ChromID identifies the protein interactome at chromatin marks

Rodrigo Villaseñor[1], Ramon Pfaendler[1], Christina Ambrosi[1,2], Stefan Butz[1,2], Sara Giuliani[1], Elana Bryan[3], Thomas W. Sheahan[3], Annika Gable[4,2], Nina Schmolka[1], Massimiliano Manzo[1,2], Joël Wirz[1], Christian Feller[5], Christian von Mering[4], Ruedi Aebersold[5,6], Philipp Voigt[3], Tuncay Baubec[1]

[1]Department of Molecular Mechanism of Disease, University of Zurich, Zurich, Switzerland [2]Life Science Zurich Graduate School, University of Zurich and ETH Zurich, Zurich, Switzerland [3]Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK [4]Institute of Molecular Life Sciences and Swiss Institute of Bioinformatics, University of Zurich, Zurich, Switzerland [5]Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland [6]Faculty of Science, University of Zurich, Zurich, Switzerland

Abstract

Chromatin modifications regulate genome function by recruiting protein factors to the genome. However, the protein composition at distinct chromatin modifications remains to be fully characterized. Here, we use natural protein domains as modular building blocks to develop engineered chromatin readers (eCRs) selective for DNA methylation and histone tri-methylation at H3K4, H3K9 a H3K27 residues. We first demonstrate their utility as selective chromatin binders in living cells by stably expressing eCRs in mouse embryonic stem cells and measuring their subnuclear localisation, genomic distribution and histone modification–binding preference. By fusing eCRs to the biotin ligase BASU, we establish ChromID, a method for identifying the chromatin-dependent protein interactome based on proximity biotinylation, and apply it to distinct chromatin modifications in mouse stem cells. Using a synthetic dual-modification reader, we also uncover the protein composition at bivalent promoters marked by H3K4me3 and H3K27me3.

These results highlight the ability of ChromID to obtain a detailed view of protein interaction networks on chromatin.

Chromatin and numerous chemical modifications on histones and DNA play critical roles in organismal development and human health[1]. These modifications are recognised by specialised reader domains in regulatory proteins and multiprotein complexes[2,3]. Depending on the presence and composition of modifications at genomic sites, regulatory factors can associate with chromatin in a spatiotemporal manner[4]. However, a major challenge in the field remains to understand how this chemical language on chromatin defines the protein interactome of the genome.

In recent years, proteomics-based assays helped to measure the affinity of proteins to chromatin marks. Current methods probe the cellular proteome using synthetic histone peptides, methylated DNA probes or *in-vitro*-reconstituted nucleosomes[5–9]. In addition, proteins bound to specific genomic segments can be identified using enrichment via antibodies, DNA sequence-specific probes or more recently, engineered dCas9-fusion proteins[10–14]. While these studies have greatly enhanced our current knowledge about interactions between proteins and chromatin marks, the available methods rely on artificial chromatin, protein-protein crosslinking, or methods that require access to the underlying DNA, leading to chromatin disruption. Therefore, novel approaches are required that enable detection of dynamic interactions between proteins and physiological chromatin in living cells.

Here, we developed chromatin-dependent protein identification (ChromID) to identify the local protein composition at individual and combinatorial chromatin marks. To this end, we used the reader domains of well-established chromatin regulators as modules to build engineered chromatin readers (eCRs). We first quantified and functionally validated the genome-wide binding and histone-PTM interaction preferences of individual eCRs towards DNA methylation, H3K9me3, H3K4me3 and H3K27me3, demonstrating their applicability as selective binders in mouse stem cells. Finally, we used the specificity of eCRs to recruit promiscuous biotin ligases to detect proteins associated with these individual chromatin modifications in mouse embryonic stem cells, revealing similarities and differences in the protein composition between these marks. By coupling ChromID to a synthetic dual-modification reader, we furthermore detected proteins associated with genomic regions marked by the bivalent modification H3K4me3 and H3K27me3.

Results

Generation and characterisation of engineered chromatin readers in mouse embryonic stem cells

We first assembled well-characterised chromatin reader domains into synthetic reporter proteins to test their affinity and specificity for individual chromatin modifications in living cells. We used the chromo domains specific for H3K27me3 from CBX7 and Drosophila Polycomb (dPC)[15,16], the H3K9me3-specific chromodomain from CBX1[17,18], the Phd domain specific for H3K4me3 from TAF3[19], and the MBD domains from the DNA

methylation readers MBD1 and MeCP2[20,21] (Fig. 1a). cDNA sequences were assembled either as single- or dual-domain constructs into a protein expression cassette containing a biotin acceptor site for biochemical purification, a nuclear localisation signal (NLS), and eGFP for live imaging and detection (Fig. 1a and Supplementary Fig. 1a). All constructs were integrated to a defined site in the mouse genome via Recombinase-Mediated Cassette Exchange RMCE[22], enabling fast generation of stable mESC lines expressing the proteins from the same genomic location and under the control of the same promoter (Fig. 1b). Measurements of eGFP fluorescence and protein levels indicated that all generated cell lines display stable and homogenous expression of the introduced engineered Chromatin Readers (eCR) at intermediate protein levels (Supplementary Fig. 1b-d). We next performed *in vitro* differentiation of mESC to glutamatergic neurons to test if the presence of eCRs interferes with biological processes relevant for cellular identity and function. Previous work indicated the requirement of the targeted modifications for cellular differentiation[23,24]. All cell lines successfully differentiated to mature neurons, and we could not observe any differentiation defects in presence of the eCRs, suggesting that their stable expression does not interfere with cellular processes (Supplementary Fig. 1e-g).

Live imaging of stable cell lines with single chromodomain eCRs targeting histone methylation showed a diffuse nuclear localisation with accumulation in nucleoli, similar to the eGFP control lacking reader domains. In contrast, eCRs containing two chromodomains showed a defined pattern with signals for the CBX1-2xChromo eCR at DAPI-dense chromocenters and the CBX7- and dPC2-2xChromo eCRs forming discrete subnuclear aggregates at the nuclear periphery and around nucleoli (Fig. 1c and Supplementary Fig. 2a). These localisation patterns are identical to what has been reported for the subnuclear distribution of H3K9me3 and H3K27me3, respectively[25,26] (Supplementary Fig. 2b). In contrast to the chromodomains, the single and dual TAF3 Phd-domain eCR showed a homogenous signal throughout the entire nucleus (Supplementary Fig. 2a) as previously reported for H3K4me3 distribution using antibodies[26]. eCRs containing single MBD domains from MBD1 or MeCP2 co-localised with DNA-methylated, DAPI-dense chromocenters, similar to their corresponding full-length proteins (Supplementary Fig. 2a)[22]. Furthermore, live-cell imaging of the eGFP-tagged eCRs enabled us to explore their localisation during cell cycle progression and along the condensed M-phase chromosomes (Supplementary Fig. 2c-e and Supplementary Videos).

The results obtained from single and dual-domain eCRs specific for histone modifications indicate that one domain is not sufficient to promote localisation and that multivalent interactions are required. To further validate this, we made use of reconstituted nucleosomes carrying H3K4me3 or H3K27me3 marks on both histone tails. Pulldown experiments using recombinant single- or dual-domain eCRs indicate robust interactions only for dual domains, but not for single-domain eCRs, supporting the necessity of multivalent interactions for stable binding of eCRs to histone-PTMs *in vivo* (Supplementary Fig. 2f).

Functional analysis validates the interaction preference of eCRs with specific chromatin modifications

Next, we explored the genome-wide binding patterns of all eCRs by biotin-ChIP-seq[22]. By visual inspection of the binding tracks we observed eCR-specific signals corresponding to the distribution of target histone modifications and DNA methylation, indicating correct localisation to these marks (Fig. 1d-f and Supplementary Fig. 3a-b). Their selective binding preference to chromatin modifications was also confirmed by genome-wide enrichments and direct comparison to histone modifications, DNA methylation and endogenous reader proteins (Fig. 1g, Supplementary Fig. 3c-d and 4a-d). The signals obtained from individual eCR datasets indicated their clear distinction in binding to genomic elements modified by the corresponding target modifications (Supplementary Fig. 5a-d), where eCRs specific for H3K4me3 preferentially associated with gene promoters, H3K9me3 readers with repetitive elements, and DNA methylation readers with methyl-CpG-dense exons (Supplementary Fig. 5e). Notably, and in accordance with live-cell imaging, only experiments using eCRs with two histone-PTM reader domains resulted in detectable binding signals (Supplemental Fig. 3a-b and 5f), highlighting again the necessity for multivalent interactions for stable target engagement.

To investigate the specificity of eCRs we introduced mutations to the reader domains known to disrupt binding: CBX1-W42A[27], CBX7-W35A[28], MBD1-R22A[29] and furthermore, various Rett syndrome mutations in the MeCP2 MBD (R106W, R133C, T158M[30]) (Supplementary Fig. 6a). In all tested instances, we observed that mutations led to a partial or complete disruption of subnuclear localisation (Supplementary Fig. 6b), as well as loss of genome-wide binding to chromatin modifications (Fig. 2a-b, d and Supplementary Fig. 6c). The same disruption of localisation was observed for wild type readers in absence of the respective chromatin marks, highlighting that binding is fully dependent on the target modification. This dependency was shown by loss of binding of MeCP2-1xMBD in cell lines lacking DNA methylation (*Dnmt*-TKO), and CBX7-2xChromo in cell lines lacking H3K27me3 (*Eed*-KO) (Fig. 2c and Supplementary Fig. 6d-f).

Next, we employed mass spectrometry as an orthogonal approach to identify the histone PTMs that are preferentially bound by the eCRs in living cells. Towards this, we detected and quantified the modifications on histones enriched in ChIP experiments using a synthetic reference peptide library including 87 individual and combined marks on histone tails from H2A, H3 and H4[31] (Supplemental Fig. 7a and b). Overall, the enriched histone PTMs reflect the genome-wide correlations described above, further corroborating the specific affinity of the reader domains for their substrates in living cells (Fig. 2e). In case of the TAF3-2xPhd eCR, we detect histone H3 tails that carry di- and tri- but not mono-methylation marks at the lysine 4 residue and furthermore acetylated H3 and H4 (K9, K14 on histone H3 and K5, K8, K12, K16 on histone H4) (Fig. 2e-f and Supplementary Fig. 7c). Histone H3 tails containing methylated K9, K27 or K36 residues were generally depleted in the TAF3-2xPhd eCR pulldowns (Fig. 2e). In contrast, CBX1-2xChromo eCR-enriched histone tails predominantly carry the H3K9 tri-methyl mark as well as H4K20me3 (Figure 2e-f), a modification co-existing with H3K9me3 at repetitive heterochromatin[32] (Supplementary Fig. 7d-e). In

addition, we also detect that H3S10-phosphorylation prevents binding of the CBX1-eCR to H3K9me3, as previously reported for CBX1 *in vitro*[33] (Fig. 2e,g).

Taken together, these experiments validate the target specificity of the introduced eCRs to chromatin marks. Furthermore, the obtained results highlight the suitability of eCRs as multi-purpose cellular probes to detect the distribution of chromatin modifications in living cells by live imaging, genomics and proteomics.

ChromID reveals the protein composition at H3K9- and DNA-methylated sites via eCR-mediated proximity biotin labelling

Having fully characterised the *in vivo* binding specificity of eCRs, we wanted to exploit their genomic localisation to detect the protein composition at distinct chromatin modifications via proximity biotin ligation (Fig. 3a). We tested three different promiscuous biotin ligases for their labelling efficiency during 24 hours in murine ES cells: BirA* R118G-mutant[34], BioID2[35] and BASU[36] (Supplementary Fig. 8a-b). BioID2 and BASU showed the highest labelling efficiency under these conditions, therefore we used these ligases in combinations with the specific H3K9me3-reader to establish the optimal conditions using quantitative label-free LC-MS/MS. The protein interactome of this mark has been well-described in mammalian cells[5,6], and served as a proof of concept to define optimal settings for ChromID (Supplementary Fig. 8c-e and Online Methods). Finally, based on the signal to noise ratio we chose BASU with 12 hours biotin-treatment followed by highly stringent washing with SDS as the most-optimal condition.

Using these conditions, we identified 58 high-confidence H3K9me3-associated proteins, which were enriched for Gene Ontology terms linked to pericentric or telomeric heterochromatin, confirming proteins found in other proteomic approaches[5,6,37] (Fig. 3b-c, Supplementary Fig. 8e and Supplementary Table 1). Found factors include the H3K9 methyltransferases SETDB1, EHMT1 and EHMT2[38,39], the HUSH complex component MPP8[40], the chromatin remodeller ATRX[41], MeCP2 and UHRF1[42,43]. Besides these factors, our method enabled us to identify zinc finger proteins, which have been linked to heterochromatin (POGZ, WIZ[44,45]), and multiple instances that have not been characterised in the context of heterochromatin (CASZ1, ZNF24, ZNF292, ZNF512B, ZNF518B, ZNF280B and ZNF280D). To test if the newly identified proteins localize to H3K9me3-marked chromatin, we further validated the localization of the endogenous ZNF280D protein. For this, we endogenously tagged ZNF280D in mouse ES cells and performed biotin ChIP-seq (Supplementary Fig. 9a-b). Genome-wide binding of ZNF280D shows a strong localisation preference to H3K9me3 sites, confirming that ChromID indeed reveals proteins associated with specific chromatin marks (Supplementary Fig. 9c-d).

We next used ChromID in conjunction with the 5-methyl-CpG-specific eCR (MBD1-1xMBD) resulting in the identification of proteins associated with DNA methylation such as DNMT1 and UHRF1 and proteins enriched for Gene Ontology terms related to heterochromatin or recognition of DNA replication (Fig. 3d-e, Supplementary Fig. 10a and Supplementary Table 2). Besides known factors, we also observed several factors that have not been associated with DNA methylation in ES cells, such as TIF1A (also known as TRIM24), CASZ1, ZNF512B or TEAD1 (Fig. 3d). The latter was recently found to bind

to methylated DNA in HT-SELEX experiments[46]. To test the specificity of these interactions for DNA methylation readout, we repeated these experiments using the mutant 5mC-reader (MBD1-1xMBD-R22A) fused to BASU. We did not detect any significantly enriched proteins with the mutant 5mC-reader, suggesting that the identified proteins associate with DNA methylation (Fig. 3d and Supplementary Fig. 10a-c).

Engineered readout of combinatorial histone PTMs enables identification of proteins associated with monovalent and bivalent chromatin

Nucleosomes bivalently modified by H3K4me3 and H3K27me3 are found at developmental gene promoters and are thought to poise their expression for timely activation[47,48]. Addressing the genomic distribution and/or protein composition of bivalently modified sites and other combinatorial modifications has been a major challenge due to lack of tools that enable simultaneous detection of both marks. To overcome this limitation, we first characterised synthetic readers engineered for simultaneous detection of H3K4me3 and H3K27me3 on the same nucleosome. eCRs containing the CBX7-Chromodomain or the dPC-Chromodomain fused to the TAF3-Phd domain were stably expressed in ES cells as described above (Supplementary Fig. 11a-b). Genome-wide binding analysis indicates preferential binding of these bivalent eCRs to genomic sites marked by both H3K4me3 and H3K27me3 modifications, while regions containing either H3K4me3 or H3K27me3 were not enriched to the same levels (Fig. 4a-c and Supplementary Fig. 11c-e). Monovalent eCRs with affinity to H3K4me3 or H3K27me3 only, showed reduced enrichments to bivalent regions, while being predominantly recruited to sites modified by H3K4me3 or H3K27me3, respectively (Fig. 4a-b and Supplementary Fig. 11c and f). To test the requirement of both domains for the observed binding, we introduced mutations in either the TAF3-Phd (DW890/891AA[19]) or the CBX7-Chromo domains (W35A[28]) of the bivalent reader. We observe loss of binding at bivalent sites for both mutant variants (Supplemental Fig. 12a-c). To further evaluate the requirement of both histone modifications for recruitment of the bivalent readers, we have introduced the TAF3-CBX7-bivalent eCR to *Eed*-KO ES cells lacking H3K27me3. In the absence of H3K27me3, the bivalent reader fails to bind to the genome (Supplemental Fig. 12d-e), further supporting the finding that its binding is dependent on multivalent readout of both modifications by the two reader domains. Taken together, the modular architecture of eCRs opens new possibilities to study and manipulate combinatorial modifications in living cells.

The differences in genomic binding observed for the monovalent and bivalent eCRs encouraged us to perform ChromID with eCRs specific to H3K4me3, H3K27me3 and bivalent nucleosomes. In total, 136 unique proteins that directly or indirectly interact with the chromatin marks were found significantly enriched across these three datasets (Supplementary Fig. 13a-b, Supplementary Table 1 and 2). A total of 125 proteins were detected at H3K4me3 (TAF3-eCR), enriching for GO terms related to transcriptional regulation and H3K4me3 (Fig. 5a), including several transcription factors, bromodomain proteins, histone modifier and chromatin remodelling complexes, as well as members of the Transcription Factor IID (TFIID), Integrator-, Mediator- and Super Elongation-complexes (Supplementary Fig. 13a-b and Supplementary Table 1). Notably, we also detect proteins involved in co-transcriptional processes such as the RNA-specific adenosine deaminase

ADAR1 and the histone mRNA 3' end processing factor CASP8AP2/FLASH[49] to be associated with H3K4me3, which we confirmed by comparing genomic co-localisation of FLASH and H3K4me3 at transcribed histone genes (Fig. 5b).

The H3K27me3-reader enabled us to identify 20 high-confidence hits, enriching for GO terms associated with Polycomb repressive complexes and histone methyltransferases (Fig. 5a, Supplementary Fig. 13a-b and Supplementary Table 1). Among those hits we observed well-studied subunits of PRC1 and PRC2 (RING2, EZH2, MTF2 and JARD2). Notably, we also identified factors involved in H3K9 methylation like SETDB1 or the zinc finger proteins WIZ and ZNF518B, suggesting a potential crosstalk between proteins bound at H3K9me3 and H3K27me3 sites. Notably, this is not due to unspecific localisation of the H3K27me3 readers to H3K9me3 or vice versa, since we do not observe this cross-reactivity from our ChIP-seq data (Supplementary Fig. 13c). Furthermore, by performing ChromID with the H3K27me3-specific readers in ES cells lacking H3K27me3 (*Eed*-KO), we would expect that such unspecific interactions would persist. However, we fail to detect any enriched proteins, indicating that the reported interactions indeed originate from H3K27me3 sites (Supplementary Fig. 13d).

Finally, the combinatorial recognition of bivalent H3K4me3 and H3K27me3 loci by the CBX7-TAF3-eCR enabled us to discover 33 high-confidence factors associated with bivalent chromatin, enriched in GO terms related to transcriptional activation and repression (Fig. 5a, Supplementary Fig. 13a-b and Supplementary Table 1). These included catalytic subunits or components of the MLL1/MLL2, the NSL histone acetyltransferase and the TFIID basal transcription factor complex, although TFIID components were detected at lower levels compared to results obtained with the monovalent H3K4me3 reader. Other factors include enhancer of Polycomb homolog (EPC1 and EPC2) and components of the NuA4 histone acetyltransferase (HAT) complex[50]. Corroborating our findings, a recent study mapped the catalytic subunit of the NuA4 complex (TIP60) to bivalent regions in mouse ES cells[51]. We also identified the histone lysine 9 and 36 demethylase KDM4C/JMJD2C[52] that colocalises with EZH2 in mouse ES cells[53], and PHF8, a demethylase involved in removal of H3K9me2, H3K27me2 and H4K20me1-residues[54]. Among the core components of the PRC1 and PRC2 complexes, we also detected BCOR, MGAP and LMBL2 which are part of the alternative PRC1.1 and PRC1.6 complexes[55]. Notably, we also observe the methylcytosine dioxygenase TET1 and the transcriptional repressor SIN3A being associated to bivalent sites, in line with previous genomic studies showing TET1 and SIN3A at bivalent promoters in ES cells[56]. Finally, we introduced the bivalent reader in ES cells lacking H3K27me3 and performed ChromID to control for false-positive proteins stemming from unspecific interactions of the readers with marks outside of nucleosomes modified by H3K27me3 and H3K4me3. In this case we could not identify any significantly enriched proteins, indicating that the reported proteins are indeed localised to bivalently-modified regions in the genome (Supplementary Fig. 13e).

To exclude that the BASU biotin ligase could influence genomic localisation of the readers and therefore falsely report proteins from sites not decorated by the targeted modifications, we performed biotin-ChIP-seq of the eCR-BASU constructs and compared their binding to the previously-obtained datasets of eGFP-fusion constructs (Supplementary Fig. 14a-c).

Based on our genome-wide analysis, binding was highly correlated between the same readers fused to either eGFP or BASU, indicating that the reader domains are not influenced by the addition of the biotin ligase. In summary, these results highlight the applicability of modular eCRs as a platform for biotin ligase recruitment, enabling successful identification of the associated proteins of chromatin subtypes.

Integrative analysis of ChromID datasets reveals the chromatin preference of regulatory proteins

Based on the combined datasets from all ChromID experiments we investigated the distribution of proteins between the different chromatin states, revealing proteins shared between multiple chromatin states, and proteins specific to single chromatin modifications. The latter was most prominent for H3K4me3-associated proteins (Fig. 5c, Supplementary Fig. 15a). Notably, several proteins identified at H3K27me3 or bivalent regions were also associated with H3K4me3, which is expected given the overlap of these modifications in ES cells[47]. In addition, we found multiple proteins shared between H3K9me3 and H3K27me3, while little overlap was found between the H3K9me3 set and proteins detected by H3K4me3 or bivalent regions (Fig. 5c and Supplementary Fig. 15a). Given the well-established crosstalk between DNA methylation and H3K9me3, we identified a substantial overlap between these sets (Supplementary Fig. 15b). The functional relationship of the detected factors was further visualised from high-confidence interaction scores obtained from the STRING database, revealing a strong interconnectivity between proteins and complexes associated with H3K4me3 and bivalent regions or with heterochromatin marked by H3K9me3 or H3K27me3 (Fig. 5d).

To obtain a quantitative view on the distribution of regulatory proteins along the interrogated chromatin marks, we clustered factors from different regulatory groups based on their enrichment across all datasets (Fig. 5e). For transcription factors, we observed several associations with H3K4me3 regions (e.g. SP2, MAX, FOXK2, ZFX). In addition, several TFs, mainly uncharacterised ZNF proteins, are associated with DNA methylation (ZNF280B, ZNF292, ZNF462, CASZ1, TCF20) and we also recover TFs previously identified to interact with methylated DNA in pull-down or HT-SELEX assays (KLF4, RREB1, ZNF191/24)[9,46] (Fig. 5e). Similar to TFs, chromatin remodellers separate into a group predominantly associated with H3K4me3 (e.g. BRD2, BRD4, INO80, CECR2), and a group preferentially associated with closed chromatin (e.g. ATRX, BAZ2A, SMARCA1). Chromatin writers such as H3K4-specific methyltransferases (KMT2A, KMT2B) and histone acetyltransferases (EP300) were preferentially located at H3K4me3, while writers of repressive marks associated with DNA methylation, H3K9me3 and/or H3K27me3 (DNMT1, EHMT1/2, NSD1, EZH2). Furthermore, and in line with genome-wide binding data, erasers such as KDM2A, KDM2B, TET1 or KDM5A were predominantly found at H3K4me3 sites, and we find several DNA repair factors associated with repressive chromatin marks (Fig. 5e). Taken together, these datasets obtained by ChromID provide a valuable resource of chromatin-mediated protein interactions in the ES cell genome.

Discussion

Here we present ChromID, a quantitative approach that enables identification of proteins associated with individual and combinatorial chromatin modifications in living cells. ChromID takes advantage of the affinity of engineered chromatin readers, which we obtained from natural reader domains of well-characterised chromatin regulators (CBX1, CBX7, dPC, TAF3, MBD1 and MeCP2). First, we characterised and functionally validated the binding selectivity of all eCRs using a series of quantitative and functional methods in mouse ES cells. The obtained results highlight the applicability of eCRs as an alternative to antibodies for studying subnuclear localisation, genome-wide distribution and histone-PTM combinations in living cells. Single domain eCRs were often insufficient to achieve binding to histone modifications under physiological conditions. This is in line with several well-known examples where binding of full-length proteins or complexes to chromatin rely on multivalent interactions[4], including recent studies that introduced synthetic multivalent chromatin readers for immunofluorescence or activation of reporter genes[57–59]. We made use of the required multivalent interactions to generate synthetic readers that recognise two modifications on the same nucleosome. Here, a short NLS was sufficient as a linker between the two reader domains, since our dual-reader eCRs target modifications on two different histone tails. This allowed us to directly target genomic sites that are bivalently marked by H3K4me3 and H3K27me3, providing novel tools to study and manipulate chromatin modifications in a context-dependent manner.

Finally, to identify the chromatin-associated protein interactome, we developed ChromID where we use the eCRs to tether promiscuous biotin ligases to chromatin, resulting in biotinylation of proteins in a ~ 35 nm radius around the modification of interest. This allowed us to detect proteins that directly and indirectly associate with chromatin modifications including DNA methylation, H3K4me3, H3K27me3 and H3K9me3, resulting in a total of 518 identified proteins. Among these we identified 180 high-confidence proteins enriched across all datasets, enabling us to assign factors based on their preference towards single or multiple chromatin marks. By employing the bivalent reader, we further achieved specific identification of proteins bound at sites marked by H3K4me3 and H3K27me3, revealing the presence of activating and repressing proteins from Trithorax and Polycomb complexes and additional factors that could play a role in chromatin regulation at bivalent sites.

Overall, the results from the individual and combinatorial measurements highlight ChromID as an approach to uncover how protein recruitment is influenced by chromatin modifications in living cells. The usage of natural reader domains in ChromID mimics physiological engagement of proteins with chromatin. This has several benefits, since the eCR-mediated interactions do not require crosslinking or single-stranded DNA. Another benefit of ChromID is the usage of proximity biotin ligation to label and subsequently identify the proteins associated with different chromatin flavours in a unified manner. This enables comparative studies between different chromatin modifications, circumventing the necessity of antibodies, which have always been limiting in such assays due to their variation in affinity and avidity, lack of availability and cost. Furthermore, once biotinylated, the proteins

are enriched using highly-stringent washing and elution conditions, ensuring effective removal of background signals and reproducible detection and multi-sample comparison.

We expect ChromID to be used to chart the protein interactome at multiple chromatin modifications and in numerous cell types in order to understand how the chemical language on chromatin directs protein recruitment in a spatiotemporal manner. The applicability of ChromID in living cells, as well as eCRs as synthetic readers, further opens the possibility to perform similar experiments in a tissue-specific manner in living animals to chart the epi-proteome during dynamic cellular processes and development.

Online Methods

Molecular cloning

Reader domains were amplified from cDNA or synthesized (IDT technologies) based on available domain annotations (Uniprot). Coding sequences are introduced in-frame to the RMCE-targeting vector parbit-v6 by Gibson assembly. The final construct expresses the N-terminal biotin-tagged domain of interest fused in-frame to a cassette containing an NLS signal followed by eGFP, an internal ribosome entry site (IRES) and the puromycin-N-acetyltransferase gene. All coding sequences are under control of a constitutive CAG promoter. BioID2-HA and the 13X-Linker were PCR amplified from MCS-13X-Linker-BioID2-HA plasmid (addgene #80899), HA-BASU was PCR amplified from BASU-RaPID plasmid (kindly provided by P. Khavari; equivalent to Addgene #107250). PCR-amplified products were cloned into RMCE-targeting vector L1-CAG-NLS-IRES-pac-1L (parbit-v9) by Gibson assembly. For bacterial expression of eCR-eGFP-6xHis fusion proteins, sequences spanning the domains of interest along with the NLS and eGFP were PCR amplified from parbit-v6 and subcloned into a modified pET-28 vector encoding an in-frame C-terminal 6xHis affinity tag.

Cell culture and cell line generation

Mouse embryonic stem cells (HA36CB1, 129×C57BL/6) were cultured as previously described[22]. Cell lines were obtained by recombinase-mediated cassette exchange (RMCE). Briefly, RMCE constructs were co-transfected with a CRE recombinase expression plasmid (1: 0.6 μg DNA ratio) to RMCE-competent and biotin ligase (BirA)-positive mouse ES cell lines (HA36CB1)[22]. Two rounds of selection were applied to yield a homogenous population of eCR expressing cells: 3mM ganciclovir for 4 days and 2mM puromycin for 2 days. Homogenous and stable protein expression was then monitored by eGFP expression using flow cytometry, immunofluorescence (IF) and immunoblotting. Transfections were carried out using Lipofectamine 3000 reagent (Thermo Fisher Scientific, L3000015) at a 2:1 μg DNA ratio in OptiMEM medium (Thermo Fisher Scientific, 31985070). The *Eed*-KO cell line was generated by co-transfecting pX330-U6-Chimeric_BB-CBh-hSpCas9 (addgene #42230) with a guide (GGTGAAAAAATAATGTCCTG) targeting exon 8 together with pRR-Puro recombination-reporter[60] (addgene #65853). 36 hours after transfection, cells were treated with 2 ug/ml puromycin for 36 h. Positive KO clones were validated by Sanger sequencing and Western blot. Endogenously tagged *Zfp280D* cell line was generated with a guide (AGTAGACCTGGCAGATGGAG) targeting exon 22 and co-transfecting pRR-EGFP

recombination-reporter (addgene #65852[60]). 72 hours after transfection, single GFP-positive cells were sorted by flow cytometry and validated by Sanger sequencing. Neuronal differentiation of ES cells was performed as previously described[61].

Flow cytometry

Cells were resuspended in DPBS and incubated with LIVE/DEADTM Fixable Near-IR Dead Cell Stain (Invitrogen, L34975) to discriminate cell viability. Samples were analysed on a FACSCanto (BD Biosciences). Cells were gated for viable and individual cells, channel voltages for eGFP (Alexa Fluor 488-A) and live/dead (APC-Cy7A) signals were set regarding verified negative and positive eGFP-expressing control cells. Raw files were analysed and visualised using FlowJo software (Tree Star; version 10.0.7). For CD24 measurements in neuronal progenitors, single cell suspensions were obtained from neuronal progenitors after 8 days of differentiation as previously described[61]. For cell surface staining cells were incubated for 30 min at 4°C with saturating concentration of anti-CD24a monoclonal antibody in the presence of anti-CD16/CD32 (eBioscience). Samples were acquired using FACSFortessa (BD Biosciences) and data were analysed using FlowJo software (Tree Star).

Western blotting

For eCR detection, 20 μg of protein were resolved in NuPAGE-Novex Bis-Tris 4–12% gradient gels (Invitrogen) and transferred on polyvinylidene fluoride (PVDF). Membranes were blocked with 5% (w/v) BSA in TBST (10mM Tris pH 8.0, 150mM NaCl, and 0.1% Tween-20), and stained with the corresponding primary antibody anti-HP1b/CBX1 (1:1,000, CST; #8676) or Lamin B1 (1:1000, Santa Cruz, sc-374015) at 4°C overnight, followed by detection with species-specific horseradish peroxidase (HRP)-conjugated secondary antibodies. For validation of *Eed*-KO cell lines, cells were lysed with NETN buffer (20 mM Tris (pH 8), 0.5 % (v/v) NP-40, 100 mM NaCl, 1 mM EDTA (pH 8)) supplemented with 1 x protease inhibitor cocktail (Roche; COEDTAF-RO) and 1 mM DTT (Sigma Aldrich, DTT-RO). Nuclei were pelleted at 6,500 x g for 10 min at 4 °C, washed once with NETN. Histones were acid-extracted overnight at 4 °C in 0.2 N HCl at a density of 4 x 10^7 nuclei per ml. Histone extracts were then centrifuged at 6,500 x g for 10 min at 4 °C to pellet debris and 5 ug were loaded onto a NuPAGE-Novex 16% Tris-Glycine Gel (Invitrogen). Western Blot and protein detection were performed as above with a transfer buffer containing no SDS, but 20 % MeOH, and membrane was stained with primary antibody anti-H3K27me3 (Diagenode, C15410195), anti-Histone H1 (Millipore, 05-457) and anti-Histone H3 (Abcam, ab1791).

Live-cell imaging and image processing

2x 10^4 eCR-eGFP fusion expressing cells were seeded on 0.2% gelatin-coated 35-mm chambered coverslip (Ibidi; 80826) one day before imaging. Next day, cells were stained with Hoechst 33342 (Invitrogen; 62249) for 10 min, washed twice with DPBS, and covered with ES cell medium containing DMEM w/o phenol red (Invitrogen, 31053028). Randomly selected cells were imaged with sequential acquisition settings on a Leica SP5 inverted confocal laser scanning microscope equipped with a climate chamber, an Argon laser for 453, 476, 488, 496, and 514 nm, and a diode laser for 561 nm. Filters for fluorescence

imaging were GFP (ex BP 470/40, em BP 525/50) and N3 (ex BP 546/12, 600/400). Confocal images were acquired with an HCX PL APO Leica 63× oil immersion objective with HyD detectors. Z stacks were acquired per site using a 0.3 μm step size. Time-lapse fluorescence microscopy was performed with a confocal spinning disk imaging system (Olympus IXplore SpinSR10,) equipped with a CSU-W1 unit (YOKOGAWA) and a 60× UPLSAPO UPlan S Apo silicon oil objective of 1.3 NA (Olympus Corporation). 11 z planes were acquired per site (1μm step size) every 5 min for approximately 12.5 hours. A 488nm laser was used to excite the GFP probe while emitted light was filtered by a 525/50 band pass filter and captured by a Prime BSI Scientific CMOS camera (2048 × 2048 pixels, Teledyne Photometrics). Images were deconvolved using Huygens Professional 19.10 software (Scientific Volume Imaging) using up to 40 iterations of the Classic Maximum Likelihood Estimation algorithm with a theoretical PSF. Background correction was automatic. The signal-to-noise ratio setting was adjusted empirically to 16 to give satisfactory results. Image analysis was performed on the resulting image series using FIJI (version 2.0.0) and the Bio-Formats Importer plugin. Appropriate single z-planes were then selected for further image analysis and display.

Fixed-cell immunofluorescence

Cells were seeded and grown as described above, fixed in 4% formaldehyde in PBS for 10 min at room temperature, washed three times in PBS, permeabilized for 5 min at room temperature in PBS supplemented with 0.1% Triton X-100 and 0.25% BSA (Sigma-Aldrich), and washed twice in PBS. Corresponding primary antibodies anti-H3K9me3: ab8898 (abcam), anti-H3K27me3: C15410195 (diagenode), anti-GFP: 11814460001 (Millipore) and secondary antibodies (Alexa Fluor 488 anti-mouse and 568 anti-rabbit IgGs from ThermoFisher) were diluted in PBS containing 2% FBS and 0.02% BSA. Primary antibody incubations were performed overnight at 4°C. Secondary antibody incubations were performed for 1h at room temperature. Following antibody incubations, cells were washed once with PBS and incubated for 10 min with PBS containing 4',6-diamidino-2-phenylindole dihydrochloride (DAPI, 0.5 μg/ml) for 10 min at room temperature to stain DNA. Randomly selected cells were imaged with sequential acquisition settings on a Leica SP5 inverted confocal laser scanning microscope.

Generation of recombinant nucleosomes

Core histones (*Xenopus* H3 and H4, human H2A and H2B) and truncated histone H3 for native chemical ligation (NCL) were expressed in E. coli and purified as previously described[48,62]. NCL reactions to generate H3K4me3, H3K4me1, and H3K9me3-modified histone H3 were carried out as described [62] with truncated H3 lacking residues 1-31 after the initiator methionine and containing a threonine-to-cysteine substitution at position 32 and a cysteine-to-alanine substitution at position 110 (H3 1–31 MT32C C110A) and the corresponding synthetic thioester peptide spanning residues 1–31 of histone H3.1 and containing the desired modification (Peptide Protein Research Ltd., Fareham, UK). For generation of H3K27me3-modified histones, a similarly truncated Xenopus H3 construct was used, lacking the first 44 residues and carrying a threonine-to-cysteine mutation at residue 45 (H3 1–45 MT45C C110A). Histone octamers were assembled and reconstituted into mononuclesomes carrying 601 DNA as described[48].

Bacterial expression of eCR-eGFP-His fusion proteins

His-tagged eCR-eGFP fusion proteins were expressed in BL21 (DE3) *E. coli* by induction for 3 h at 37°C with 0.5 mM IPTG in the presence of 20 μM $ZnCl_2$. Cells were lysed by sonication in lysis buffer (20 mM Tris HCl pH 8, 500 mM NaCl, 0.1% NP-40, 0.5 mM PMSF). His tagged protein was bound to Sepharose 6 Fast Flow Ni-NTA resin (GE Healthcare), washed with 300 mM wash buffer (50 mM NaH_2PO_4 pH 8, 300 mM NaCl, 20 mM imidazole, 0.1 mM PMSF) and 1 M wash buffer (1 M instead of 300 mM NaCl), and eluted in 50 mM NaH_2PO_4 pH 8, 300 mM NaCl, 250 mM imidazole. Fractions containing the desired eCR-eGFP fusion protein were pooled and dialysed against BC100 (20 mM HEPES KOH pH 8, 100 mM KCl, 10% glycerol, 0.5 mM DTT).

Nucleosome pulldown assays

For pulldown assays with recombinant modified nucleosomes and eCR-eGFP fusion proteins, streptavidin sepharose high performance beads (GE Healthcare) were blocked with 1 mg/ml BSA in pulldown buffer (20 mM HEPES KOH pH 7.9, 150 mM NaCl, 10% glycerol, 1 mM EDTA, 1 mM DTT, 0.2 mM PMSF, 0.1% NP-40, 0.1 mg/ml BSA) before three washes with pulldown buffer. All centrifugation steps were carried out at 1,500 g for 2 min at 4°C. All incubation steps were carried out at 4°C. Beads were incubated overnight with 3 μg of assembled recombinant nucleosomes in pulldown buffer. After three washes, bead-bound nucleosomes were incubated with increasing amounts of eCR-eGFP fusion proteins for 2 h. Beads were then six times washed with pulldown buffer by 5-min incubation under rotation before elution of bound proteins by boiling with 1.5x SDS sample buffer (95 mM Tris HCl pH 6.8, 15% glycerol, 3% SDS, 75 mM DTT, 0.15% bromophenol blue). Protein binding was analysed by Western Blotting with anti-His antibody (Sigma, H10229), and corresponding histone modification antibodies.

Chromatin immunoprecipitation sequencing

For cross-linking and chromatin extraction, 30–50 × 10^6 cells were fixed for 8 min with 1% formaldehyde at room temperature followed by the addition of glycine (final concentration 0.12 M) and incubation for 10 min on ice. Cell lysis, chromatin extraction and fractionation, followed by antibody or streptavidin-based enrichment was performed as previously described[63]. For biotin-ChIP we used 90 μl pre-blocked streptavidin-M280 per 150–200 μg chromatin, for antibody-ChIP we used 5 ug of antibody for 100 μg chromatin. Antibodies used: H3K27me3 (Diagenode, C15410195), H3K4me3 (abcam, ab8580), H3K9me3 (abcam, ab8898). ChIP-seq libraries were prepared using the NEB-next ChIP-seq library Kit (E6240) following the standard protocols. Up to 8 samples with different index barcodes were combined at equal molar ratios and sequenced as pools. Sequencing of library pools was performed on Illumina HiSeq 4000 or Nova Seq machines according to Illumina standards, with 75- to 150-bp single-end sequencing. Library demultiplexing was performed following Illumina standards.

Genomics data analysis

Sequencing samples were filtered for low-quality reads and adaptor sequences removed using Trim Galore (https://github.com/FelixKrueger/TrimGalore). Filtered reads were

mapped using QuasR[64] in R to the mouse genome (version mm9) using the BOWTIE algorithm allowing for two mismatches and only unique mappers were used (-m 1 --best --strata). Identical reads from PCR duplicates were filtered out.

To obtain genome-wide 1kb intervals, we partitioned the entire genome into 1 kb sized tiles. Intervals overlapping with satellite repeats (Repeatmasker), ENCODE black-listed and low mappability scores[65] (below 0.5) were removed in order to reduce artefacts due to annotation errors and repetitiveness. To detect eCR-enriched regions, we utilised MACS2 using the eGFP ChIP-seq as background and applying the following parameters: --broad -g mm --broad-cutoff 0.1. To detect antibody-specific peaks for histone modifications, we applied the same approach but using input chromatin as a background signal. Obtained histone modification peaks were further filtered according to qval >= 2 and pileup >= 3.4 scores. Peaks were overlapped with genomic features and coverages were calculated using the following hierarchy: promoters, enhancers, exons, repeats and introns. Promoters were defined as +/- 1kb around RefSeq gene TSS, enhancers were defined based on DHS peaks where H3K4me1 was higher than H3K4me3[66], exons and introns were retrieved based on RefSeq annotations, and repetitive elements using Repeatmasker. ChromHMM segmentation[67] of the mouse genome was obtained from http://compbio.mit.edu/ChromHMM/, as part of ENCODE. For Figure 1e, a genomic range object containing all peaks was generated, overlapping peak regions were merged, and finally used to compute correlations between eCR and antibody signals. To define H3K4me3-monovalent, H3K27me3-monovalent and bivalently-marked peaks, we calculated the H3K4me3 and H3K27me3 enrichments at these sites and selected all H3K4me3 peaks devoid of H3K27me3 as H3K4me3-monovalent peaks. H3K27me3 lacking H3K4me3 signals were selected as H3K27me3-monovalwent while H3K4me3 peaks positive for H3K27me3 were selected as bivalent peaks. ChIP enrichments at genomic segments and peaks were calculated as $\log_2$-fold changes over input chromatin (for antibodies) and over eGFP (for eCRs) after library size normalization and using a constant of eight pseudo counts to reduce sampling noise. Heatmap and average density-profiles around peaks were generated using genomation in R[68].

Mass spectrometry analysis of histone modifications

Histones were processed as described in[31,69] with modifications as described below. De-crosslinked histones were separated by SDS-PAGE on 16 % Novex Tris-Glycine gels (Invitrogen, XP00165BOX), stained with InstantBlue (Expedeon, ISB1L) and bands corresponding to core histones were excised. Gel pieces were washed twice with water and twice with 100 mM ammonium bicarbonate. Gel pieces were destained by incubating three times for each 10 min with 50 mM ammonium bicarbonate/50% acetonitrile at 37 °C with shaking at 800 rpm in a Thermomixer. Gel pieces were successively dehydrated by incubating with 100 mM ammonium bicarbonate, once 20 mM ammonium bicarbonate and three times with acetonitrile. Histones were twice derivatized by chemical acetylation by reacting 5 µL of d6-acetic anhydride ((CD3CO)2O, Sigma-Aldrich), 15 µL of 100 mM ammonium bicarbonate and 1 M ammonium bicarbonate buffered with 1:2 diluted ammonium hydroxide solution to keep the pH at 8. The reactions were performed for 45 min at 37 °C with shaking at 800 rpm in a Thermomixer. After the derivatization reactions,

histones were washed four times with 100 mM ammonium bicarbonate, two times with water, and three times with acetonitrile. Histone gel pieces were rehydrated with a 25 ng/ul trypsin solution in 100 mM ammonium bicarbonate (sequencing-grade trypsin from Promega) and digested overnight at 37 °C. Processing of tryptic peptides, mass spectrometric measurements and data analysis as previously described[69].

Nuclear extraction for ChromID

Cells were cultured with ES medium and induced for the corresponding time periods (12-24 hours) with 50μM biotin (Sigma) dissolved in DPBS. Cells were grown to about 90% confluency on 15cm dishes (approximately 50 x 10^6 cells), harvested by trypsinisation, and pelleted by centrifugation at 1000 rpm for 5 min. The subsequent steps were either performed on ice or at 4°C. Pellets were gently resuspended (by shaking) in 5 pellet volumes (PV) of nuclear extract buffer 1 (NEB1; 10mM HEPES pH7.5, 10mM KCl, 1mM EDTA, 1.5mM MgCl2, 1mM dithiothreitol (DTT), and 1x PIC) and swelled on ice for 10 min, followed by centrifugation at 2000 g for 10 min. Pellets were then gently resuspended in 2x PV of NEB1, followed by dounce homogenisation using a loose pistil (10 times up and down). Nuclei were collected by centrifugation at 2000 g for 10 min and resuspended in 1x PV of NEB1 + 12μl/ml of Benzonase (Millipore, 71206) to digest genomic DNA, followed by overhead rotation at 4°C for 3 hours. Nuclei were then pelleted by centrifugation at 2000 g for 10 min, resuspended in 1x PV of NEB2-450 (20mM HEPES pH 7.5, 0.2mM EDTA, 1.5mM MgCl2, 20% glycerol, 450mM NaCl, 1mM DTT, and 1x PIC), Dounce homogenised using a tight pistil (10 times up and down), vortexed, followed by overhead rotation at 4°C for one hour. Cell debris were removed by centrifugation at 2000 g for 10 min. The salt concentration of the nuclear extracts (NE) was adjusted to 150mM by drop-wise addition of 2x residual volume of NEB2-NS (see above, without NaCl), and NP40 levels were adjusted to 0.3%. Subsequently, protein concentrations were measured using QubitTM Protein Assay Kit (Thermo Fisher Scientific, Q33211). Equal amounts of proteins were used per IP (standard: 2 mg) and protein lysate volumes were adjusted to equal volumes with IP buffer (IPB; NEB2-150, 0.3% NP40, 1mM DTT, and 1x PIC).

Streptavidin beads preparation for affinity purification

Streptavidin M-280 Dynabeads (Thermo Fisher) were equilibrated three times with IPB (see above) by overhead rotation for 10 min at 4°C and subsequently pre-blocked in IPB + 1% cold fish gelatin rotating 4°C for 1 hour. Finally, beads were taken up in IPB (starting volume). 40 μl of pre-blocked Streptavidin M-280 beads were added to lysates and incubated overnight rotating at 4°C.

High stringency washes and on-bead digestion for ChromID

After incubation of nuclear lysates with beads rotating at 4°C overnight, beads were separated from the unbound fraction on a magnetic rack and washed twice with 2% SDS in TE (+ 1mM DTT, 1x PIC) for 10 min rotating overhead at room temperature (RT), once with high salt buffer (HSB; 50mM HEPES pH 7.5, 1mM EDTA, 1% Triton X-100, 0.1% deoxycholate, 0.1% SDS, 500mM NaCl, 1mM DTT, and 1x PIC) for 10 min at RT, once with DOC buffer (250mM LiCl, 10mM Tris pH 8.0, 0.5% NP40, 0.5% deoxycholate, 1mM EDTA, 1mM DTT, and 1x PIC) for 10 min at 4°C, and twice with TE buffer (+ 1mM DTT,

1x PIC) for 10 min at 4°C. After the washes, beads were isolated from the last TE wash on a magnetic rack and the proteins were digested with 0.5 μg trypsin (Promega; V5111) in 40ul digestion buffer (1M Urea in 50mM Tris pH 8.0, 1mM Tris-(2-carboxyethyl)-phosphin (TCEP)) directly on beads, overnight at 26°C and shaking at 600 rpm. Next day, the digested protein-peptide mix was isolated from beads and reduced with 2mM TCEP for 45 minutes at RT, and then alkylated with 10mM Chloroacetamide (ClAA) for 30min at RT in the dark. The digestion was stopped by acidifying the peptides with Trifluoracetic acid (TFA) to a final concentration of 0.5%, and the acetonitrile (ACN) concentration was adjusted to 3% prior loading on C18 StageTips.

C18 StageTips clean-up

Obtained peptides were cleaned-up using in-house produced (Functional Genomics Center Zurich, FGCZ) C18-StageTips. StageTips were humidified with 100% methanol (MeOH), cleaned twice with 60% ACN; 0.1% TFA, and conditioned twice with 3% ACN; 0.1% TFA. Peptides were loaded onto the StageTips, and the collected flow-through was loaded again. Afterwards, the peptides were desalted twice with 3% ACN; 0.1% TFA, and finally eluted twice with 60% ACN; 0.1% TFA. Desalted peptides were shock frozen in liquid nitrogen (N2), completely dried in a speed vacuum centrifuge, and subsequently resolved in 3% ACN; 0.1% formic acid (FA), containing internal retention time standard peptides (iRT Kit Ki-3002-1, Biognosys).

Detection of biotinylated proteins by data-dependent acquisition (DDA) mass spectrometry

We used an Easy-nLC 1000 HPLC system operating in trap / elute mode (trap column: Acclaim PepMap 100 C18, 3um, 100A, 0.075x20mm; separation column: EASY-Spray C18, C18, 2um, 100A, 0.075x500mm, Temp: 50°C) coupled to an Orbitrap Fusion mass spectrometer (Thermo Scientific). Trap and separation column were equilibrated with 12 ul and 6 ul solvent A (0.1 % FA in water), respectively. 2 μl of the resuspended sample solution was injected onto the trap column at constant pressure (500 bar) and peptides were eluted with a flow rate of 0.3 μl/min using the following gradient: 2 % - 25 % B (0.1 % FA in ACN) in 50 min, 25 % - 32 % B in 10 min an 32 % - 97 % B in 10 min. After 10 min of washing by 97 % B. High accuracy mass spectra were acquired with an Orbitrap Fusion mass spectrometer (Thermo Scientific) using the following parameter: scan range of 300-1500 m/z, AGC-target of 4e5, resolution of 120'000 (at m/z 200), and a maximum injection time of 50 ms. Data-dependent MS/MS spectra were recorded in top speed mode in the linear ion trap using quadrupole isolation (1.6 m/z window), AGC target of 1e4, 300 ms maximum injection time, HCD-fragmentation with 30 % collision energy, a maximum cycle time of 3 sec, and all available parallelizable time was enabled. Mono isotopic precursor signals were selected for MS/MS with charge states between 2 and 7 and a minimum signal intensity of 5e3. Dynamic exclusion was set to 25 sec and an exclusion window of 10 ppm. After data collection, the peak lists were generated using automated rule-based converter control[70] and Proteome Discoverer 2.1 (Thermo Scientific).

Protein identification and label-free protein quantification of DDA data

Raw data were processed with MaxQuant (version 1.5.3.30) and its built-in Andromeda search engine for feature extraction, peptide identification and protein inference[71]. The

mouse reference proteome (UniProtKB/Swiss-Prot and UniProtKB/TrEMBL) version 2018_12 combined with manually annotated contaminant proteins was searched with protein and peptide false discovery rates (FDR) values set to 1%. Match-between-runs algorithm was enabled. All MaxQuant parameters can be found in the uploaded parameterfile: rpx40_mqpar.xml (deposited in the PRIDE repository). Perseus (versions 1.6.1.1) was used for statistical analysis[72]. Results were filtered to remove reverse hits and proteins only identified by site. Further, only proteins found in at least 3 replicates were kept. Missing values were imputed from a 1.8 standard deviations left-shifted Gaussian distribution with a width of 0.3 (relative to the standard deviation of measured values). Potential interactors were determined using a two sample t-test using s0 = 0.1 and 1 (details shown in volcano plots) and permutation-based FDR = 0.01[73] and visualised by volcano plots. Obtained results were exported and further visualised using the statistical computer language R (version 3.5.2).

Estimation of protein abundance by data-independent acquisition (DIA)

Cells were grown to about 90% confluency on 10 cm dishes (approximately 15 x 10^6 cells), harvested by trypsinisation, and pelleted by centrifugation at 1000 rpm for 5 min. Cell nuclei were extracted following the nuclear extraction procedure (described above) until digestion of genomic DNA. Nuclei were then pelleted by centrifugation at 2000 g for 10 min and resuspended in 30 μl lysis buffer (4% (w/v) SDS, 100 mM Tris/HCL pH 8.2). Lysate was incubated at 95 °C for 5 min under 1000 rpm shaking, followed by centrifugation at 16000 g for 10 min at RT. Supernatant was processed immediately using FASP[74] using 50 μg of total protein as measured by QubitTM Protein Assay. Tryptic peptides were cleaned-up using in-house produced C18-StageTips. Peptides were resuspended in 3% acetonitrile, 0.1% formic acid in water including iRT standard peptides.

Data-independent acquisition (DIA) was performed on an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, San Jose, CA) coupled online to an Acquity UPLC M-class (Waters, Milford, MA) using a PicoView 565 nanospray source (New Objective). Peptide mixtures were separated in a single-pump trap/elute mode, using a trapping (nanoEase Symmetry C18, 5 μm, 180 μm × 20 mm) and an analytical column (nanoEase HSS T3 C18, 100 Å, 1.8 μm, 75 μm x 250 mm). Solvent A was water, 0.1% formic acid and solvent B was acetonitrile, 0.1% formic acid. 0.5 μg peptides/sample were loaded with a constant flow of 0.5% solvent B, at 15 μL/min onto the trapping column. Trapping time was 0.5 min. Peptides were eluted via the analytical column with a constant flow of 300 nL/min. During the elution step, the percentage of solvent B increased in a nonlinear fashion from 8% to 22% in 82 min and 22% to 32% in 8 min. In brief, following MS settings were applied: MS1 scan at 120'000 Orbitrap resolution with an AGC target of 1×10^6 and max. injection time of 118 ms in the mass range of 350 to 1205 m/z, followed by 50 DIA scans covering a precursor mass range of 400 to 1000 m/z with isolation window widths of 12 m/z. The scan resolution in the Orbitrap was set to 15'000 with an AGC target of 1×10^6 and max. injection time of 25 ms. The HCD collision energy was set to 30%.

Data Analysis

In brief, raw files were analysed in Spectronaut Pulsar (13.9.191106.43655, Biognosys) by library-free DirectDIA. The basic principles of DirectDIA analysis have been previously described by[75]. The searches were done against the mouse reference proteome (UniProtKB/ Swiss-Prot) and a eGFP reference. Search results were filtered at 1% FDR on precursor and protein group level. Only the top 3 peptides were used for label-free protein intensity calculation. The protein group report of significant proteins was further used for plotting in R.

Functional gene set enrichment and network visualisation

All proteins identified previously were mapped to human STRING identifiers via the gene names and sequence similarity. Functional gene set enrichment was performed using the "Proteins with Values/Ranks" functionality in STRINGv11[76] for each chromatin reader. The log2 fold changes over background were used. From all terms enriched in any of the chromatin readers, nine Gene Ontology Cellular Component terms (The Gene Ontology Consortium, 2019) which were significantly enriched in at least one of the readers were selected. Cytoscape (version 3.7.1) was used to layout the 79 proteins that were identified in ChromID experiments and are members of at least one of the selected GO terms. Visualisation was based on GO term membership only. Each protein was represented by a pie chart which signifies in which reader the protein was significantly detected after LS-MS/MS. STRING interaction confidences were added as links between proteins, with a cutoff set at confidence 0.4. For foreground protein network visualisation, all proteins with a positive log2 fold change in any of the chromatin readers compared to nBASU were considered as foreground. Their protein-protein interaction network was retrieved from STRINGv11[76] with an interaction confidence threshold of 0.7. The network was imported into Cytoscape (version 3.7.1) and visualized using the "Prefuse Force Directed OpenCL Layout".

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Allis CD, Jenuwein T. The molecular hallmarks of epigenetic control. Nat Rev Genet. 2016; 17:487–500. [PubMed: 27346641]

2. Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol. 2007; 14:1025–1040. [PubMed: 17984965]

3. Musselman CA, Lalonde M-E, Côté J, Kutateladze TG. Perceiving the epigenetic landscape through histone readers. Nat Struct Mol Biol. 2012; 19:1218–1227. [PubMed: 23211769]

4. Ruthenburg AJ, Li H, Patel DJ, David Allis C. Multivalent engagement of chromatin modifications by linked binding modules. Nat Rev Mol Cell Biol. 2007; 8:983–994. [PubMed: 18037899]

5. Vermeulen M, et al. Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers. Cell. 2010; 142:967–980. [PubMed: 20850016]

6. Bartke T, et al. Nucleosome-Interacting Proteins Regulated by DNA and Histone Methylation. Cell. 2010; 143:470–484. [PubMed: 21029866]

7. Nikolov M, et al. Chromatin Affinity Purification and Quantitative Mass Spectrometry Defining the Interactome of Histone Modification Patterns. Molecular & Cellular Proteomics. 2011; 10

8. Eberl HC, Spruijt CG, Kelstrup CD, Vermeulen M, Mann M. A Map of General and Specialized Chromatin Readers in Mouse Tissues Generated by Label-free Interaction Proteomics. Molecular Cell. 2013; 49:368–378. [PubMed: 23201125]

9. Spruijt CG, et al. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. Cell. 2013; 152:1146–1159. [PubMed: 23434322]

10. Mittler G, Butter F, Mann M. A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. Genome Research. 2008; 19:284–293. [PubMed: 19015324]

11. DEjardin J, Kingston RE. Purification of proteins associated with specific genomic Loci. Cell. 2009; 136:175–186. [PubMed: 19135898]

12. Liu X, et al. In Situ Capture of Chromatin Interactions by Biotinylated dCas9. 2017; :1–36. DOI: 10.1016/j.cell.2017.08.003

13. Myers SA, et al. Discovery of proteins associated with a predefined genomic locus via dCas9–APEX-mediated proximity labeling. Nat Meth. 2018; :1–8. DOI: 10.1038/s41592-018-0007-1

14. Schmidtmann E, Anton T, Rombaut P, Herzog F, Leonhardt H. Determination of local chromatin composition by CasID. Nucleus. 2016; 7:476–484. [PubMed: 27676121]

15. Fischle W, et al. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes & Development. 2003; 17:1870–1881. [PubMed: 12897054]

16. Bernstein E, et al. Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Mol Cell Biol. 2006; 26:2560–2569. [PubMed: 16537902]

17. Bannister AJ, et al. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature. 2001; 410:120–124. [PubMed: 11242054]

18. Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature. 2001; 410:116–120. [PubMed: 11242053]

19. Vermeulen M, et al. Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4. Cell. 2007; 131:58–69. [PubMed: 17884155]

20. Meehan RR, Lewis JD, Bird AP. Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. Nucleic Acids Research. 1992; 20:5085–5092. [PubMed: 1408825]

21. Nan X, Meehan RR, Bird A. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. Nucleic Acids Research. 1993; 21:4886–4892. [PubMed: 8177735]

22. Baubec T, Ivanek R, Lienert F, Schübeler D. Methylation-dependent and -independent genomic targeting principles of the MBD protein family. Cell. 2013; 153:480–492. [PubMed: 23582333]

23. Pasini D, Bracken AP, Hansen JB, Capillo M, Helin K. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. Mol Cell Biol. 2007; 27:3769–3779. [PubMed: 17339329]

24. Tsumura A, et al. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes to Cells. 2006; 11:805–814. [PubMed: 16824199]

25. Peters AHFM. Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. Nat Genet. 2001; 30:77–80. [PubMed: 11740497]

26. Xu J, et al. Super-Resolution Imaging of Higher-Order Chromatin Structures at Different Epigenomic States in Single Mammalian Cells. CellReports. 2018; 24:873–882.

27. Hiragami-Hamada K, et al. Dynamic and flexible H3K9me3 bridging via HP1β dimerization establishes a plastic state of condensed chromatin. Nature Communications. 2016; 7:1–16.

28. Yap KL, et al. Molecular Interplay of the Noncoding RNA ANRIL and Methylated Histone H3 Lysine 27 by Polycomb CBX7 in Transcriptional Silencing of INK4a. Molecular Cell. 2010; 38:662–674. [PubMed: 20541999]

29. Jørgensen HF, Ben-Porath I, Bird AP. Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. Mol Cell Biol. 2004; 24:3387–3395. [PubMed: 15060159]

30. Brown K, et al. The molecular basis of variable phenotypic severity among common missense mutations causing Rett syndrome. Human Molecular Genetics. 2016; 25:558–570. [PubMed: 26647311]

31. Feller C, Forné I, Imhof A, Becker PB. Global and specific responses of the histone acetylome to systematic perturbation. Molecular Cell. 2015; 57:559–571. [PubMed: 25578876]

32. Schotta G, et al. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes & Development. 2004; 18:1251–1262. [PubMed: 15145825]

33. Fischle W, et al. Regulation of HP1–chromatin binding by histone H3 methylation and phosphorylation. Nat Cell Biol. 2005; 438:1116–1122.

34. Roux KJ, Kim DI, Raida M, Burke B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. The Journal of Cell Biology. 2012; 196:801–810. [PubMed: 22412018]

35. Kim DI, et al. An improved smaller biotin ligase for BioID proximity labeling. Mol Biol Cell. 2016; 27:1188–1196. [PubMed: 26912792]

36. Ramanathan M, et al. RNA–protein interaction detection in living cells. Nat Meth. 2018; 15:207–212.

37. Saksouk N, et al. Erratum to Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation [Molecular Cell 56 (2014) 580-594]. Molecular Cell. 2015; 57:202.

38. Tachibana M, et al. G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes & Development. 2002; 16:1779–1791. [PubMed: 12130538]

39. Peters AHFM, et al. Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. MOLCEL. 2003; 12:1577–1589.

40. Tchasovnikarova IA, et al. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. Science. 2015; 348:1481–1485. [PubMed: 26022416]

41. Goldberg AD, et al. Distinct Factors Control Histone Variant H3.3 Localizationat Specific Genomic Regions. Cell. 2010; 140:678–691. [PubMed: 20211137]

42. Agarwal N, et al. MeCP2 interacts with HP1 and modulates its heterochromatin association during myogenic differentiation. Nucleic Acids Research. 2007; 35:5402–5408. [PubMed: 17698499]

43. Arita K, et al. Recognition of modification status on a histone H3 tail by linked histone reader modules of the epigenetic regulator UHRF1. Proc Natl Acad Sci USA. 2012; 109:12950–12955. [PubMed: 22837395]

44. Ueda J, Tachibana M, Ikura T, Shinkai Y. Zinc finger protein Wiz links G9a/GLP histone methyltransferases to the co-repressor molecule CtBP. Journal of Biological Chemistry. 2006; 281:20120–20128. [PubMed: 16702210]

45. Nozawa R-S, et al. Human POGZ modulates dissociation of HP1α from mitotic chromosome arms through Aurora B activation. Nature Publishing Group. 2010; 12:719–727.

46. Yin Y, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science. 2017; 356

47. Mikkelsen TS, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007; 448:553–560. [PubMed: 17603471]

48. Voigt P, et al. Asymmetrically Modified Nucleosomes. Cell. 2012; 151:181–193. [PubMed: 23021224]

49. Yang X-C, Burch BD, Yan Y, Marzluff WF, Dominski Z. FLASH, a Proapoptotic Protein Involved in Activation of Caspase-8, Is Essential for 3′ End Processing of Histone Pre-mRNAs. MOLCEL. 2009; 36:267–278.

50. Doyon Y, Selleck W, Lane WS, Tan S, Côté J. Structural and Functional Conservation of the NuA4 Histone Acetyltransferase Complex from Yeast to Humans. Mol Cell Biol. 2004; 24:1884–1896. [PubMed: 14966270]

51. Ravens S, Yu C, Ye T, Stierle M, Tora L. Tip60 complex binds to active Pol II promoters and a subset of enhancers and co-regulates the c-Myc network in mouse embryonic stem cells. Epigenetics & Chromatin. 2015; :1–16. DOI: 10.1186/s13072-015-0039-z [PubMed: 25621012]

52. Whetstine JR, et al. Reversal of Histone Lysine Trimethylation by the JMJD2 Family of Histone Demethylases. Cell. 2006; 125:467–481. [PubMed: 16603238]

53. Das PP, et al. Distinct and Combinatorial Functions of Jmjd2b/Kdm4b and Jmjd2c/Kdm4c in Mouse Embryonic Stem Cell Identity. MOLCEL. 2014; 53:32–48.

54. Horton JR, et al. Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. Nat Struct Mol Biol. 2010; 17:38–43. [PubMed: 20023638]

55. Aranda S, Mas G, Di Croce L. Regulation of gene transcription by Polycomb proteins. Science Advances. 2015; 1:e1500737–15. [PubMed: 26665172]

56. Williams K, et al. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature. 2011; 473:343–348. [PubMed: 21490601]

57. Delachat AMF, et al. Engineered Multivalent Sensors to Detect Coexisting Histone Modifications in Living Stem Cells. Cell Chemical Biology. 2018; 25:51–56.e6. [PubMed: 29174541]

58. Mauser R, Kungulovski G, Keup C, Reinhardt R, Jeltsch A. Application of dual reading domains as novel reagents in chromatin biology reveals a new H3K9me3 and H3K36me2/3 bivalent chromatin state. Epigenetics & Chromatin. 2017; :1–19. DOI: 10.1186/s13072-017-0153-1 [PubMed: 28149326]

59. Tekel SJ, et al. Tandem Histone-Binding Domains Enhance the Activity of a Synthetic Chromatin Effector. ACS Synth Biol. 2018; 7:842–852. [PubMed: 29429329]

60. Flemr M, Bühler M. Single-Step Generation of Conditional Knockout Mouse Embryonic Stem Cells. CellReports. 2015; :1–16. DOI: 10.1016/j.celrep.2015.06.051

61. Bibel M, Richter J, Lacroix E, Barde Y-A. Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. Nat Protoc. 2007; 2:1034–1043. [PubMed: 17546008]

62. Abad MA, et al. Borealin–nucleosome interaction secures chromosome association of the chromosomal passenger complex. The Journal of Cell Biology. 2019; 218:3912–3925. [PubMed: 31570499]

63. Manzo M, et al. Isoform-specific localization of DNMT3A regulates DNA methylation fidelity at bivalent CpG islands. EMBO J. 2017; 36:3421–3434. [PubMed: 29074627]

64. Gaidatzis D, Lerch A, Hahne F, Stadler MB. QuasR: quantification and annotation of short reads in R. Bioinformatics. 2015; 31:1130–1132. [PubMed: 25417205]

65. Derrien T, et al. Fast Computation and Applications of Genome Mappability. PLoS ONE. 2012; 7:e30377–16. [PubMed: 22276185]

66. Heintzman ND, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. 2007; 39:311–318. [PubMed: 17277777]

67. Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. Nature Publishing Group. 2017; 12:2478–2492.

68. Akalin A, Franke V, Vlahovi ek K, Mason CE, Schübeler D. genomation: a toolkit to summarize, annotate and visualize genomic intervals. Bioinformatics. 2014; doi: 10.1093/bioinformatics/btu775

69. Feller C, et al. Histone epiproteomic profiling distinguishes oligodendroglioma, IDH-mutant and 1p/19q co-deleted from IDH-mutant astrocytoma and reveals less tri-methylation of H3K27 in oligodendrogliomas. Acta Neuropathologica. 2019; 139:211–213. [PubMed: 31773240]

70. Barkow-Oesterreicher S, Türker C, Panse C. FCC – An automated rule-based processing tool for life science data. Source Code for Biology and Medicine. 2013; 8:3. [PubMed: 23311610]

71. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature Biotechnology. 2008; 26:1367–1372.

72. Tyanova S, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Meth. 2016; 13:731–740.

73. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA. 2001; 98:5116–5121. [PubMed: 11309499]

74. Wi niewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nat Meth. 2009; 6:359–362.

75. Tsou C-C, et al. DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. Nat Meth. 2015; 12:258–264.

76. Szklarczyk D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Research. 2019; 47:D607–D613. [PubMed: 30476243]

77. Sokolova M, et al. Genome-wide screen of cell-cycle regulators in normal and tumor cells identifies a differential response to nucleosome depletion. Cell Cycle. 2017; 16:189–199. [PubMed: 27929715]

78. Ibarra A, Benner C, Tyagi S, Cool J, Hetzer MW. Nucleoporin-mediated regulation of cell identity genes. Genes & Development. 2016; 30:2253–2258. [PubMed: 27807035]

79. Stadler MB, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature. 2011; 480:490–495. [PubMed: 22170606]

80. Liu Z, Scannell DR, Eisen MB, Tjian R. Control of Embryonic Stem Cell Lineage Commitment by Core Promoter Factor, TAF3. Cell. 2011; 146:720–731. [PubMed: 21884934]

81. Morey L, Aloia L, Cozzuto L, Benitah SA, Di Croce L. RYBP and Cbx7 Define Specific Biological Functions of Polycomb Complexes in Mouse Embryonic Stem Cells. CellReports. 2013; 3:60–69.

Editorial summary

The protein complexes associated with specific chromatin marks in living cells are identified using engineered binding proteins.
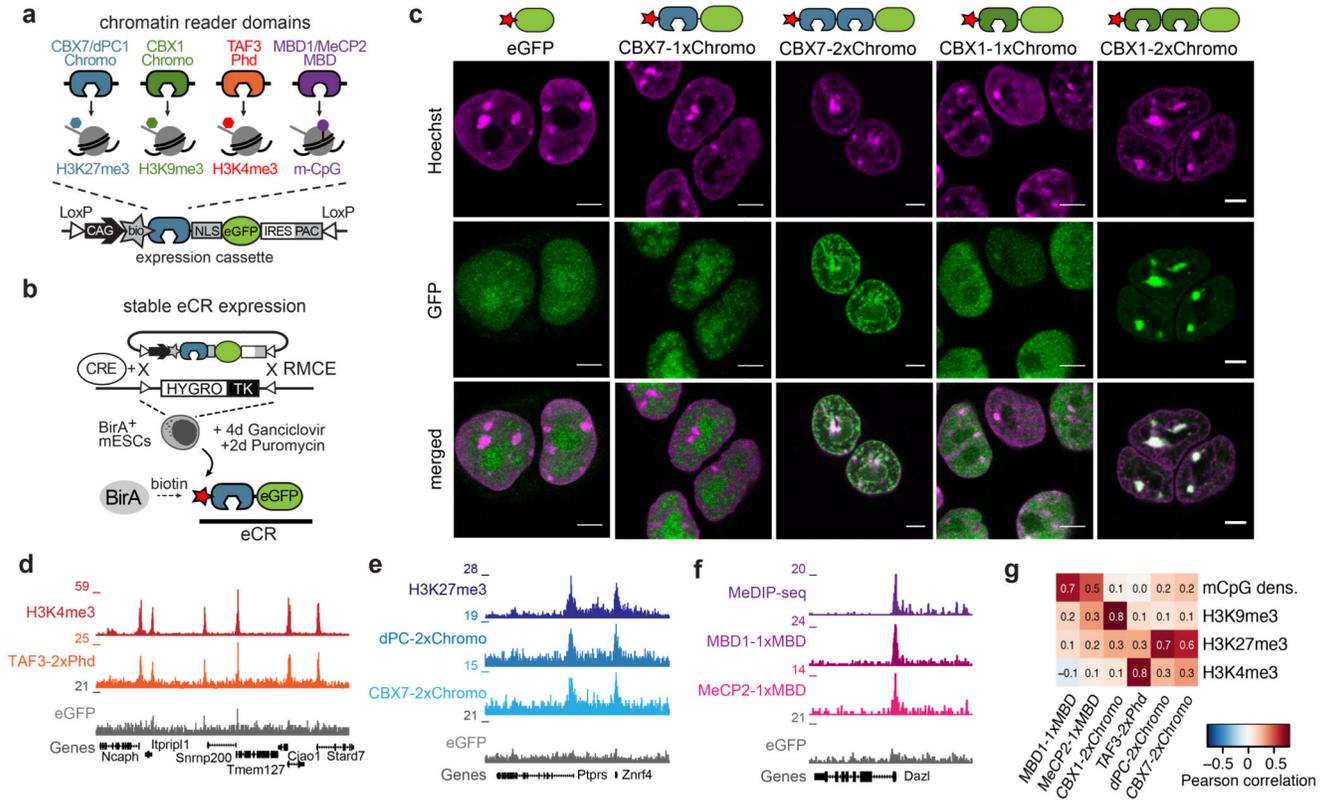
Figure 1. Nuclear and genomic localisation indicates correct eCR interactions with the genome
a) Top: chromatin reader domains used in this study and their specificities towards chromatin marks. Bottom: schematics of constructs utilised to generate mouse embryonic stem cells expressing engineered chromatin readers (eCRs). eCRs are composed of single or dual chromatin reader domains fused in frame to eGFP and a biotin acceptor site (see also Supplementary Fig. 1a). **b)** Recombinase-Mediated Cassette Exchange (RMCE) followed by double selection (Ganciclovir and Puromycin) was applied to generate stable integration of the expression construct at a defined site in the mouse genome. Stably expressed eCRs are *in vivo* biotinylated by a bacterial BirA ligase. **c)** Live imaging shows nuclear localisation of single and dual eCRs in mouse ES cells. Nuclear eGFP serves as control (See also Supplementary Fig. 2). Size bars = 5 μm. Similar results were obtained from 2 independent experiments. **d-f)** Genome browser examples showing correct localisation of eCRs according to chromatin modifications detected by antibody-ChIP-seq and DNA methylation by MeDIP-seq. For histone-PTM readers, only eCRs with two reader domains are shown (See supplementary Fig. 3 for single-domain eCRs). Shown is the library-normalised read density at 100bp intervals. Gene models and the position of CpG islands and repetitive elements are indicated. **g)** Pearson correlation score obtained from comparisons of eCRs with chromatin modifications at selected genomic intervals positive for the interrogated chromatin modifications (N=340291).
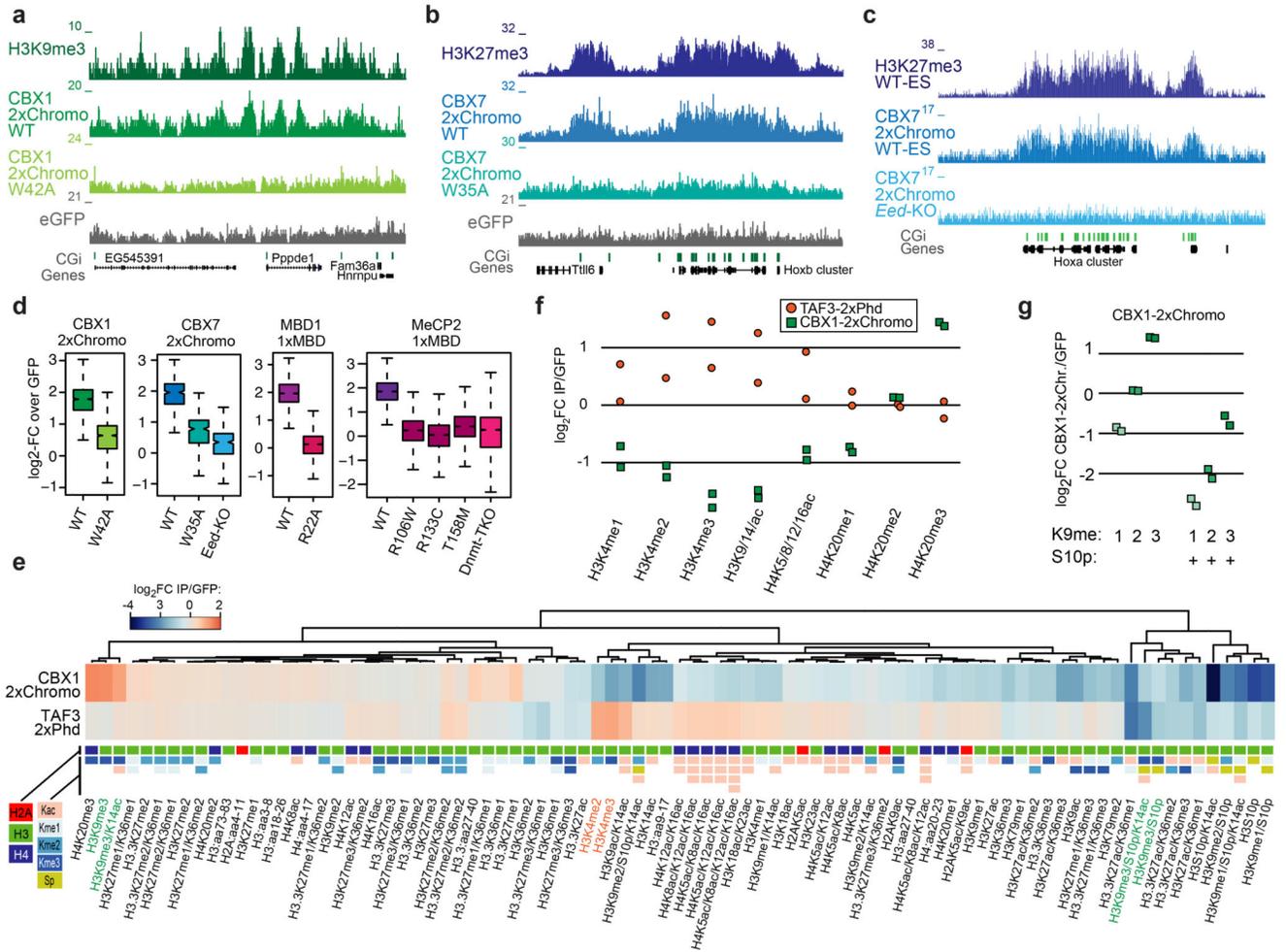
Figure 2. Functional analysis indicates dependency on reader domains and modifications for correct eCR localisation.

a-b) Genome browser examples for loss of binding of mutant-eCRs to sites enriched by wild type eCRs. Shown is the library-normalised read density at 100bp intervals. Gene models and the position of CpG islands and repetitive elements are indicated. **c)** Genome browser example for loss of CBX7-2xChromo-eCR binding to the genome in absence of H3K27me3 in *Eed*-KO cells. **d)** Box plots showing loss of binding of mutant eCRs to sites bound by their wild-type versions (CBX1 N=29879, CBX7 N=3204, MBD1 N=141443, MeCP2 N=27176). In addition, removal of binding substrates such as DNA methylation (*Dnmt*-TKO) or H3K27me3 (*Eed*-KO) results in loss of binding of wild-type eCRs to the corresponding sites. Shown are log2-FC enrichments at peak regions identified for the wild-type eCRs in wild type cells. Boxes denote the interquartile range (IQR) and whiskers 1.5 x IQR. Median values are indicated. **e)** Heatmap showing the histone-PTM log2-FC enrichment scores obtained from CBX1-2xChromo and TAF3-2xPhd eCR ChIP. Histone-PTMs are clustered based on enrichment scores. Colour code below the heatmap indicates histone isoforms and modification types detected in this assay. **f-g)** Dot plots showing enrichment/depletion for selected histone-PTMs, and the effect of combinatorial serine-10

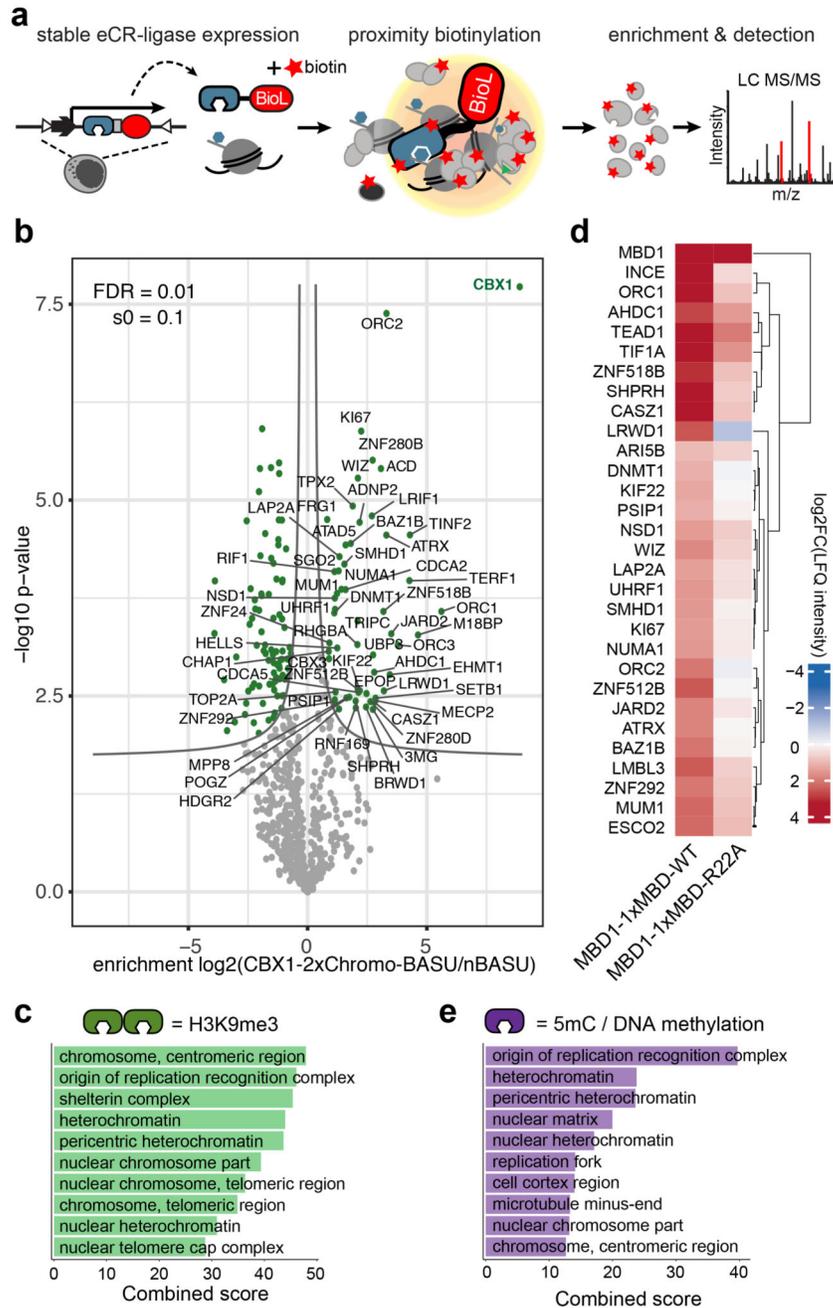phosphorylation on CBX1-2xChromo binding. Results from two independent measurements are shown.

Figure 3. ChromID identifies proteins associated with H3K9me3 and DNA methylation.
a) Schematic describing ChromID using engineered chromatin readers fused to promiscuous biotin ligases (BioL). **b)** Volcano plot showing ChromID results obtained using the CBX1-eCR-BASU targeting H3K9me3 over a reader-free nuclear BASU (nBASU) control. Statistically-enriched proteins are indicated (FDR-corrected two-tailed t-test, FDR = 0.01, s0 = 0.1, log2-FC > 0, n = 4 independent replicates). Peptides used to identify CBX1 match the Chromodomain used in the eCR. **c)** Bar plots representing the top 10 cellular component GO terms summarizing the proteins (N=58) enriched by CBX1-eCR-BASU. The combined

score is calculated by multiplying the ln(p-value) from Fisher's exact test (two-tailed) and the z-score. **d)** Heatmap representation of significantly enriched proteins captured with the 5mC-reader (MBD1-eCR-BASU) and compared to results obtained using a mutant 5mC-reader (R22A-eCR-BASU). Shown are average LFQ intensities (log2-FC) from four independent measurements. Peptides used to identify MBD1 match the MBD domain used in the eCR. **e)** Same as in c, but for proteins enriched by MBD1-eCR-BASU.
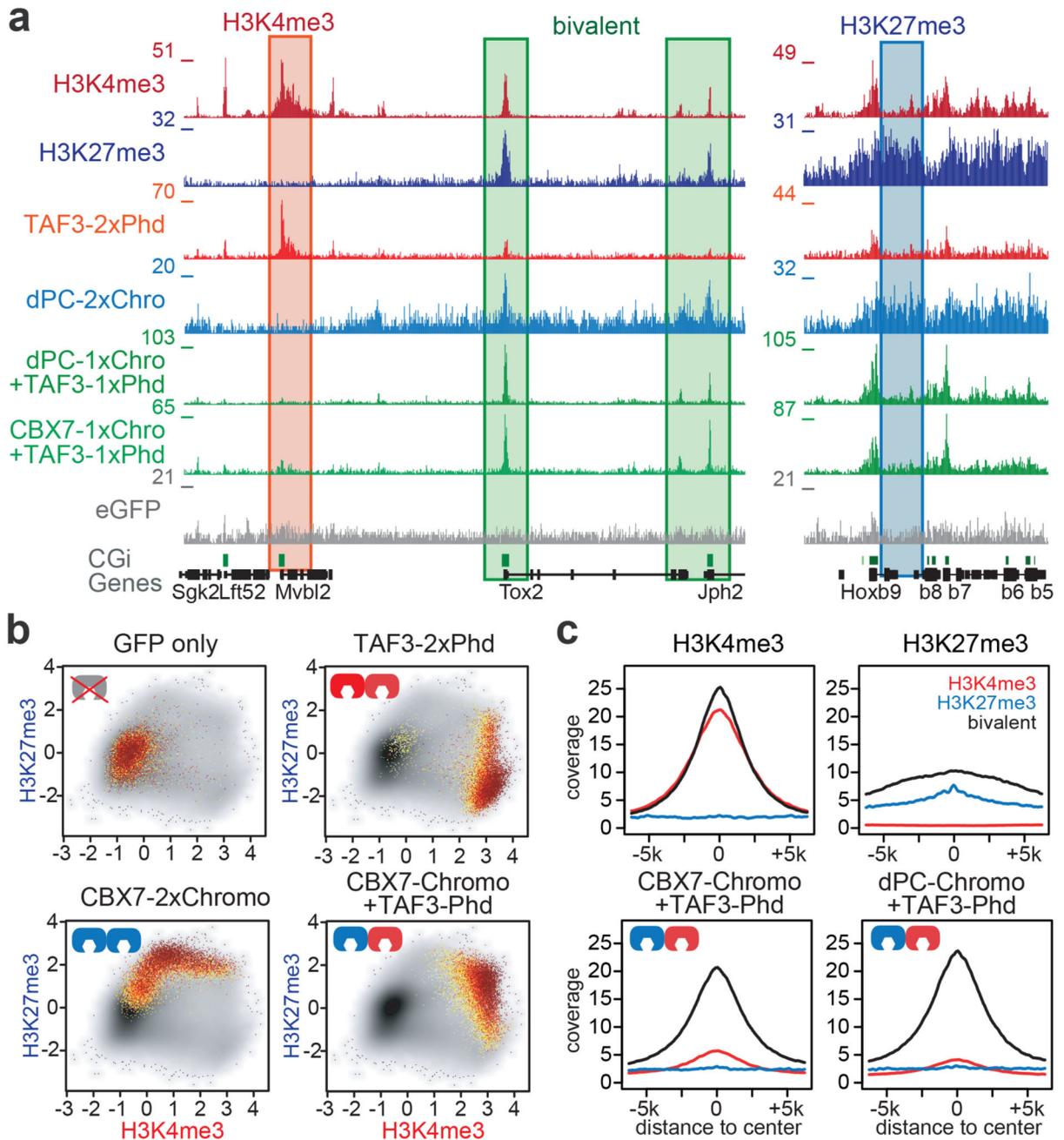
Figure 4. Generation and validation of eCRs reading bivalent H3K4me3 and H3K27me3 marks.
a) Genome browser example showing context-dependent localisation of dual-reader eCRs to bivalent sites decorated by H3K4me3 and/or H3K27me3. Binding is preferentially directed to bivalent sites, while regions modified by H3K4me3- or H3K27me3-only show less recruitment. Gene models and the position of CpG islands and repetitive elements are indicated. **b)** Scatter plots indicating the distribution (highlighted data points) and enrichment (colour) for the tested eCRs along the mouse genome based on H3K27me3 and/or H3K4me3 marks. Shown is the enrichment of H3K27me3 and H3K4me3 at 1kb

windows covering the entire genome (grey). Coloured data points indicate the top 1% genomic windows enriched by the indicated eCR. eCRs specific for one modification separate towards their respective substrates, while the dual reader localises predominantly to the bivalent-modified sites (See also Supplementary Fig. 11d). **c)** Average density profiles around H3K4me3- (red) and H3K27me3-monovalent (blue) or bivalent-peaks (black). Data indicates increased preference of the dual-reader eCR for the bivalent peaks while binding at H3K4me3- and H3K27me3-only peaks is strongly reduced (See also Supplementary Fig. 11e).
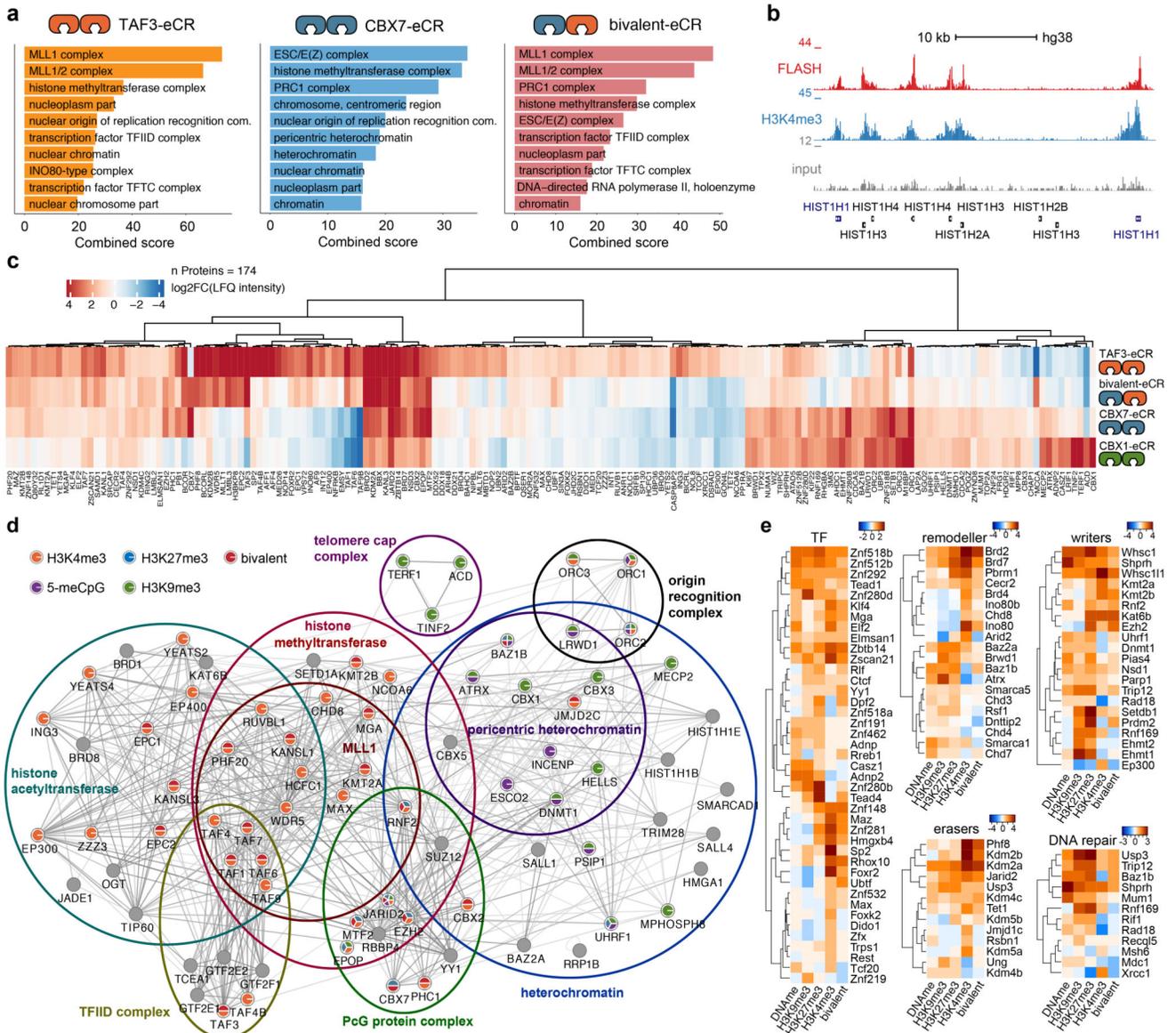
Figure 5. ChromID identifies the proteins associated with key chromatin marks in mouse ES cells.

a) Bar plots representing the top 10 cellular component GO terms enriched at H3K4me3-, H3K27me3- and bivalently-modified chromatin. **b)** Genome browser example for FLASH/CASP8AP2 co-localising at transcribed histone genes marked with H3K4me3. **c)** Heatmap representation of proteins significantly-enriched in either of the ChromID experiments specific for H3K9me3-, H3K4me3-, H3K27me3- or bivalently-modified chromatin in mouse ES cells (FDR-corrected, two-tailed t-test, FDR = 0.01, n = 4 independent replicates). Significance threshold was set to s0 = 0.1 and FDR = 0.01. The LFQ intensities (log2-FC) over nBASU are shown. **d)** Network analysis based on proteins belonging to major cellular component GOs terms identified in at least one ChromID experiments (N proteins = 79). Individual proteins are shown as nodes, edges indicate interactions retrieved from the

STRING database (interaction score > 0.9). Proteins detected in ChromID experiments belonging to the selected GO terms, but not called significant by the two-tailed t-test are shown as grey nodes. Significantly enriched proteins are coloured according to the reader they have been identified. **e)** Heatmap representation of identified factors classified based on their functionality and clustered according to the computed LFQ intensities (log2-FC/nBASU). Proteins were selected based on min 0.5 log2-FC in at least one ChromID experiment. TF: Transcription Factor.