoters, and the exclusive ability to differentiate into primordial germ cells, while losing the expression of certain naive genes (1,10). Finally, at the primed phase, PSCs are partially fate determined, yet still share a core regulatory circuitry with earlier pluripotency phases (3,11–14).

Current systems-wide knowledge of pluripotency is primarily based on transcriptome and epigenome analyses, and chromatin accessibility data (1,10,14–16). For instance, previous studies revealed that major chromatin reorganization and compaction occur at the formative phase (10). However, how this chromatin reorganization affects chromatin proteome composition, the chromatome (17), remains unknown. Moreover, although the expression of chromatin binders, such as transcription factors, has been extensively studied in PSCs (18–21), changes in expression do not inevitably entail changes in chromatin association. The latter has not been studied comprehensively on a global scale and instead mostly has been studied by focusing on specific transcription factors or histone PTM-associated proteins (22–26). Therefore, the complete picture of the chromatome structure and dynamics in functional phases of pluripotency is still largely missing.

Previous attempts to quantify global chromatomes combined high-resolution mass spectrometry (MS) with the biochemical purification of native (27,28) or formaldehyde (FA) crosslinked chromatin (29–32). Although these methods greatly contributed to the understanding of the chromatome, they offer limited insights as they cannot detect low-abundant DNA-binding factors that are known to play key regulatory roles despite low abundance. Furthermore, current sample preparation strategies require millions of cells (15–50 mio.) and multiple purification steps, which impairs overall protein recovery and quantification (30,31). Therefore, the current view of the chromatome remains incomplete.

To overcome these difficulties, we developed a method that combines a new streamlined chromatin purification strategy, Chromatin Aggregation Capture (ChAC), with Data-Independent Acquisition (DIA) MS-based proteomics, a powerful strategy for rapid, accurate, and reproducible proteomics analysis with a broad dynamic range that allows identification of low-abundant proteins starting with 100–250k cells. Using this method, we generated accurate and comprehensive chromatome maps of mouse naive, formative and primed PSCs that cover 80% of transcribed chromatin binders in single MS runs. Our analysis of these datasets revealed striking chromatome changes between different functional phases of pluripotency and provided evidence for novel, low-abundant chromatin binders that are dynamically regulated in pluripotency transitions. Additionally, by comparing the abundance of proteins in chromatomes and proteomes, we were able to infer chromatin reorganizations mediated by differential affinities or subcellular localizations. Finally, we applied this approach

to chromatomes of human PSCs to provide a mouse-to-human comparison of the pluripotency chromatome. Collectively, we present a comprehensive atlas of proteomes and chromatomes for the three pluripotency phases, thus revealing previously unknown details about how cell identity governing proteins are recruited to or evicted from chromatin in the process of pluripotency transitions. We have made the datasets available and searchable on an interactive web application, accessible on: <https://pluripotency.shinyapps.io/ChromatomeAtlas/>.

MATERIALS AND METHODS

Cell culture

Naive J1 mESCs were cultured in serum-free media consisting of: N2B27 (50% neurobasal medium (Life Technologies), 50% DMEM/F12 (Life Technologies)), 2i (1 μ M PD032591 and 3 μ M CHIR99021 (Axon Medchem, Netherlands)), 1000 U/ml recombinant leukemia inhibitory factor (LIF, Millipore), and 0.3% BSA (Gibco), 2 mM L-glutamine (Life Technologies), 0.1 mM β -mercaptoethanol (Life Technologies), N2 supplement (Life Technologies), B27 serum-free supplement (Life Technologies), and 100 U/ml penicillin, 100 μ g/ml streptomycin (Sigma). Formative EpiLCs were derived by differentiating naive mESCs (33) for 48 h using the same serum-free media for naive mESCs devoid of 2i, LIF, and BSA and supplemented with 10 ng/ml Fgf2 (R&D Systems), 20 ng/ml Activin A (R&D Systems) and 0.1 \times Knockout Serum Replacement (KSR) (Life Technologies). Both, naive mESCs and EpiLCs, were cultured on 0.2% gelatin-treated flasks. The media of EpiLCs was changed once after 24 h and all cells were harvested after 48 h. Cells were tested negative for Mycoplasma contamination by PCR.

Identical culture conditions for mouse formative and primed as well as human ESCs

129S2C1a mouse EpiSCs (34) and J1 EpiLCs that were compared directly to human ESCs H9 were cultured in UPPS medium consisting of StemMACS iPS Brew XF (Miltenyi Biotec) supplemented with 1 μ M IWR-1 (Sigma) and 0.5 μ M CHIR (Tocris) (35). ESCs, EpiSCs and compared EpiLCs were cultured on plates coated with Matrigel (Corning) diluted 1:100 in DMEM/F-12 (Thermo Fisher Scientific).

For all experiments, cells were differentiated/cultured in three independent flasks and are therefore considered to be three biological replicates. Cells were split upon harvesting for total proteome (5×10^6 cells per replicate) and chromatome (15×10^6 cells per replicate) analyses and flash-frozen. The following descriptions are based on the above-mentioned amounts. Systematic downscaling showed that as few as 1×10^4 to 1×10^5 cells per replicate may suffice (see also Materials and Methods details).

Total proteome sample preparation

Previously flash-frozen samples were quickly placed on ice and pellets were solubilized in 200 μ l lysis buffer (6 M guanidinium Chloride, 100 mM Tris-HCl pH 8.5, 2 mM

DTT) and heated for 10 min at 99°C under constant shaking at 1400 rpm. Subsequently, samples were sonicated at 4°C in 30 s on/off intervals for 15 cycles using a Bioruptor® Plus sonication instrument (Diagenode) at high-intensity settings. If the viscosity of the samples was sufficiently reduced, protein concentrations were estimated, otherwise, sonication was repeated. For concentration measurements, the Pierce™ BCA Protein Assay Kit (23225, Thermo Fisher Scientific) was employed following the manufacturer's instructions. After at least 20 min of incubation with 40 mM chloroacetamide, 30 µg of each proteome sample was diluted in a 30 µl lysis buffer supplemented with CAA and DTT. Samples were diluted in 270 µl digestion buffer (10% acetonitrile, 25 mM Tris-HCl pH 8.5, 0.6 µg Trypsin/sample (Pierce™ Trypsin Protease, 90058, Thermo Fisher Scientific) and 0.6 µg/sample LysC (Pierce™ LysC Protease, 90051, Thermo Fisher Scientific) and proteins digested for 16 h at 37°C with constant shaking at 1100 rpm.

To stop protease activity 1% (v/v) trifluoroacetic acid (TFA) was added the next day and samples were loaded on self-made StageTips consisting of three layers of SDB-RPS matrix (Empore) (36) that were previously equilibrated by 0.1% (v/v) TFA. After loading, two washing steps with 0.1% (v/v) TFA were scheduled, and peptides were eluted by 80% acetonitrile and 2% ammonium hydroxide. Upon evaporation of the eluates in a SpeedVac centrifuge, samples were resuspended in 20 µl 0.1% TFA and 2% acetonitrile. After complete solubilization of peptides by constant shaking for 10 min at 2,000 rpm, peptide concentrations were estimated on a Nanodrop™ 2000 spectrophotometer (Thermo Fisher Scientific) at 280 nm.

Chromatin aggregation capture

Previously flash-frozen samples were quickly placed on ice and pellets were solubilized in 1 ml cellular lysis buffer (20 mM HEPES pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% NP40, freshly added 1× cOmplete™ EDTA-free Protease Inhibitor Cocktail (04693132001, Roche)) and incubated for 10 min on ice. Nuclei were pelleted by centrifugation (2300 g, 5 min, 4°C) and the supernatant was discarded. In the differential fraction analysis (Figure 2A), the supernatant was saved as the cytosolic fraction. Upon a second wash of the nuclei pellet with the cellular lysis buffer, the nuclei were taken into 3 ml crosslinking buffer (PBS pH 7.4 (806552, Sigma), 1× cOmplete™ EDTA-free Protease Inhibitor Cocktail). Formaldehyde (28906, Thermo Fisher Scientific) was added to a final concentration of 1% and samples were incubated for 10 min on an orbital shaker at room temperature. Excess formaldehyde was then quenched by 125 mM Glycine for 5 min and crosslinked cells were washed twice with ice-cold PBS. Nuclei were lysed in 300 µl SDS buffer (50 mM HEPES pH 7.4, 10 mM EDTA pH 8.0, 4% UltraPure™ SDS Solution (24730020, Invitrogen), freshly added 1× cOmplete™ EDTA-free Protease Inhibitor Cocktail) by gentle pipetting. After 10 min incubation at room temperature, 900 µl freshly prepared Urea buffer (10 mM HEPES pH 7.4, 1 mM EDTA pH 8.0, 8 M urea (U4883, Sigma)) was added. Tubes were carefully inverted 7 times and centrifuged at 20 000 g and room temperature for 30 min. The supernatant was discarded without

perturbing the pellet. The pellet was resuspended in 300 µl Sonication buffer (10 mM HEPES pH 7.4, 2 mM MgCl₂, freshly added 1× cOmplete™ EDTA-free Protease Inhibitor Cocktail). Before sonication, two additional wash steps can be scheduled (one SDS and urea wash and one SDS only wash) (30), but to our hands, this did not notably improve the chromatin enrichment efficiency. The chromatin samples were sonicated using a Bioruptor® Plus at 4°C for 15 cycles (30 s on, 60 s off). The protein concentration was estimated by the Pierce™ BCA Protein Assay Kit.

Next, protein aggregation capture (PAC) was performed. Here 1000 µg of undiluted Sera-Mag™ beads (1 mg, GE24152105050250, Sigma) per 100 µg chromatin solution were washed three times by 70% acetonitrile. 300 µl of the chromatin solution corresponding to 100 µg was added after the last wash to the beads and 700 µl 100% acetonitrile was added to each sample. Chromatome-bead mixtures were vortexed. After 10 min incubation on the bench, the samples were again vortexed and rested on the bench. Samples were then placed into a magnetic rack. A first wash followed this with 700 µl 100% acetonitrile, a second wash with 1 ml 95% acetonitrile, and a third wash with 1 ml 70% ethanol. The remaining ethanol was allowed to evaporate and beads were resuspended in 400 µl 50 mM HEPES pH 8.5 supplemented with fresh 5 mM TCEP and 5.5 mM CAA. Samples were incubated for 30 min at room temperature upon which LysC (1:200) and Trypsin (1:100) were added. Proteins were digested overnight at 37°C. From this step on, samples were treated exactly like the total proteome samples.

Chromatin aggregation capture of <1 million cells

Chromatin aggregation capture for sub-million amounts of cells was performed with some additional modifications to the standard protocol. Here, cells were directly harvested into a DNase-/RNase-free 1.5 ml tube (0030108051, Eppendorf). Nuclei were then isolated by 0.5 ml of cellular lysis buffer and the nuclei pellet was resuspended in 666 µl crosslinking buffer. After crosslinking with 1% formaldehyde and subsequent formaldehyde quenching with 125 mM Glycine, the chromatin extraction was performed again by SDS and Urea washes with careful pipetting so that nothing would stick to the pipette tip. Of note, with <100 000 cells the chromatin is not visually pelleted but rather a smear that spreads at the wall of the tube. For 10 000 cells even this smear is not visible anymore and it is advised to use a thermal shaker at 1,500 rpm instead of pipetting. For 10 000–250 000 cells the protein yield after sonication was between 10–16 µg. Here, we used 10 µg as input for the PAC purification and 1500 µg magnetic beads per replicate since smaller amounts require a higher bead-to-protein ratio (37). After the peptide cleanup, these samples were resuspended in 8 µl of 0.1% TFA and 2% acetonitrile.

Chromatin immunoprecipitation for MS analysis

Chromatin immunoprecipitation for subsequent MS analysis (ChIP-MS) using a KAT7 (Abcam, ab70183), H3K4me3 (Abcam, ab8580), H3K9me3 (Abcam, ab8898) or normal rabbit IgG (Cell Signaling Technology, #2729) antibody

was performed in triplicates in naive, formative and primed PSCs. ChIP-MS was performed like previously described (38–40), but without nuclei isolation and MNase digestion. Briefly, for each replicate, independently grown 10×10^6 cells were harvested and crosslinked in 1% paraformaldehyde. Lysis of cells was performed in IP buffer (1.7% Triton X-100, 100 mM NaCl, 50 mM Tris-HCl pH 8.0, 5 mM EDTA pH 8.0, 0.3% SDS, and freshly added $1 \times$ protease inhibitor cocktail). After 10 min incubation on ice, samples were sonicated for 15 min in a Bioruptor Plus (30 s on/off cycles, Diagenode). Shearing efficiency was checked after overnight reverse crosslinking and proteinase K digestion of samples on a 1% agarose gel. Shearing had to be repeated twice to reach an average DNA length of ~ 150 –1000 bp. Protein concentrations were estimated by BCA assay (Thermo). Samples were subsequently diluted to 1 mg/ml in 1 ml. 2 μ g of the antibody was added to each replicate and samples were incubated O/N at 4°C under constant rotation. 80 μ l of protein A sepharose bead slurry volume was added to each sample. After two hours of incubation at 4°C and under constant rotation, beads were washed three times by a low salt buffer (50 mM HEPES pH 7.5, 140 mM NaCl, 1% Triton X-100) and once by a high salt buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 1% Triton X-100). In case of histone pulldowns, a third wash buffer was used (50 mM HEPES pH 7.5, 250 mM LiCl, 1% Triton X-100) after the high salt wash. Samples were then washed three times by TBS. Supernatants were discarded and beads were resuspended in 50 μ l 2 mM DTT for 30 min at 37°C and subsequently 40 mM CAA for 5 min at 37°C (both diluted in 2 M Urea and 50 mM Tris-HCl pH 7.5). Then proteins were on-bead digested by Trypsin (20 μ g/ml) O/N at 25°C. The next day, protease activity was stopped by 1% TFA and peptides were cleaned up on StageTips consisting of three layers of C18 material (Empore) (36). After elution from StageTips peptides were speedvac dried and resuspended in 20 μ l of A* buffer (0.1% TFA and 2% acetonitrile). Peptide concentrations were estimated on a Nanodrop™ 2000 spectrophotometer (Thermo Fisher Scientific) at 280 nm.

Acid histone extraction

5 Mio. cells were harvested and nuclei were isolated by cellular lysis buffer (20 mM HEPES pH 7.4, 10 mM NaCl, 3 mM MgCl_2 , 0.1% NP40, freshly added $1 \times$ cOmplete™ EDTA-free Protease Inhibitor Cocktail (04693132001, Roche)) and histones were extracted by 0.2 N HCl at a density of 5 Mio. nuclei/500 μ l. Samples were incubated O/N at 4°C under constant rotation. After spinning at 16 000 g for 10 min at 4°C, the histone containing supernatant was acetone precipitated (5 volumes acetone: 1 volume histones). Histones were solubilized in DNase- and RNase-free water (Thermo Fisher Scientific, 10977035).

SDS-PAGE and western blot

8 μ g of the chromatome and full proteome extracts and 1 μ g of acid histone extracts were separated on SDS-PAGE. Proteins were transferred onto a nitrocellulose membrane

and incubated with an antibody against QSER1 (Abcam, ab86072, 1:1000) or H3K9me3 (Abcam, ab8898, 1:1000). The secondary antibody of goat-anti-rabbit IgG (H + L)–HRP conjugate was used with a dilution of 1:5000. Blots were developed with the Pierce ECL western blotting substrate (Thermo Scientific, 32109) and scanned by the Amersham™ Imager 600 system.

Nanoflow LC–MS/MS measurements for proteomes and chromatomes

Peptides were separated prior to MS by liquid chromatography on an Easy-nLC 1200 (Thermo Fisher Scientific) on in-house packed 50 cm columns of ReproSilPur C18-AQ 1.9- μ m resin (Dr Maisch GmbH). By employing a binary buffer system (buffer A: 0.1% formic acid and buffer B: 0.1% formic acid and 80% acetonitrile) with successively increasing buffer B percentage (from 5% in the beginning to 95% at the end) peptides were eluted for 120 min under a constant flow rate of 300 nl/min. Via a nanoelectrospray source, peptides were then injected into an Orbitrap Exploris™ 480 mass spectrometer (Thermo Fisher Scientific). Samples were scheduled in triplicates and a subsequent washing step while the column temperature was constantly at 60°C. Thereby the operational parameters were monitored in real-time by SprayQc.

DDA-based runs consisted of a top12 shotgun proteomics method within a range of 300–1650 m/z , a default charge state of 2, and a maximum injection time of 25 ms. The resolution of full scans was set to 60 000 and the normalized AGC target was set to 300%. For MS2 scans the orbitrap resolution was set to 15 000 and the normalized AGC target to 100%. The maximum injection time was 28 ms.

DIA-based runs employed an orbitrap resolution of 120 000 for full scans in a scan range of 350–1400 m/z . The maximum injection time was set to 45 ms. For MS2 acquisitions the mass range was set to 361–1033 with isolation windows of 22.4 m/z . A window overlap of 1 m/z was set as default. The orbitrap resolution for MS2 scans was at 30 000, the normalized AGC target was at 1000%, and the maximum injection time was at 54 ms. The tested DIA methods varied within the range of the isolation windows which were 37.3 m/z for in total of 18 windows and 16.8 m/z for in total of 40 windows.

MS data quantification

DIA-NN-based analysis of raw MS data acquired in DIA mode was performed by using version 1.7.17 beta 12 in ‘high accuracy’ mode. Instead of a previously measured precursor library, spectra and RTs were predicted by a deep learning-based algorithm and spectral libraries were generated from FASTA files. Cross-run normalization was established in an RT-dependent manner. Missed cleavages were set to 1. N-terminal methionine excision was activated and cysteine carbamidomethylation was set as a fixed modification. Proteins were grouped with the additional command ‘-relaxed-prot-inf’. Match-between runs was enabled and the precursor FDR was set to 1%.

The DIA raw files were analyzed with the Spectronaut Pulsar X software package (Biognosys, version 14.10.201222.47784) (41) applying the default Biognosys factory settings for DIA analysis (Q -value cutoff at precursor and protein level was set to 0.01). Imputation of missing values was disabled.

The DDA raw files were analyzed with MaxQuant 1.6.11.0 (42). ‘Match between runs’ was enabled and the FDR was adjusted to 1%, including proteins and peptides. The MaxLFQ algorithm was enabled for the relative quantification of proteins (43). Contaminants were defined by using the Andromeda search engine (44).

Statistical analyses

Downstream analysis of raw data output was performed with Perseus (version 1.6.0.9) (45). For the calculation of CVs, proteins or precursors with <2 out of 3 valid values were filtered out. For GO term counts the filtering was more strict and 3 out of 3 valid values were required. GO enrichment analyses of differentially enriched proteins (Figure 2A) were performed against the background of total identified proteins by employing a Benjamini-Hochberg FDR-corrected Fisher’s Exact test. The analysis was thereby performed individually for each cluster. The functional enrichment analysis of proteins enriched by ChAC-DIA versus total proteome was performed by ranking proteins according to their enrichment in the ChAC-DIA fraction. The functional enrichment analysis was thereby based on STRING (46).

Student’s t -tests were performed after imputation of missing values. The latter was always performed based on a Gaussian distribution relative to the standard deviations of measured values (width of 0.2 and a downshift of 1.8 standard deviations). Both, one- and two-sided t -tests were calculated with a permutation-based FDR of 0.05 and an $s_0 = 1$ if not otherwise declared. For the multiple sample test based on an ANOVA (Figure 2A) we chose a minimal 1.5-fold change. We performed imputation for missing values, except for supplementary heatmaps that represent the data without imputation (Supplementary Figures S4–S9). Student’s t -tests of normalized chromatomes were performed after calculating pairwise differences of ChAC-DIA and total proteome values. The complete catalog of proteins found in the naive, formative, and primed states can be found in Supplementary Table 3.

Correlations between samples in the differential fraction analysis experiment were calculated with Perseus, and the correlations between transcriptomes, proteomes, and chromatomes were calculated with GraphPad Prism (version 9.1.0).

Analysis of ChIP-MS experiments was performed by first filtering out proteins that were identified less than twice in a set of triplicates. A two-sided Student’s t -test of the \log_2 transformed LFQ intensities (specific pulldown vs normal IgG pulldown) was performed to obtain significantly enriched proteins. By definition, a permutation-based false discovery rate of 5% and a fold change cut-off of $\log_2 = 1$ were applied. For stoichiometry calculations of the HBO1 complex, iBAQ values were \log_2 transformed and normalized to KAT7.

Web application development

Row-normalized z -scores for each significantly changing protein across the ChAC-DIA purification steps were generated for an interactive profile plot representation of the data. Significant chromatome and proteome changes during pluripotency were represented in an interactive heatmap as mean row differences of \log_2 intensities.

The web application was programmed using R Shiny with the following libraries besides base R packages for data processing and visualization: shiny (1.7.1), shinydashboard (0.7.2), shinyHeatmaply (0.2.0), plotly (4.10.0), heatmaply (1.3.0) and png (0.1–7). From the tidyverse (1.3.1) family we further utilized tidyr (1.2.0), dplyr (1.0.9), and ggplot2 (3.3.6).

RESULTS

Chromatin aggregation capture (ChAC) followed by data-independent MS acquisition (DIA) enables near-complete chromatome identification and high-precision quantification

We hypothesized that accurate and comprehensive chromatin proteomics could be accomplished by combining Chromatin Aggregation Capture (ChAC) with Data Independent Acquisition (DIA). The method comprises nuclei isolation and formaldehyde crosslinking followed by an initial chromatin enrichment under denaturing conditions similar to the Chromatin enrichment for proteomics (ChEP) protocol (30). This is followed by an additional purification based on the protein aggregation capture (PAC) technique (37) to generate specific and pure chromatin fractions, and achieve highly accurate quantification by DIA-based MS using the DIA-NN software package (47). Briefly, in DIA, all peptide precursors that fall into a pre-defined mass-to-charge (m/z) window are fragmented and acquired on the MS2-level compared to selecting the top N most abundant peptide ions in a typical Data-Dependent MS Acquisition experiment (DDA) (41,48–51). The application of DIA is especially relevant for the analysis of enriched cellular structures that consist of highly repetitive structural elements such as nucleosomes. Here, DIA is much more sensitive and accurate for lower abundant proteins than the more semi-stochastic DDA-based approach (52,53). To improve chromatome quantification accuracy and comprehensiveness, we optimized the protocol, MS acquisition strategy (Supplementary Figure S1A–C), and raw data analysis (Supplementary Figure S1D–H) (Supplementary Table 1).

To benchmark the chromatome protocol, we performed ChAC-DIA in naive mouse embryonic stem cells (mESCs) and compared it to a recent ChEP-based chromatome data set of mESCs (PRIDE: PXD011782) (54). ChAC-DIA identified over 2.5 times more proteins in half of the MS acquisition time (Figure 1B). In addition, ChAC-DIA quantified proteins more reproducibly with median coefficients of variation (CVs) of 4% compared to 16% in the previous study (Figure 1B and Supplementary Table 1). The CV differences were even more pronounced at the peptide ion level (Supplementary Figure S1E).

Next, we classified nuclear, DNA-binding, RNA-binding, or chromatin-binding proteins based on their

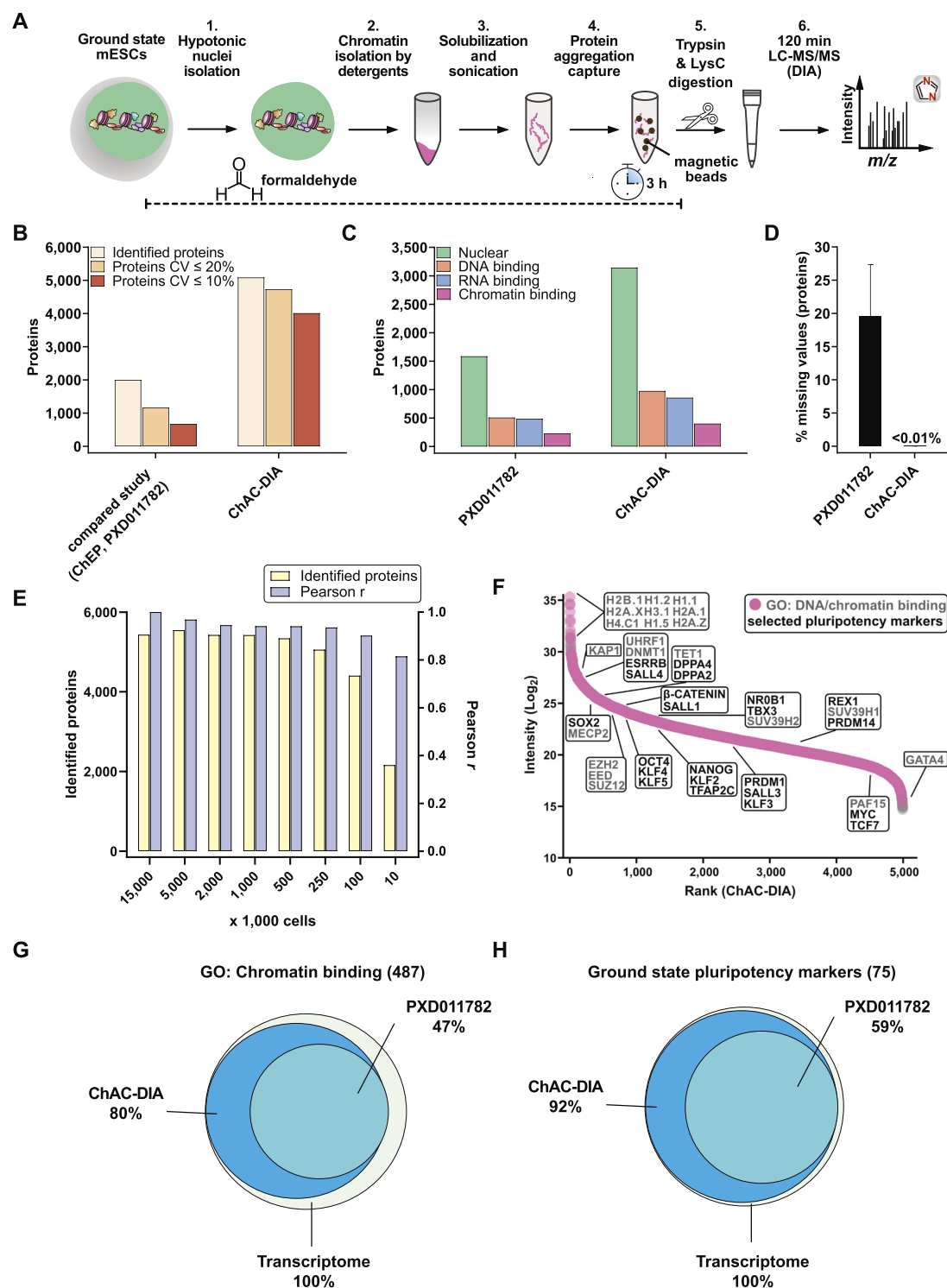


Figure 1. Chromatin aggregation capture (ChAC) followed by data-independent MS acquisition (DIA) enables near-complete chromatinome identification and high-precision quantification. (A) Schematic workflow of ChAC-DIA. (B) Total numbers of identified proteins with representations of the coefficient of variation (CV) below 20% and 10%. ChAC-DIA results obtained in library-free mode by DIA-NN were benchmarked against a previous study based on the ChEP protocol (PRIDE: PXD011782). In both cases, mouse naive PSCs were used. (C) Total numbers of proteins falling into a gene ontology (GO) category. (D) Percentage of missing intensity values on protein level across replicates. (E) Total numbers of identified proteins and Pearson correlation coefficients of ChAC-DIA applied on different cell amounts. Pearson *r* reflects the correlation with the standard protocol comprising 15 Mio cells. (F) Protein abundance rank based on the ChAC-DIA-derived naive PSC chromatinome. Chromatin binding proteins are highlighted in pink. Protein names in black indicate examples of *bona fide* pluripotency factors. Protein names in gray indicate other chromatin binders and the highest ranked nine proteins. (G) Venn diagram of proteins annotated as chromatin binding in ChAC-DIA, the compared study, and a transcriptome data set of naive PSCs (ArrayExpress: E-MTAB-6797). (H) Venn diagram of literature derived *bona fide* naive pluripotency factors identified by ChAC-DIA, the compared study, and a transcriptome data set of naive PSCs (ArrayExpress: E-MTAB-6797). See also Supplementary Figures S1 and S2.

Gene Ontology (GO) annotations (55). ChAC-DIA identified more than twice the number of nuclear and DNA-binding proteins, and three times more unique peptides of DNA-binding proteins as the previous ChEP method despite half of the required MS time (Figure 1C and Supplementary Figure S1E). Furthermore, annotated chromatin proteins had significantly fewer missing values across replicates (Figure 1D) and smaller CVs (Supplementary Figure 1H).

To make the method applicable to rare stem cell populations, we examined how input amounts affect the performance of our method. Cell numbers between 100K to 5 Mio. correlated well with the original protocol comprising 15 Mio. cells (Pearson correlation > 0.9) and 250k to 5 Mio. cells were sufficient for stable identification rates of over 5000 proteins (Figure 1E). Notably, ChAC-DIA with as few as 10k cells still resulted in over 2000 protein identifications. Ranking proteins quantified by ChAC-DIA according to their abundance revealed specific enrichment of histones and *bona fide* naive pluripotency factors as compared to a full proteome (Figure 1F, Supplementary Figure S2A, and Supplementary Table 1).

To further assess the comprehensiveness of ChAC-DIA, we compared the results to naive mESC transcriptome data. Among approximately 13000 expressed transcripts, 487 encode proteins annotated as chromatin binders, of which 80% were identified by ChAC-DIA (Figure 1G). Among *bona fide* naive pluripotency factors, 92% were identified by ChAC-DIA. Given that not all transcripts are translated into proteins with the same efficiency, we also compared the results obtained by ChAC-DIA to a full proteome analysis covering around 7000 proteins and observed that ChAC-DIA identified the same number of known chromatin binders that were also present in the full proteome data (Supplementary Figure S2B–D). We speculated that these annotated chromatin binding proteins might be missed due to overall low expression levels. However, we found that only some of these transcripts are lowly expressed (Supplementary Figure S2E). We, therefore, checked whether these missing proteins harbor additional cellular localizations and thus might not be frequently nuclear in naive mESCs. Indeed, these missing proteins are more often annotated cytoplasmic or membrane-associated proteins (Supplementary Figure S2F). Half of the missed proteins were identified and enriched in purified cytoplasmic fractions of naive mESCs (Supplementary Figure S2G).

Taken together, our results validated ChAC-DIA as a rapid and highly accurate method for analyzing the chromatome that uses only 100–250K cells and achieves unprecedented, almost complete chromatome coverage, including low-abundant proteins.

Chromatome mapping reveals a specific enrichment of chromatin-associated proteins in ground state PSCs

To define high-confidence chromatomes of ground state PSCs and thereby assess the specificity of chromatin enrichment by ChAC-DIA, we analyzed all fractions obtained during the chromatin purification in triplicates (i.e. whole cell lysate, cytoplasmic and nuclei fractions, ChAC-DIA after 1–3 washes). In total, we identified 8567 proteins, and the

triplicates correlated well with each other ($R^2 > 0.95$). We observed that the correlation between the chromatin and nuclei fractions was weak ($R^2 = 0.66$) (Supplementary Figure S3A–D). Filtering for proteins with significantly different quantities between the fractions (ANOVA FDR < 0.05, fold change difference ≥ 1.5), resulted in 5464 proteins which explains the low correlation between the fractions. Unsupervised hierarchical cluster analysis of these proteins revealed nine distinct clusters (Figure 2A and Supplementary Table 2).

Two clusters (II and III), harboring 1141 proteins, were significantly enriched in the chromatomes (ChAC-DIA after 1–3 washes), but not in the nuclei or any other fraction. Therefore, proteins in clusters II and III comprise high-confidence chromatin binders. Importantly, well-known pluripotency proteins such as DNMT1, ESRRB, SALL4 or SOX2 are most abundant within these two clusters. Cluster II contained the highest enrichment of general chromatin-specific GO categories such as ‘nucleosome’ or ‘nucleosomal DNA binding’ (Supplementary Figure S3E and Supplementary Table 2). Euchromatic and heterochromatic proteins were equally enriched within this cluster. In cluster III, mitotic chromatin binders were over-represented, resulting in GO categories such as ‘mitotic prometaphase’. Clusters I and IV revealed significant enrichment of proteins in the nuclei fraction and a strong depletion in the chromatomes indicating that these two clusters captured nucleoplasmic proteins (Figure 2B). In line with this, well-characterized nucleoplasmic proteins such as RANGAP1 or CDK11B were categorized within these two clusters. In contrast, proteins in clusters V–IX were enriched for cytoplasm-specific GO categories (e.g. ‘Golgi membrane’, ‘structural constituent of the ribosome’ or ‘Mitochondrion’) (Supplementary Figure S3F). PCA analysis of the six different fractions confirmed that the three chromatin fractions are distinct from the nuclei fraction (Figure 2C).

Pluripotency phases are guided by distinct signaling pathways that lead to the translocation of otherwise cytoplasmic transcription factors into the nucleus (56–59). For example, naive pluripotent stem cells harbor active WNT and LIF pathways, while the GSK, FGF2 and Activin A pathways are inactive. Our data captured these features accurately, as we observed the chromatin-association of transcription factors linked to the WNT and LIF pathways, while those related to GSK, FGF2 and Activin A were mostly cytoplasmic (Figure 2D–H). For instance, β -CATENIN, the effector of WNT signaling, was equally distributed between the cytoplasmic and chromatin fractions, while being less abundant in the nuclear fraction (Figure 2D). We also observed chromatin enrichment of the LIF pathway transcription factors like KLF4 and KLF5, as well as STAT1 and STAT3, which, although being less abundant at chromatin than in the cytoplasm, still showed chromatin enrichment over the nuclear fraction (Figure 2E). In contrast, GSK, FGF2 and Activin A-related transcription factors were depleted from the chromatin fractions (Figure 2F–H). Taken together, we confirmed that ChAC-DIA selectively enriched components of the chromatome by reducing background proteins, even hard to separate mitochondrial or ribosomal proteins. This enabled the identification of not

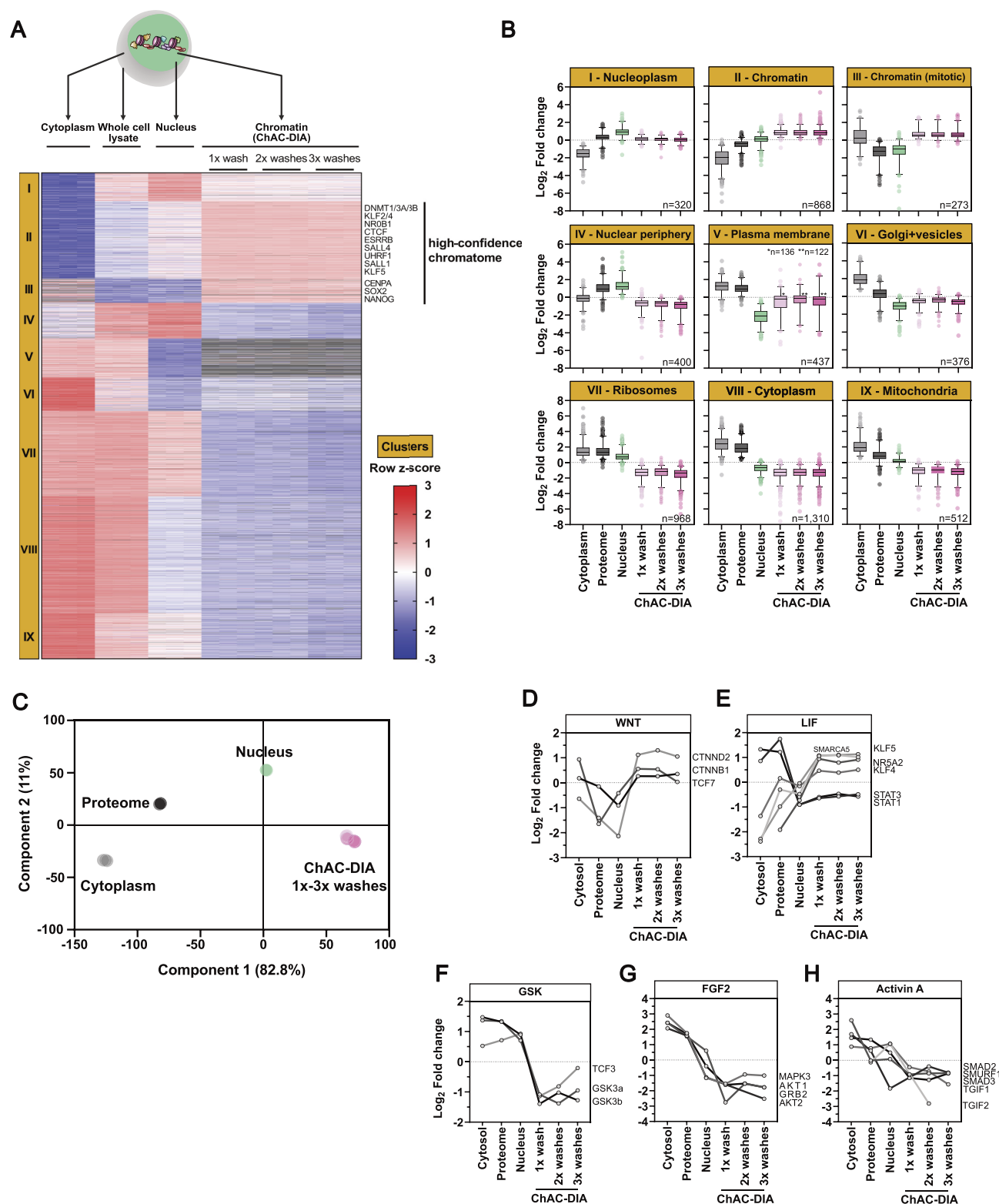


Figure 2. Chromatome mapping reveals a specific enrichment of chromatin-associated proteins in ground state PSCs. (A) Different fractions along the ChAC-DIA protocol were processed and measured. After ANOVA testing ($FDR < 0.05$, fold change difference ≥ 2) results were visualized in a heatmap generated by unsupervised hierarchical k-means clustering of z-scored intensities. In total nine clusters were identified. Proteins that are enriched only in the chromatome fractions are highlighted as the high-confidence chromatome. (B) Boxplot representation of row-scaled fold changes within each cluster. Cluster names are based on the most prominent GO-enriched terms (see Supplementary Figure 3E). (C) Principal component analysis (PCA) of the six different fractions. (D–H) Individual intensity profile plots of several proteins that are components of the WNT, LIF, GSK, FGF2 or Activin A pathways. See also Supplementary Figure S3.

only the majority of the annotated chromatinome, but the expansion of the existent GO annotations. Thus, ChAC-DIA provides a high-confidence global map of the chromatinome. Furthermore, analyzing chromatinome data in combination with the overall proteome, and proteomes derived from different cellular fractions, allowed us to dissect events such as nuclear translocation and chromatin binding of proteins related to pluripotency-regulating pathways.

Chromatinome atlas of mouse naive, formative and primed pluripotent stem cells identifies groups of chromatin proteins with distinct binding patterns

Two recent studies provided evidence that the formative phase is a discrete pluripotent state during embryonic development that is transcriptionally distinct from naive and primed pluripotency phases (1,10). To examine this further, we analyzed chromatinomes of naive, formative, and primed PSCs (Figure 3A). We observed that 1403 proteins significantly changed in the chromatinome during the differentiation of naive to formative PSCs, while the proteome revealed 1683 significantly regulated proteins (P value < 0.05 , $FC \geq 2$) (Figure 3A). In contrast, between formative and primed PSCs, only 859 proteins were significantly regulated on chromatinome level and 1451 on proteome level. This suggests a more drastic reorganization of the chromatinome during the transition from naive to formative pluripotency.

Next, we analyzed the chromatinome changes based on a list of PSC phase-specific factors that we derived from the literature (Supplementary Table 3) (1,4,6,7,10,13,15,33,54,60–67). ChAC-DIA data confirmed that the abundance of the core pluripotency circuitry (OCT4, MYC, SOX2 and SALL4) is maintained throughout pluripotency; whereas state-specific markers displayed phase-dependent selective enrichment in the chromatinome (Figure 3B–D and Supplementary Table 3). The naive chromatinome was characterized by high levels of REX1, ESRRB, KLF4 and TET2 while the *de novo* methyltransferases DNMT3A and DNMT3B, OTX2, and OCT6 (or POU3F1), were highly enriched in the formative chromatinome (Figure 3C). We observed a slight enrichment of lineage-specific transcription factors such as NES as early as the formative state.

In contrast to the formative chromatinome, the primed chromatinome was characterized by lower levels of early post-implantation-specific proteins like DPPA4 (15) and OCT6 (7) and higher levels of *bona fide* primed-specific transcription factors such as SOX1 (10) and SALL3 (60). Similarly, naive factors like ESRRB, HMCES and TET2 were further decreased in the primed chromatinome while lineage-specific factors such as RAI1 and SIX6 (Figure 3D) were significantly enriched, which fits the partially fate-determined identity of primed PSCs. Among the primed-specific chromatin constituents, several histone H1 variants and high mobility group (HMG) proteins were also observed. The enrichment of these proteins governing chromatin structure and compaction could in part account for the previously described reduced chromatin plasticity and accessibility at the primed phase (1,5,10). Although major chromatinome changes were already established at the formative state, these results demonstrate that formative and primed

pluripotency are characterized by distinct chromatin landscapes.

These findings point to gradual chromatin recruitment or eviction of pluripotency governing factors during naive to primed transition. Interestingly, we observed similar chromatin-enrichment patterns for proteins related to epigenetic regulation, transcriptional regulation, and chromatin remodeling, as well as hundreds of zinc finger proteins with mostly unknown functions in pluripotency regulation (Supplementary Figures S4–S9). Approximately 70% of proteins harboring a zinc finger domain significantly change between naive and primed pluripotency, which fits well with the recently reported zinc finger protein-driven regulation of transposable elements during early embryonic development (68,69).

In summary, we provide the first systematic and near-comprehensive chromatinome atlas of naive, formative, and primed PSCs (Supplementary Figures S4–S9, Supplementary Table 3) and provide an interactive web application for easy access to the data set (Supplementary Figure S10). We show that the chromatinome reflects distinct features of pluripotency phases and a tightly regulated pluripotency phase transition process.

Identification of novel pluripotency phase-specific proteins through chromatinome analysis

Using the comprehensive chromatinome dataset we next sought to pinpoint novel pluripotency phase-specific proteins that bind chromatin in a similar manner to *bona fide* phase-specific proteins such as TBX3, OCT6 or SOX1 (Figure 4A–C). To achieve this, we ranked proteins according to their fold change between each pluripotency phase and observed differential enrichments of proteins associated with H3K4me3 or H3K9me3. For instance, we found that QSER1 increases at chromatin from naive to formative and decreases from formative to primed (Figure 4D, E). Previous studies have shown that QSER1, along with TET1, protects bivalent promoters from *de novo* methylation in human ESCs. (70). Our chromatinome data shows that QSER1 and the *de novo* methyltransferases peak at the formative phase, potentially indicating a conserved role of QSER1 in mouse PSCs. Other H3K4me3-related proteins are preferentially enriched in the naive chromatinome (e.g. KAT6B) or the primed chromatinome (e.g. KAT6A, ZNF800).

Among the H3K9me3-associated proteins, we observed that two trimethyltransferases of H3K9, SUV39H1 and SUV39H2, increase at chromatin from naive to formative, while SUV39H1 decreases from formative to primed. To test whether SUV39H1/2 inhibition by their specific inhibitor Chaetocin could provide evidence for increased catalytic activity of these enzymes in formative vs naive pluripotent stem cells, we treated wild-type PSCs with or without Chaetocin and compared to Suv39h double knockout mESCs in both naive and formative states. We then quantified H3K9me3 abundance by western blot, which revealed lower levels of H3K9me3 in formative PSCs upon 0.1 μ M Chaetocin treatment than in naive PSCs (Figure 4F). Our results suggest increased catalytic activities of SUV39H1/2 in formative PSCs, consistent with the increased chromatin binding of both enzymes revealed by

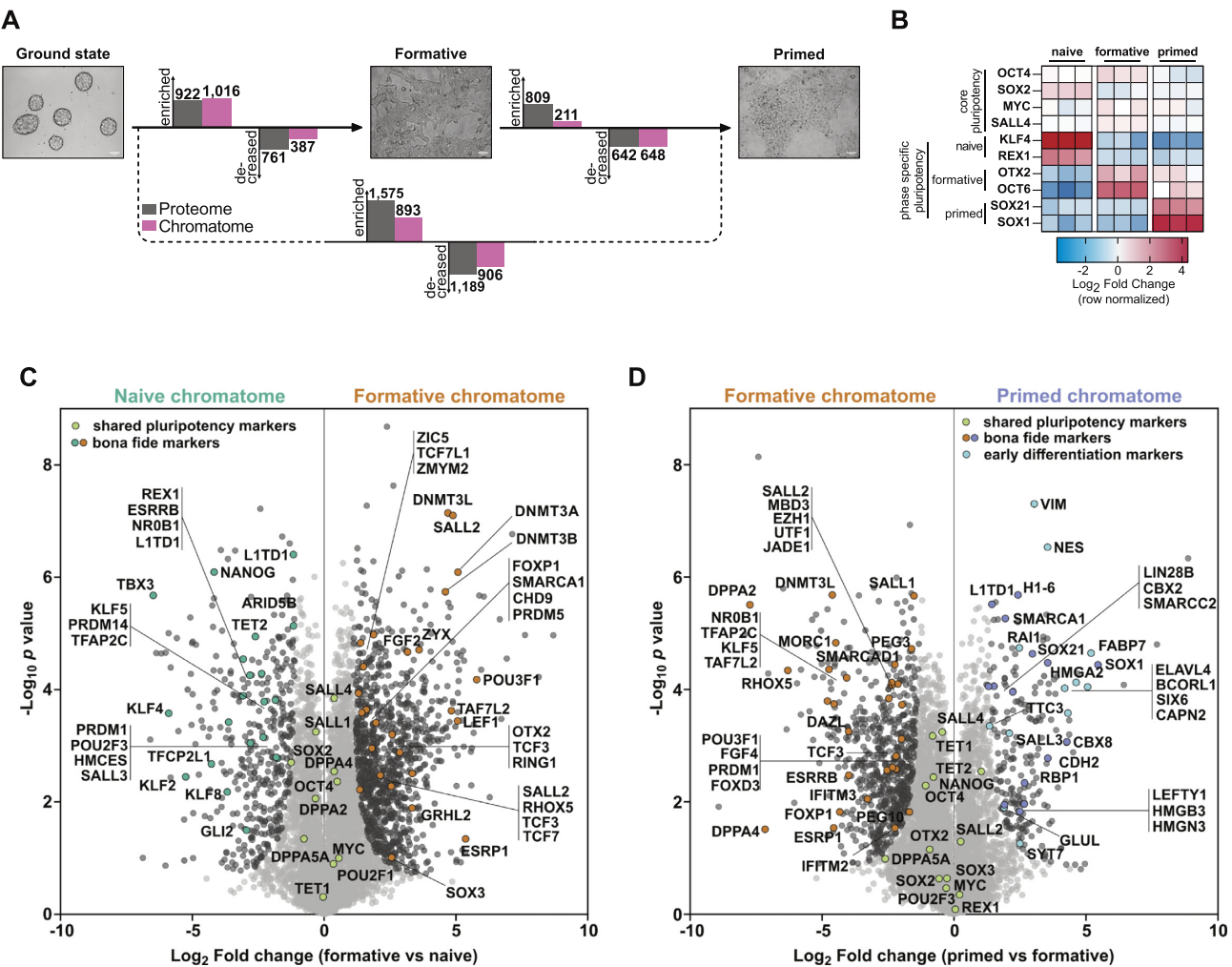


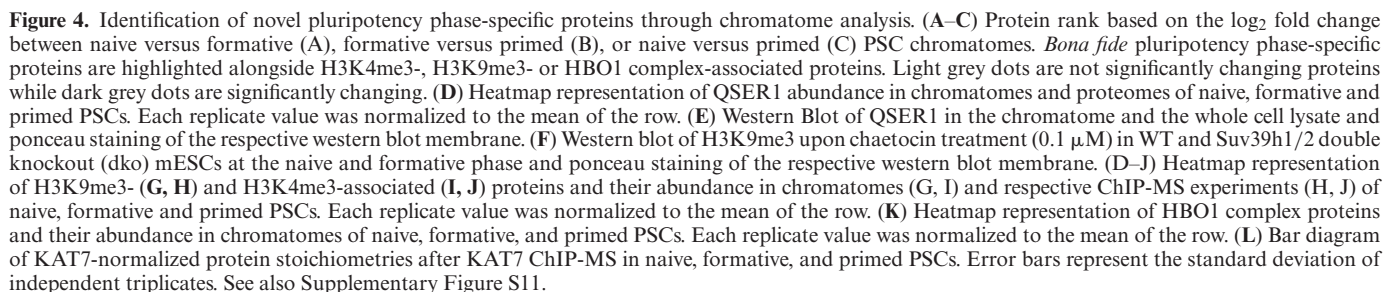
Figure 3. Chromatome atlas of mouse naive, formative, and primed pluripotent stem cells identifies groups of chromatin proteins with distinct binding patterns. (A) Schematic representation of compared cell lines and total significant changes between respective proteomes and chromatomes (Student's *t*-test, *P* value < 0.05, FC ≥ 2). (B) Heatmap representation of *bona fide* pluripotency factors. Fold changes are row-normalized by subtracting the mean log₂ fold-change from each value. (C, D) Volcano plots of chromatomes based on Student's *t*-test displayed in (A). Light grey dots: not significantly enriched proteins. Black dots: significantly enriched proteins. Green dots: shared pluripotency factors. Blue dots: early differentiation markers. *n* = 3 biological replicates, meaning independently cultured/differentiated PSCs of the same genetic background. See also Supplementary Figures S4–S10.

ChAC-DIA. We further observed a SUV39H1-like pattern for DNMT3L and ZNF462. Proteins that continuously decreased in their chromatin association from naive to primed included LIRE1 and PHF11, while FLYWCH1, SUV39H2, UHRF2, CBX3, CBX5 and MKI67 increased from naive to primed.

To validate the global chromatome change of the described H3K4me3- and H3K9me3-associated proteins, we performed ChIP-MS of both histone PTMs and compared the ChAC-DIA results to the ChIP-MS data (Figure 4G–J). We observed a high level of similarity between the two datasets for well-described H3K4me3- or H3K9me3-associated proteins. However, some proteins showed slightly different levels in the global chromatome compared to specific regions with H3K9me3. A good example is FLYWCH1, a low-abundant chromatin binder at H3K9me3-rich regions which has not been detected in previous chromatome or proteome studies of PSCs (60,71). FLYWCH1 chromatin binding increases along with H3K9me3 from

naive to primed PSCs (Supplementary Figure S6C) but is most abundant at H3K9me3 sites in formative PSCs, suggesting alternative mechanisms of chromatin association beyond H3K9me3 binding.

We further observed several chromatin-associated complexes among these phase-specific proteins (Supplementary Figure S9). One interesting example is the HBO1 complex, which acetylates several lysines at histones H3 and H4 and by this co-regulates the origin of replication licensing and MCM complex formation (72,73). The specificity of the complex is determined by the association of the mutually exclusive accessory subunits JADE1/2/3 and BRPF1/3 (74). Our chromatome data suggests that the core HBO1 complex (KAT7, ING4/5, MEAF6) remains at a constant level from naive to primed, while the accessory subunits are dynamically regulated. JADE1, BRPF1 and BRPF3 were mostly enriched in the naive chromatome, while JADE3 peaked at the formative phase and JADE2 peaked in the primed phase (Figure 4A–C, K). Since global



chromatome changes might not reflect the actual changes within the HBO1 complex, we calculated the complex stoichiometries after performing ChIP-MS on the HBO1 catalytic subunit KAT7 (Figure 4L and Supplementary Figure S11). The ChIP-MS data revealed that KAT7 indeed interacts in a stable ratio with ING4/5 and MEAF6, but selectively interacts with JADE1/2/3 and hardly with BRPF1. This latter finding might hint towards a cell-type dependent BRPF1/3 interaction with KAT7 or more frequent interactions of BRPF1/3 with other complexes (e.g. MOZ/MORF complex, Supplementary Figure S9). The switch between JADE1/2/3 across pluripotency implies that the complex might target different lysines in a pluripotency phase-specific manner.

Collectively, we used the comprehensive chromatome dataset to identify novel pluripotency phase-specific proteins that bind chromatin in a manner similar to known phase-specific proteins. We found that especially proteins associated with H3K4me3 and H3K9me3 show phase-specific enrichment patterns and that these patterns can be confirmed by ChIP-MS.

Determination of relative chromatin binding reveals regulatory changes along pluripotency phases

Next, we correlated the transcriptome changes during the naive to formative transition (75) with the respective proteome and chromatome changes. As expected and previously reported (60,76,77) the proteome showed a moderately positive correlation with the transcriptome (Figure 5A), due to mechanisms regulating translation and protein stability. Consequently, transcriptome and chromatome showed the lowest correlation (Figure 5B) indicating that transcriptional data can only provide limited coverage of regulatory chromatin changes. Interestingly, the comparison of proteome and chromatome changes revealed also a moderate positive correlation (Figure 5C), pointing to mechanisms controlling chromatin binding and dissociation. In line with these observations, proteins related to active signaling pathways in postimplantation pluripotency like the FGF2, Activin A, and Notch pathways were differentially enriched in the chromatome, while they changed neither on transcriptome nor on proteome level.

Proteome-independent changes in the chromatome contain valuable information and point to either altered chromatin affinity or subcellular localization and availability of individual proteins (Figure 5D). We, therefore, computed proteome normalized chromatome changes to estimate the relative changes in chromatin binding. We subtracted the Log₂ chromatome-intensity of a protein from its mean Log₂ proteome intensity across triplicates and subsequently filtered for significant proteins by ANOVA testing (FDR < 0.05 and FC > 2) (Figure 5E and Supplementary Table 3). Based on our differential chromatin fraction analysis, we defined high-confidence chromatin binders as proteins that are significantly enriched in the chromatome over the proteome.

We observed that 1518 proteins significantly changed in relative chromatin binding from naive to primed pluripotency. Hierarchical clustering yielded five distinct clusters

harboring proteins with different trends in relative chromatin binding across pluripotency phases. GO analysis of these five clusters against the background of total identified proteins revealed distinct functional categories (Benjamini-Hochberg FDR < 0.05) (Figure 5F and Supplementary Table 3). In the cluster of proteins with a peak in relative chromatin binding at the formative phase (cluster II) categories related to signaling pathways like ‘β-catenin degradation’ or ‘RAF activation’ were enriched (Figure 5F). Importantly, cluster III showed an increased relative chromatin binding at the formative and primed phases and was enriched for categories associated with a repressive chromatin state like ‘heterochromatin’ or ‘transcription corepressor activity’. More specifically, this cluster harbored essential heterochromatic proteins such as SETDB1, SETDB2, KAP1, CBX3 and CBX5 suggesting a functional relation of their formative and primed specific enrichment to the incremental heterochromatinization towards the exit from pluripotency. Interestingly, this cluster III was also enriched for GO categories related to ‘SUMOylation of transcription factors’, ‘SUMOylation of chromatin organization proteins’, and SUMOylation-dependent ‘PML bodies’. In line with this observation, SUMOylation was reported to regulate heterochromatinization in naive mouse PSCs (78). Notably, histone H1.0, whose function in chromatin compaction depends also on its SUMOylation (79), peaked in its relative chromatin binding at the primed phase. These results suggest that besides the binding of classical heterochromatin factors, SUMOylation also contributes to heterochromatin formation at the formative and primed phases. Among the proteins with decreasing relative chromatin binding (clusters IV and V) are enzymes involved in DNA and histone demethylation or DNA repair like TDG, APOBEC3, NTHL1, KDM4C and KDM6A. Thus, lower levels of these proteins would translate into an increase of repressive epigenetic marks, which is expected to promote repressive chromatin states and reduce chromatin plasticity.

These findings are indicative of an increased chromatin affinity of heterochromatic proteins at the formative and primed phases which may enhance in turn further heterochromatinization and prepare pluripotent stem cells for differentiation.

The chromatome of conventionally cultured human ESCs is most similar to the mouse primed state

Previous reports compared the epigenome, transcriptome, and proteome of conventional human ESCs (hESCs) with mouse PSCs and have shown that hESCs are more similar to post-implantation mouse PSCs (34,60,80,81). Here, we used our method to examine the correspondence between different pluripotency states of hESCs and mouse PSCs. A Venn diagram representation of the high-confidence chromatomes for all three mouse PSCs and hESCs revealed an overlap of approximately 75% (Figure 6A and Supplementary Table 4). The strongest overlap was between proteins related to chromatin remodeling, histone modifications, and developmental processes (Supplementary Figure S12A). A PCA of the high-confidence chromatomes resulted in a clear separation of all three mouse PSCs from hESCs on PC1. PC2 in turn separates hESCs and

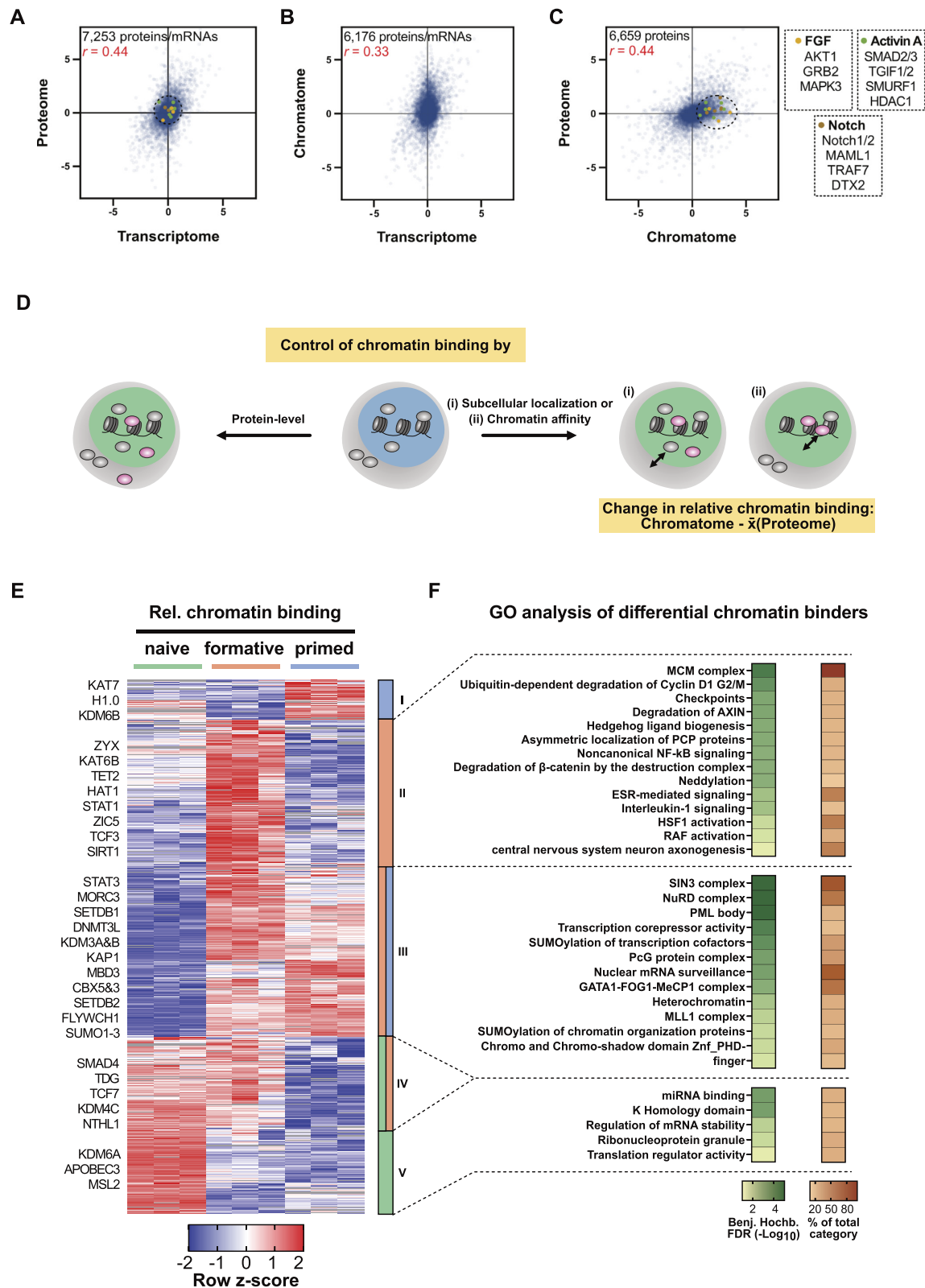


Figure 5. Relative chromatin binding reveals higher chromatin affinity of heterochromatic proteins in formative and primed PSCs. (A–C) Correlations of transcriptomes (ArrayExpress: E-MTAB-6797), proteomes, and chromatomes of formative vs naive PSCs of isogenic background (J1). Only proteins/mRNAs that were identified in both compared data sets are displayed. Pearson correlation coefficients are indicated in red. (D) Schematic representation of the relative chromatin binding concept. (E) Row z -scored relative chromatin binding changes between naive, formative, and primed PSCs filtered for ANOVA significant changes ($FDR < 0.05$, $FC \geq 2$) and high-confidence chromatin binders. The relative chromatin binding was computed by subtracting the FC on chromatome level by the mean proteome FC of either formative versus naive or primed versus formative PSCs, respectively. (F) GO analyses of proteins enriched in clusters II, III or V of the hierarchical clustering from (E). As a comparison, the whole set of identified proteins was utilized.

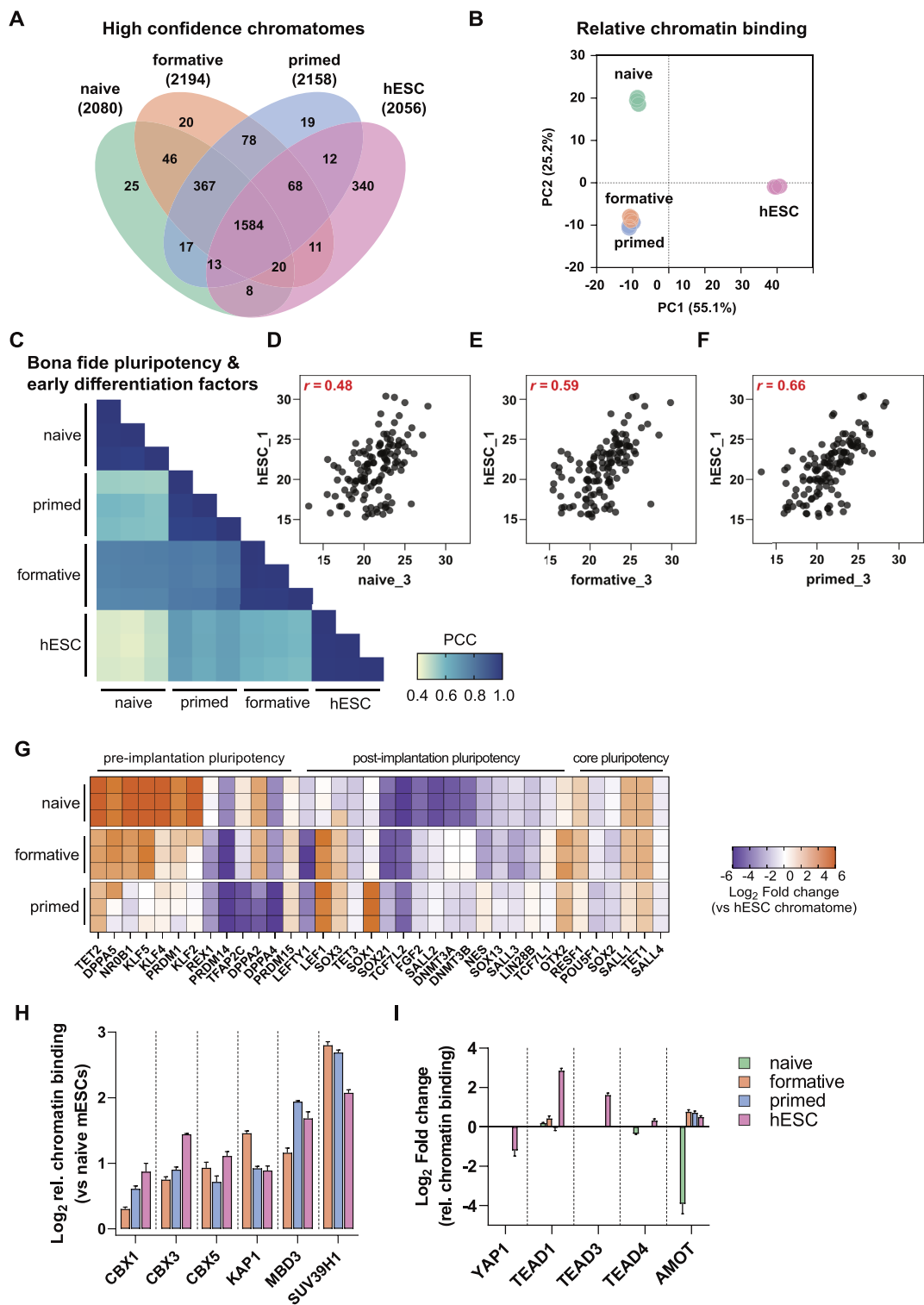


Figure 6. The chromatome of conventionally cultured human ESCs is most similar to the mouse primed state. (A) Venn diagram of high-confidence chromatomes in all tested cell lines. The high-confidence chromatome was defined by a Student's T-test between each cell line's chromatome vs proteome (Student's *t*-test, *P* value < 0.05 and FC ≥ 1.5). (B) PCA of the high-confidence chromatomes of the tested four cell lines on relative chromatin binding level. (C) Pearson correlations of chromatomes filtered for literature-derived *bona fide* pluripotency and differentiation factors. (D–F) Scatter plots of one replicate of hESCs versus naive (D), formative (E) or primed PSCs (F) and Pearson correlation coefficient from (C) are displayed in red. (G) hESC-normalized chromatomes from each mouse PSC to hESCs in log₂. The selection comprises *bona fide* pluripotency factors. (H) Relative chromatin bindings of a selection of heterochromatic proteins after normalization to their respective relative chromatin bindings in naive mESCs. The bars represent mean values and the error bar is based on the standard error of the mean. SUV39H1 was not identified in the full proteome of naive mESCs which is why the relative chromatin binding was imputed by a fixed value: 0. (I) Relative chromatin bindings of proteins related to the HIPPO signaling pathway in all analyzed cell lines. Bars represent mean values and the error bar is based on the standard error of the mean. See also Supplementary Figure S12.

mouse formative and primed PSCs from mouse naive PSCs (Figure 6B).

To further dissect whether hESCs correspond more to the early or late mouse post-implantation stage, we computed correlations between the chromatomes of all four cell lines selected for *bona fide* pluripotency and early differentiation factors (Figure 6C). We noted an incremental increase in the correlation of hESCs with naive, formative, and primed PSCs (Pearson, $r = 0.48$ for naive, 0.59 for formative, and 0.66 for primed PSCs) (Figure 6D–F), while chromatomes of formative and primed PSCs correlated better to each other (Pearson, $r = 0.78$) than to naive PSCs (Pearson, $r = 0.74$ and $r = 0.57$, respectively). We observed similar differences on the relative chromatin binding and total proteome levels (Supplementary Figure S12B, C).

For an in-depth view of pluripotency factors and their contribution to cell identity, we computed the chromatome difference between a given mouse PSC-line and hESCs for each *bona fide* pluripotency factor (Figure 6G and Supplementary Table 4). A step-wise loss of pre-implantation pluripotency markers was observed from naive to primed PSCs with some remarkable exceptions; TFAP2C, DPPA2, DPPA4 and PRDM14 were more similar in their chromatin abundance between both naive and formative PSCs and hESCs. These proteins are indicative of germline competence, a capability that mouse formative PSCs and conventional hESCs harbor, while mouse naive PSCs first require differentiation to the formative state (33,82–85). Moreover, REX1, a well-characterized naive pluripotency and germline marker, was more strongly associated with the hESC chromatome than mouse formative and primed PSCs, likely reflecting the more heterogeneous nature of hESCs or species-specific differences (86). In a PCA based on these *bona fide* pluripotency factors only, mouse formative PSCs were even further separated from primed PSCs but not from naive PSCs (Supplementary Figure S12D, F). A scatter plot of the protein loading values uncovered that the main causes of this separation were naive pluripotency factors such as NR0B1, KLF2 and KLF4 (Supplementary Figure S12E). Thus, these naive factors were less associated with chromatin in hESCs and mouse primed PSCs than formative or naive PSCs. Conversely, post-implantation pluripotency factors contributed to the higher similarity between hESCs and primed PSCs. Of note, we did not observe differences in the chromatin association of the core pluripotency circuitry such as OCT4 or SALL4 (Supplementary Figure S12F, G).

The relative chromatin binding of well-known heterochromatic proteins (CBX1, CBX3, CBX5, KAP1, MBD3 and SUV39H1) revealed similar high levels in hESCs as in formative and primed PSCs (Figure 6H, see also Figure 5). An increased relative chromatin binding of heterochromatic proteins seems thus to be a common hallmark of post-implantation PSCs, indicating that higher chromatin compaction involves enhanced chromatin association of heterochromatic proteins. However, we also observed notable differences between hESCs and mouse post-implantation PSCs, like for the HIPPO signaling pathway (Figure 6I). This pathway is highly active in pluripotent epiblast cells and upon its activation the downstream proteins YAP1 and TAZ are kept cytoplasmic (56,87). Inter-

estingly, we observed YAP1 and TAZ only in the full proteome fractions, except for hESCs where YAP1 was also present in the chromatin fraction. This was in agreement with a higher relative chromatin binding of the YAP1 co-factors TEAD1/3/4 in hESCs, likely suggesting a more inactive state of the HIPPO pathway in hESCs than in closely related mouse pluripotency phases.

In summary, the conventional hESC chromatome is similar to mouse PSC chromatomes reflecting post-implantation, particularly the mouse primed stage. This is largely due to lower levels of naive-specific transcription factors in these chromatomes. However, hESCs differ from mouse primed PSCs in the chromatin association of e.g. essential germline factors and the HIPPO pathway, indicating that hESCs have some similarities to mouse formative-like chromatomes and that the HIPPO pathway is regulated differently between mouse and human PSCs.

DISCUSSION

Previous studies have established methods for chromatin purification and measurement (29–32,88,89). These techniques, however, require large numbers of cells and have limited accuracy and comprehensiveness, often failing to detect low-abundant proteins such as regulatory factors. In this study, we combined a stringent and simple chromatin preparation strategy of crosslinked nuclei with an additional purification step by protein aggregation capture (PAC) and optimized DIA-based MS. Our method only requires three hours of experimental hands-on time and confidently reduces non-chromatin proteins while identifying more than twice the number of DNA-binding proteins compared to other methods in half of the MS acquisition time (54,90). In addition, recent deep neural network-based computational processing of DIA measurements without a peptide library (direct DIA) can now outperform DDA in accuracy and comprehensiveness (47,50,91,92). Thus, our direct DIA measurements additionally decreased instrument time, while providing a near-complete chromatome coverage. However, it is possible that a library-based analysis would increase the current chromatome depth further, and may represent a potential future opportunity.

The datasets generated here allowed us to perform several different types of analysis. Given that ChAC-DIA selectively enriched components of the chromatome, we were able to assemble a high-confidence global map of the chromatome. By comparing chromatome and proteome data, including proteomic data derived from different cellular fractions, for different pluripotency phases, we identified proteins affected by nuclear translocation or chromatin binding. For example, we observed chromatin enrichment of cytoplasmic transcription factors such as those involved in WNT and LIF pathways, and not GSK, FGF2 and Activin A pathways in naive PSCs, which has implications for their role in pluripotency regulation. Furthermore, normalizing the chromatome to protein levels enabled a global assessment of changes in relative chromatin binding which may be caused by either altered chromatin affinity and accessibility or differential subcellular localization and availability. Our method thus enables accurate and comprehensive chromatome and relative chromatin binding measure-

ments despite limited cell numbers, making it ideally suited for analyzing minute tissue samples or rare subpopulations of cells.

Additionally, ChAC-DIA enables the quantification of low-abundant transcriptional or epigenetic regulators, and we identified several low-abundant chromatin binders that are pluripotency phase-specific. Besides well-described factors, we find many phase-specific proteins with still unknown functions in pluripotency regulation. Given their phase-specific chromatin association, many of them are likely to contribute to the regulation of cellular identity. One such example is EZHIP which was only identified in the formative phase. EZHIP was recently described to inhibit H3K27me3 by mimicking the H3K27M oncohistone and thus preventing the PRC2 complex from spreading along chromatin (93,94). Bulk levels of H3K27me3 are known to be downregulated from naive to primed pluripotency while bivalent sites harboring H3K4me3 and H3K27me3 are enriched (10,95). In our chromatome data set, we observed that EZH1 increases at chromatin between the naive and formative PSCs which does not fit a global down-regulation of H3K27me3. Interestingly, this goes along with an increase in EZHIP in the formative chromatome implying a possible role of PRC2 inhibition or redirection to other regions by EZHIP in formative PSCs. Moreover, low-abundant epigenetic writers such as SUV39H1/2, SUV420H1/2, SETDB2 or TET1–TET3 featured phase-specific enrichment at chromatin. Remarkably, all three TET proteins showed a distinct redistribution along the exit from pluripotency, starting with TET1 and TET2 being most abundant in the naive state and TET3 being mostly chromatin-associated in the primed state. This was also observed in conventional hESCs where TET2 and TET1 are even less associated with chromatin than in mouse primed PSCs.

The chromatome correlates weakly with the transcriptome and proteome and is, therefore, an important complement to previous studies of pluripotency. Our results provide a system-wide view of pluripotency by offering a chromatome atlas with specifically enriched proteins for each analyzed pluripotency phase. Our observations are in line with the recent finding that formative pluripotency is an essential state which is transcriptionally and epigenetically distinct from naive pluripotency and to a smaller degree also from primed pluripotency (1,3,10,13,62,96). The underlying chromatome changes fit in with the phased progression model of pluripotency (3). Moreover, formative and primed PSCs share the majority of open chromatin sites while there is little overlap between formative and naive PSCs (1). Our data support this observation by showing that the chromatome undergoes larger changes from naive to formative, than from formative to primed pluripotency. The chromatin composition is further reorganized between formative and primed PSCs, mainly driven by transcription factors triggering early differentiation as well as histone H1 and HMG variants guiding chromatin compaction. The histone H1 chromatin enrichment is in agreement with an increased relative chromatin binding of SUMO1–3 and SUMOylating enzymes of chromatin organizing proteins. SUMOylation of histone H1 was recently described as a mechanism for heterochromatinization in ESCs (79), thus

suggesting a role for SUMOylation in further chromatin compaction from formative to primed pluripotency. An increased relative chromatin binding was observed for additional heterochromatic proteins, such as KAP1 and CBX3, at the formative and primed phases. Surprisingly, this increased relative chromatin binding of heterochromatic proteins was conserved in conventional hESCs. We conclude that heterochromatic proteins not only become more abundant towards the exit from pluripotency, but also have a stronger affinity for chromatin. One potential explanation for this enhanced affinity is that the increase of repressive epigenetic marks during the transition from naive to primed pluripotency provides additional binding sites for heterochromatic proteins, thereby giving rise to a more repressive chromatome signature.

Conventionally cultured hESCs are reminiscent of mouse primed PSCs regarding their epigenome, transcriptome and underlying signaling cues (56,80). Still, human embryonic development comprises pluripotent phases that differ in length and growth conditions when compared to mouse (1,3,4,97–99). It remains unclear whether hESCs are the direct counterpart of mouse primed PSCs and to what extent they share unique features with mouse formative PSCs. A quantitative comparison of the high-confidence chromatomes revealed that mouse primed PSCs correlated best with hESCs. Of note, a comparable correlation range was previously described on transcriptome and full proteome levels (33,60). In our hands, the correlation between hESCs and mouse primed PSCs increased even further when only *bona fide* pluripotency and early differentiation factors were considered. Here, chromatome-levels of naive pluripotency factors were the main difference between mouse primed PSCs and hESCs on the one side and mouse formative and naive PSCs on the other side. One major distinction between hESCs and mouse primed PSCs was the high chromatin association of essential germline factors like DPPA2, PRDM14 and TFAP2C in hESCs which resembles formative pluripotency in the mouse. This finding may explain the differential developmental capacities of hESCs and mouse primed PSCs. In addition, the hESC chromatome provided evidence for a less active HIPPO pathway compared to all three mouse PSCs, likely reflecting more species-specific signaling mechanisms.

Our study sheds light on the important question of whether cell identity-defining transcription factors coexist, suggesting an ongoing competition with each other (100,101), or abruptly change across pluripotency phases (4). For all three phases and especially for the formative phase we observed that transcription factors were gradually recruited or evicted from chromatin. For instance, OTX2, a key transcription factor of formative pluripotency (15,102), peaks in abundance at the formative state, but is still associated with chromatin in naive and primed PSCs. Thus, our findings support the model of coexisting phase-specific transcription factors that ultimately define cellular identity if a certain critical threshold is exceeded.

In conclusion, we present a robust chromatin proteomics method to detect changes in the abundance and affinity of even low-abundant proteins. We offer a rich resource for the proteomes, chromatomes and relative chromatin bindings in mouse naive, formative and primed PSCs, as well

as hESCs that are a basis for identifying and investigating novel regulatory mechanisms of pluripotency. Further investigations of candidate phase-specific proteins highlighted herein may help detangle the connection between pluripotency and lineage priming and support clinical applications of iPSCs. The dramatically improved sensitivity now makes it possible to also study rare subpopulations of cells. The comprehensive capture of chromatomes and chromatin affinities provides a deep and unbiased view of regulatory events underlying the establishment, maintenance, and change of cellular identity.

DATA AVAILABILITY

The mass spectrometry proteomics data has been deposited to the ProteomeXchange Consortium via the PRIDE (103) partner repository with the dataset identifiers PXD034448 for chromatomes and proteomes and PXD039556 for ChIP-MS. To make the proteome and chromatome files better comprehensible, they have been assigned to experiments (Raw data list, see PXD034448). Source data are provided in this paper.

The used RNA-Seq dataset is derived from the ArrayExpress with the following accession code: E-MTAB-6797.

CODE AVAILABILITY

The underlying custom code for the provided web application, accessible on <https://pluripotency.shinyapps.io/Chromatome.Atlas/>, can be found at the Github repository https://github.com/ugur-enes/pluripotency_chromatome_shinyapp.git.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Igor Paron and Tim Heymann for excellent MS technical assistance, Dr Johannes Bruno Müller-Reif for columns, and Jeannette Koch and Ruzica Barisic for technical assistance. We thank Dr Christopher B. Mulholland, Dr Michael D. Bartoschek as well as Dr Carsten Marr for valuable and insightful scientific discussions. We thank Dr Vadim Demichev for generously providing access to the most recent version of DIA-NN and for answering questions on setting up DIA-NN. We thank Dr Ludovic Vallier and Dr Derk ten Berge for kindly providing mEpiSCs. E.U. gratefully acknowledges the International Max Planck Research School for Molecular Life Sciences (IMPRS-LS) and the research training group 1721 (RTG 1721) for training and support.

Author contributions: E.U. and M.W. designed the study; H.L., M.W. and M.M. supervised the study. E.U. performed all experiments and MS data analysis, and E.U., H.L. and M.W. interpreted the data. E.U., H.L., M.W. and M.M. wrote the manuscript and prepared the figures. A.d.I.P. established the same culture conditions for human and mouse ESCs and conducted the cell culture under the supervision of M.D. W.Q. performed western blots and A.I. performed

KAT7-ChIP-MS. E.U. and S.B. programmed the web application. All authors discussed the results, then read and approved the manuscript.

FUNDING

Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) [213249687 – SFB1064 to H.L. (A17) and S.B. (A22)]. Funding for open access charge: Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) [213249687 – SFB1064 to H.L. (A17)].
Conflict of interest statement. None declared.

REFERENCES

- Kinoshita, M., Barber, M., Mansfield, W., Cui, Y., Spindlow, D., Stirparo, G.G., Dietmann, S., Nichols, J. and Smith, A. (2021) Capture of mouse and human stem cells with features of formative pluripotency. *Cell Stem Cell*, **28**, 453–471.
- Takahashi, S., Kobayashi, S. and Hiratani, I. (2018) Epigenetic differences between naïve and primed pluripotent stem cells. *Cell. Mol. Life Sci.*, **75**, 1191–1203.
- Smith, A. (2017) Formative pluripotency: the executive phase in a developmental continuum. *Development*, **144**, 365–373.
- Boroviak, T., Loos, R., Lombard, P., Okahara, J., Behr, R., Sasaki, E., Nichols, J., Smith, A. and Bertone, P. (2015) Lineage-specific profiling delineates the emergence and progression of naïve pluripotency in mammalian embryogenesis. *Dev. Cell*, **35**, 366–382.
- Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T. and Misteli, T. (2006) Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell*, **10**, 105–116.
- Neagu, A., van Genderen, E., Escudero, I., Verwegen, L., Kurek, D., Lehmann, J., Stel, J., Dirks, R.A.M., van Mierlo, G., Maas, A. *et al.* (2020) In vitro capture and characterization of embryonic rosette-stage pluripotency between naïve and primed states. *Nat. Cell Biol.*, **22**, 534–545.
- Morgani, S., Nichols, J. and Hadjantonakis, A.-K. (2017) The many faces of pluripotency: in vitro adaptations of a continuum of in vivo states. *BMC Dev. Biol.*, **17**, 7.
- Melcer, S., Hezroni, H., Rand, E., Nissim-Rafinia, M., Skoultschi, A., Stewart, C.L., Bustin, M. and Meshorer, E. (2012) Histone modifications and lamin A regulate chromatin protein dynamics in early embryonic stem cell differentiation. *Nat. Commun.*, **3**, 910.
- Zylicz, J.J., Dietmann, S., Günesdogan, U., Hackett, J.A., Cougot, D., Lee, C. and Surani, M.A. (2015) Chromatin dynamics and the role of G9a in gene regulation and enhancer silencing during early mouse development. *Elife*, **4**, e09571.
- Wang, X., Xiang, Y., Yu, Y., Wang, R., Zhang, Y., Xu, Q., Sun, H., Zhao, Z.-A., Jiang, X., Wang, X. *et al.* (2021) Formative pluripotent stem cells show features of epiblast cells poised for gastrulation. *Cell Res.*, **31**, 526–541.
- Hackett, J.A. and Surani, M.A. (2014) Regulatory principles of pluripotency: from the ground state up. *Cell Stem Cell*, **15**, 416–430.
- Nichols, J. and Smith, A. (2009) Naïve and primed pluripotent states. *Cell Stem Cell*, **4**, 487–492.
- Kalkan, T., Olova, N., Roode, M., Mulas, C., Lee, H.J., Nett, I., Marks, H., Walker, R., Stunnenberg, H.G., Lilley, K.S. *et al.* (2017) Tracking the embryonic stem cell transition from ground state pluripotency. *Development*, **144**, 1221–1234.
- Xiang, Y., Zhang, Y., Xu, Q., Zhou, C., Liu, B., Du, Z., Zhang, K., Zhao, B., Wang, X., Gayen, S. *et al.* (2020) Epigenomic analysis of gastrulation identifies a unique chromatin state for primed pluripotency. *Nat. Genet.*, **52**, 95–105.
- Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T. and Wysocka, J. (2014) Reorganization of enhancer patterns in transition from naïve to primed pluripotency. *Cell Stem Cell*, **14**, 838–853.
- Chovanec, P., Collier, A.J., Krueger, C., Várnai, C., Semprich, C.I., Schoenfelder, S., Corcoran, A.E. and Rugg-Gunn, P.J. (2021) Widespread reorganisation of pluripotent factor binding and gene regulatory interactions between human pluripotent states. *Nat. Commun.*, **12**, 2098.

17. Imhof, A. and Bonaldi, T. (2005) 'Chromatomics' the analysis of the chromatome. *Mol. Biosyst.*, **1**, 112–116.
18. Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H. and Smith, A. (1998) Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell*, **95**, 379–391.
19. Loh, Y.-H., Wu, Q., Chew, J.-L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J. *et al.* (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.*, **38**, 431–440.
20. Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M. and Yamanaka, S. (2003) The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell*, **113**, 631–642.
21. Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A. *et al.* (2007) Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat. Cell Biol.*, **9**, 625–635.
22. Ji, X., Dadon, D.B., Abraham, B.J., Lee, T.I., Jaenisch, R., Bradner, J.E. and Young, R.A. (2015) Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 3841–3846.
23. Alajem, A., Biran, A., Harikumar, A., Sailaja, B.S., Aaronson, Y., Livyatan, I., Nissim-Rafinia, M., Sommer, A.G., Mostoslavsky, G., Gerbasi, V.R. *et al.* (2015) Differential Association of Chromatin Proteins identifies BAF60a/SMARCD1 as a regulator of embryonic stem cell differentiation. *Cell Rep.*, **10**, 2019–2031.
24. Rafiee, M.-R., Girardot, C., Sigismondo, G. and Krijgsvel, J. (2016) Expanding the circuitry of pluripotency by selective isolation of chromatin-associated proteins. *Mol. Cell*, **64**, 624–635.
25. Villaseñor, R., Pfandler, R., Ambrosi, C., Butz, S., Giuliani, S., Bryan, E., Sheahan, T.W., Gable, A.L., Schmolka, N., Manzo, M. *et al.* (2020) ChromID identifies the protein interactome at chromatin marks. *Nat. Biotechnol.*, **38**, 728–736.
26. Kloet, S.L., Makowski, M.M., Baymaz, H.I., van Voorthuisen, L., Karemaker, I.D., Santanach, A., Jansen, P.W.T.C., Di Croce, L. and Vermeulen, M. (2016) The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. *Nat. Struct. Mol. Biol.*, **23**, 682–690.
27. Shiio, Y., Eisenman, R.N., Yi, E.C., Donohoe, S., Goodlett, D.R. and Aebersold, R. (2003) Quantitative proteomic analysis of chromatin-associated factors. *J. Am. Soc. Mass Spectrom.*, **14**, 696–703.
28. Räsche, M., Smeenk, G., Hansen, R.K., Temu, T., Oka, Y., Hein, M.Y., Nagaraj, N., Long, D.T., Walter, J.C., Hofmann, K. *et al.* (2015) DNA repair. Proteomics reveals dynamic assembly of repair complexes during bypass of DNA cross-links. *Science*, **348**, 1253671.
29. Kustatscher, G., Hégarat, N., Wills, K.L.H., Furlan, C., Bukowski-Wills, J.-C., Hochegger, H. and Rappsilber, J. (2014) Proteomics of a fuzzy organelle: interphase chromatin. *EMBO J.*, **33**, 648–664.
30. Kustatscher, G., Wills, K.L.H., Furlan, C. and Rappsilber, J. (2014) Chromatin enrichment for proteomics. *Nat. Protoc.*, **9**, 2090–2099.
31. Ginno, P.A., Burger, L., Seebacher, J., Iesmantavicius, V. and Schübeler, D. (2018) Cell cycle-resolved chromatin proteomics reveals the extent of mitotic preservation of the genomic regulatory landscape. *Nat. Commun.*, **9**, 4048.
32. Aranda, S., Alcaine-Colet, A., Blanco, E., Borràs, E., Caillot, C., Sabidó, E. and Di Croce, L. (2019) Chromatin capture links the metabolic enzyme AHCY to stem cell proliferation. *Sci. Adv.*, **5**, eaav2448.
33. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. and Saitou, M. (2011) Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell*, **146**, 519–532.
34. Brons, I.G.M., Smithers, L.E., Trotter, M.W.B., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A. *et al.* (2007) Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature*, **448**, 191–195.
35. Stauske, M., Rodriguez Polo, I., Haas, W., Knorr, D.Y., Borchert, T., Streckfuss-Bömeke, K., Dressel, R., Bartels, I., Tiburcy, M., Zimmermann, W.-H. *et al.* (2020) Non-Human primate iPSC generation, cultivation, and cardiac differentiation under chemically defined conditions. *Cells*, **9**, 1349.
36. Rappsilber, J., Mann, M. and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.*, **2**, 1896–1906.
37. Batth, T.S., Tollenaere, M.A.X., Rütger, P., Gonzalez-Franquesa, A., Prabhakar, B.S., Bekker-Jensen, S., Deshmukh, A.S. and Olsen, J.V. (2019) Protein aggregation capture on microparticles enables multipurpose proteomics sample preparation. *Mol. Cell. Proteomics*, **18**, 1027–1035.
38. Mulholland, C.B., Nishiyama, A., Ryan, J., Nakamura, R., Yiğit, M., Glück, I.M., Trummer, C., Qin, W., Bartoschek, M.D., Traube, F.R. *et al.* (2020) Recent evolution of a TET-controlled and DPPA3/STELLA-driven pathway of passive DNA demethylation in mammals. *Nat. Commun.*, **11**, 5972.
39. Stolz, P., Mantero, A.S., Tvardovskiy, A., Ugur, E., Wange, L.E., Mulholland, C.B., Cheng, Y., Wierer, M., Enard, W., Schneider, R. *et al.* (2022) TET1 regulates gene expression and repression of endogenous retroviruses independent of DNA demethylation. *Nucleic Acids Res.*, **50**, 8491–8511.
40. Qin, W., Ugur, E., Mulholland, C.B., Bultmann, S., Solovei, I., Modic, M., Smets, M., Wierer, M., Forné, I., Imhof, A. *et al.* (2021) Phosphorylation of the HP1 β hinge region sequesters KAP1 in heterochromatin and promotes the exit from naïve pluripotency. *Nucleic Acids Res.*, **49**, 7406–7423.
41. Bruderer, R., Bernhardt, O.M., Gandhi, T., Miladinović, S.M., Cheng, L.-Y., Messner, S., Ehrenberger, T., Zanotelli, V., Butscheid, Y., Escher, C. *et al.* (2015) Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol. Cell. Proteomics*, **14**, 1400–1410.
42. Cox, J. and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **26**, 1367.
43. Cox, J., Hein, M.Y., Lubner, C.A., Paron, I., Nagaraj, N. and Mann, M. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics*, **13**, 2513–2526.
44. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V. and Mann, M. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.*, **10**, 1794–1805.
45. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M. and Cox, J. (2016) The Perseus computational platform for comprehensive analysis of (prote) omics data. *Nat. Methods*, **13**, 731.
46. Snel, B., Lehmann, G., Bork, P. and Huynen, M.A. (2000) STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.*, **28**, 3442–3444.
47. Demichev, V., Messner, C.B., Vernardis, S.I., Lilley, K.S. and Ralser, M. (2020) DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods*, **17**, 41–44.
48. Venable, J.D., Dong, M.-Q., Wohlschlegel, J., Dillin, A. and Yates, J.R. (2004) Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods*, **1**, 39–45.
49. Gillet, L.C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R. and Aebersold, R. (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics*, **11**, O111.016717.
50. Bruderer, R., Bernhardt, O.M., Gandhi, T., Xuan, Y., Sondermann, J., Schmidt, M., Gomez-Varela, D. and Reiter, L. (2017) Optimization of experimental parameters in data-independent mass spectrometry significantly increases depth and reproducibility of results. *Mol. Cell. Proteomics*, **16**, 2296–2309.
51. Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B.C. and Aebersold, R. (2018) Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol. Syst. Biol.*, **14**, e8126.
52. Hansen, F.M., Tanzer, M.C., Brüning, F., Bludau, I., Stafford, C., Schulman, B.A., Robles, M.S., Karayel, O. and Mann, M. (2021) Data-independent acquisition method for ubiquitinome analysis reveals regulation of circadian biology. *Nat. Commun.*, **12**, 254.
53. Steger, M., Demichev, V., Backman, M., Ohmayer, U., Ihmor, P., Müller, S., Ralser, M. and Daub, H. (2021) Time-resolved in vivo

- ubiquitinome profiling by DIA-MS reveals USP7 targets on a proteome-wide scale. *Nat. Commun.*, **12**, 5399.
54. van Mierlo, G., Wester, R.A. and Marks, H. (2019) A mass spectrometry survey of chromatin-associated proteins in pluripotency and early lineage commitment. *Proteomics*, **19**, e1900047.
 55. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Michael Cherry, J., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. et al. (2000) Gene ontology: tool for the unification of biology. *Nat. Genet.*, **25**, 25–29.
 56. Weinberger, L., Ayyash, M., Novershtern, N. and Hanna, J.H. (2016) Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nat. Rev. Mol. Cell Biol.*, **17**, 155–169.
 57. Fuerer, C., Nusse, R. and Ten Berge, D. (2008) Wnt signalling in development and disease. Max Delbrück Center for Molecular Medicine meeting on Wnt signaling in development and disease. *EMBO Rep.*, **9**, 134–138.
 58. ten Berge, D., Kurek, D., Blauwkamp, T., Koole, W., Maas, A., Eroglu, E., Siu, R.K. and Nusse, R. (2011) Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells. *Nat. Cell Biol.*, **13**, 1070–1075.
 59. Vallier, L., Alexander, M. and Pedersen, R.A. (2005) Activin/nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J. Cell Sci.*, **118**, 4495–4509.
 60. Yang, P., Humphrey, S.J., Cinghu, S., Pathania, R., Oldfield, A.J., Kumar, D., Perera, D., Yang, J.Y.H., James, D.E., Mann, M. et al. (2019) Multi-omic profiling reveals dynamics of the phased progression of pluripotency. *Cell Syst.*, **8**, 427–445.
 61. van Mierlo, G., Dirks, R.A.M., De Clerck, L., Brinkman, A.B., Huth, M., Kloet, S.L., Saksouk, N., Kroeze, L.I., Willems, S., Farlik, M. et al. (2019) Integrative proteomic profiling reveals PRC2-dependent epigenetic crosstalk maintains ground-state pluripotency. *Cell Stem Cell*, **24**, 123–137.
 62. Kalkan, T., Bornelöv, S., Mulas, C., Diamanti, E., Lohoff, T., Ralser, M., Middelkamp, S., Lombard, P., Nichols, J. and Smith, A. (2019) Complementary activity of ETV5, RBPJ, and TCF3 drives formative transition from naive pluripotency. *Cell Stem Cell*, **24**, 785–801.
 63. Gretarsson, K.H. and Hackett, J.A. (2020) Dppa2 and Dppa4 counteract de novo methylation to establish a permissive epigenome for development. *Nat. Struct. Mol. Biol.*, **27**, 706–716.
 64. Hackett, J.A., Huang, Y., Günesdogan, U., Gretarsson, K.A., Kobayashi, T. and Surani, M.A. (2018) Tracing the transitions from pluripotency to germ cell fate with CRISPR screening. *Nat. Commun.*, **9**, 4292.
 65. Pastor, W.A., Chen, D., Liu, W., Kim, R., Sahakyan, A., Lukianchikov, A., Plath, K., Jacobsen, S.E. and Clark, A.T. (2016) Naive human pluripotent cells feature a methylation landscape devoid of blastocyst or germline memory. *Cell Stem Cell*, **18**, 323–329.
 66. Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L. and McKay, R.D.G. (2007) New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature*, **448**, 196–199.
 67. Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficiz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W. et al. (2014) Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell*, **158**, 1254–1269.
 68. Ecco, G., Imbeault, M. and Trono, D. (2017) KRAB zinc finger proteins. *Development*, **144**, 2719–2729.
 69. Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D. and Pfaff, S.L. (2012) Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*, **487**, 57–63.
 70. Dixon, G., Pan, H., Yang, D., Rosen, B.P., Jashari, T., Verma, N., Pulecio, J., Caspi, I., Lee, K., Stransky, S. et al. (2021) QSER1 protects DNA methylation valleys from de novo methylation. *Science*, **372**, eabd0875.
 71. Santos-Barriopedro, I., van Mierlo, G. and Vermeulen, M. (2021) Off-the-shelf proximity biotinylation for interaction proteomics. *Nat. Commun.*, **12**, 5015.
 72. Miotto, B. and Struhl, K. (2010) HBO1 histone acetylase activity is essential for DNA replication licensing and inhibited by Geminin. *Mol. Cell*, **37**, 57–66.
 73. Wong, P.G., Glozak, M.A., Cao, T.V., Vaziri, C., Seto, E. and Alexandrow, M. (2010) Chromatin unfolding by Cdt1 regulates MCM loading via opposing functions of HBO1 and HDAC11-geminin. *Cell Cycle*, **9**, 4351–4363.
 74. Lalonde, M.-E., Avvakumov, N., Glass, K.C., Joncas, F.-H., Saksouk, N., Holliday, M., Paquet, E., Yan, K., Tong, Q., Klein, B.J. et al. (2013) Exchange of associated factors directs a switch in HBO1 acetyltransferase histone tail specificity. *Genes Dev.*, **27**, 2009–2024.
 75. Mulholland, C.B., Traube, F.R., Ugur, E., Parsa, E., Eckl, E.M., Schöning, M., Modic, M., Bartoschek, M.D., Stolz, P., Ryan, J. et al. (2020) Distinct and stage-specific contributions of TET1 and TET2 to stepwise cytosine oxidation in the transition from naive to primed pluripotency. *Sci. Rep.*, **10**, 12066.
 76. Kempf, J.M., Weser, S., Bartoschek, M.D., Metzeler, K.H., Vick, B., Herold, T., Völse, K., Mattes, R., Scholz, M., Wange, L.E. et al. (2021) Loss-of-function mutations in the histone methyltransferase EZH2 promote chemotherapy resistance in AML. *Sci. Rep.*, **11**, 5838.
 77. Griffin, T.J., Gygi, S.P., Ideker, T., Rist, B., Eng, J., Hood, L. and Aebersold, R. (2002) Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae** *S. Mol. Cell. Proteomics*, **1**, 323–333.
 78. Theurillat, I., Hendriks, I.A., Cossec, J.-C., Andrieux, A., Nielsen, M.L. and Dejean, A. (2020) Extensive SUMO modification of repressive chromatin factors distinguishes pluripotent from somatic cells. *Cell Rep.*, **33**, 108251.
 79. Sheban, D., Shani, T., Maor, R., Aguilera-Castrejón, A., Mor, N., Oldak, B., Shmueli, M.D., Eisenberg-Lerner, A., Bayerl, J., Hebert, J. et al. (2022) SUMOylation of linker histone H1 drives chromatin condensation and restriction of embryonic cell fate identity. *Mol. Cell*, **82**, 106–122.
 80. Davidson, K.C., Mason, E.A. and Pera, M.F. (2015) The pluripotent state in mouse and human. *Development*, **142**, 3090–3099.
 81. Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A. et al. (2013) Derivation of novel human ground state naive pluripotent stem cells. *Nature*, **504**, 282–286.
 82. Grabole, N., Tischler, J., Hackett, J.A., Kim, S., Tang, F., Leitch, H.G., Magnúsdóttir, E. and Surani, M.A. (2013) Prdm14 promotes germline fate and naive pluripotency by repressing FGF signalling and DNA methylation. *EMBO Rep.*, **14**, 629–637.
 83. Maldonado-Saldivia, J., van den Bergen, J., Krouskos, M., Gilchrist, M., Lee, C., Li, R., Sinclair, A.H., Surani, M.A. and Western, P.S. (2007) Dppa2 and Dppa4 are closely linked SAP motif genes restricted to pluripotent cells and the germ line. *Stem Cells*, **25**, 19–28.
 84. Schemmer, J., Araúz-Bravo, M.J., Haas, N., Schäfer, S., Weber, S.N., Becker, A., Eckert, D., Zimmer, A., Nettersheim, D. and Schorle, H. (2013) Transcription factor TFAP2C regulates major programs required for murine fetal germ cell maintenance and haploinsufficiency predisposes to teratomas in male mice. *PLoS One*, **8**, e71113.
 85. Lim, J.J., Shim, M.S., Lee, J.E. and Lee, D.R. (2014) Three-step method for proliferation and differentiation of human embryonic stem cell (hESC)-derived male germ cells. *PLoS One*, **9**, e90454.
 86. Graf, T. and Stadtfeld, M. (2008) Heterogeneity of embryonic and adult stem cells. *Cell Stem Cell*, **3**, 480–483.
 87. Kagiwada, S., Aramaki, S., Wu, G., Shin, B., Kutejova, E., Obridge, D., Adachi, K., Wrana, J.L., Hübner, K. and Schöler, H.R. (2021) YAP establishes epiblast responsiveness to inductive signals for germ cell fate. *Development*, **148**, dev199732.
 88. Torrente, M.P., Zee, B.M., Young, N.L., Baliban, R.C., LeRoy, G., Floudas, C.A., Hake, S.B. and Garcia, B.A. (2011) Proteomic interrogation of human chromatin. *PLoS One*, **6**, e24747.
 89. Kulej, K., Avgousti, D.C., Sidoli, S., Herrmann, C., Della Fera, A.N., Kim, E.T., Garcia, B.A. and Weitzman, M.D. (2017) Time-resolved global and chromatin proteomics during Herpes Simplex virus type 1 (HSV-1) infection. *Mol. Cell. Proteomics*, **16**, S92–S107.
 90. van Mierlo, G. and Vermeulen, M. (2021) Chromatin proteomics to study epigenetics - challenges and opportunities. *Mol. Cell. Proteomics*, **20**, 100056.
 91. Pino, L.K., Just, S.C., MacCoss, M.J. and Searle, B.C. (2020) Acquiring and analyzing data independent acquisition proteomics experiments without spectrum libraries. *Mol. Cell. Proteomics*, **19**, 1088–1103.

92. Bekker-Jensen,D.B., Bernhardt,O.M., Hogrebe,A., Martinez-Val,A., Verbeke,L., Gandhi,T., Kelstrup,C.D., Reiter,L. and Olsen,J.V. (2020) Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. *Nat. Commun.*, **11**, 787.
93. Ragazzini,R., Pérez-Palacios,R., Baymaz,I.H., Diop,S., Ancelin,K., Zielinski,D., Michaud,A., Givélet,M., Borsos,M., Aflaki,S. *et al.* (2019) EZHIP constrains Polycomb Repressive Complex 2 activity in germ cells. *Nat. Commun.*, **10**, 3858.
94. Jain,S.U., Rashoff,A.Q., Krabbenhoft,S.D., Hoelper,D., Do,T.J., Gibson,T.J., Lundgren,S.M., Bondra,E.R., Deshmukh,S., Harutyunyan,A.S. *et al.* (2020) H3 K27M and EZHIP impede H3K27-methylation spreading by inhibiting allosterically stimulated PRC2. *Mol. Cell*, **80**, 726–735.
95. Gonzales-Cope,M., Sidoli,S., Bhanu,N.V., Won,K.-J. and Garcia,B.A. (2016) Histone H4 acetylation and the epigenetic reader Brd4 are critical regulators of pluripotency in embryonic stem cells. *Bmc Genomics (Electronic Resource)*, **17**, 95.
96. Kinoshita,M. and Smith,A. (2018) Pluripotency deconstructed. *Dev. Growth Differ.*, **60**, 44–52.
97. Rossant,J. and Tam,P.P.L. (2017) New insights into early Human development: lessons for stem cell derivation and differentiation. *Cell Stem Cell*, **20**, 18–28.
98. Nakamura,T., Okamoto,I., Sasaki,K., Yabuta,Y., Iwatani,C., Tsuchiya,H., Seita,Y., Nakamura,S., Yamamoto,T. and Saitou,M. (2016) A developmental coordinate of pluripotency among mice, monkeys and humans. *Nature*, **537**, 57–62.
99. Guo,G., Stirparo,G.G., Strawbridge,S.E., Spindlow,D., Yang,J., Clarke,J., Dattani,A., Yanagida,A., Li,M.A., Myers,S. *et al.* (2021) Human naive epiblast cells possess unrestricted lineage potential. *Cell Stem Cell*, **28**, 1040–1056.
100. Loh,K.M. and Lim,B. (2011) A precarious balance: pluripotency factors as lineage specifiers. *Cell Stem Cell*, **8**, 363–369.
101. Sarkar,A. and Hochedlinger,K. (2013) The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. *Cell Stem Cell*, **12**, 15–30.
102. Yang,S.-H., Kalkan,T., Morissroe,C., Marks,H., Stunnenberg,H., Smith,A. and Sharrocks,A.D. (2014) Otx2 and Oct4 drive early enhancer activation during embryonic stem cell transition from naive pluripotency. *Cell Rep.*, **7**, 1968–1981.
103. Perez-Riverol,Y., Bai,J., Bandla,C., García-Seisdedos,D., Hewapathirana,S., Kamatchinathan,S., Kundu,D.J., Prakash,A., Frericks-Zipper,A., Eisenacher,M. *et al.* (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.*, **50**, D543–D552.