The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation

Susan L Kloet1, Matthew M Makowski1,6, H Irem Baymaz1,6, Lisa van Voorthuijsen1, Ino D Karemaker1, Alexandra Santanach2,3, Pascal W T C Jansen1, Luciano Di Croce2–4 & Michiel Vermeulen1,5

Although the core subunits of Polycomb group (PcG) complexes are well characterized, little is known about the dynamics of these protein complexes during cellular differentiation. We used quantitative interaction proteomics and genome-wide profiling to study PcG proteins in mouse embryonic stem cells (ESCs) and neural progenitor cells (NPCs). We found that the stoichiometry and genome-wide binding of PRC1 and PRC2 were highly dynamic during neural differentiation. Intriguingly, we observed a downregulation and loss of PRC2 from chromatin marked with trimethylated histone H3 K27 (H3K27me3) during differentiation, whereas PRC1 was retained at these sites. Additionally, we found PRC1 at enhancer and promoter regions independently of PRC2 binding and H3K27me3. Finally, overexpression of NPC-specific PRC1 interactors in ESCs led to increased Ring1b binding to, and decreased expression of, NPC-enriched Ring1b-target genes. In summary, our integrative analyses uncovered dynamic PcG subcomplexes and their widespread colocalization with active chromatin marks during differentiation.

PcG proteins are an evolutionarily conserved family of protein complexes that are important for cellular differentiation and development1. They play a key role in regulating the expression of Hox gene clusters, which control body-segmentation patterns2. Polycomb proteins were first discovered and characterized in Drosophila and were later identified in higher eukaryotes3. Polycomb proteins assemble into three major protein complexes: Polycomb repressive complexes 1 and 2 (PRC1 and PRC2), and the Polycomb-repressive deubiquitinase (PR-DUB)4. PRC1 contains a catalytic subunit, Ring1a or Ring1b, both of which are ubiquitin ligases for histone H2A K119 (refs. 5, 6). The PRC2 subunits Ezh1 and 2 mono-, di-, and trimethylates histone H3 K27 (refs. 7, 8). PR-DUB contains a deubiquitinase (BAP1, known as Calypso in Drosophila) for H2A K119 (K118 in Drosophila)9. These three protein complexes contain additional subunits, and several paralogous subunits define distinct subcomplexes10. Furthermore, a number of presumably substoichiometric interactors have been described11,12. In general, Polycomb proteins are associated with gene repression, although recent reports have also hinted at a role in gene activation13–16.

Polycomb protein complexes functionally cooperate in transcriptional gene silencing. For example, the H3K27me3 modification catalyzed by PRC2 is recognized by the Cbx subunits of PRC1 (refs. 17, 18). Furthermore, ubiquitinated H2AK119 (H2AK119Ub) serves as a binding scaffold for PRC2 in vivo19 and stimulates the catalytic activity of PRC2 on nucleosomes in vitro20. However, recent evidence has suggested that H2AK119Ub is not strictly required for transcriptional repression by PRC1, at least during certain stages of development21,22. Nevertheless, the enzymatic activities and cross-talk between PRC1 and PRC2 are thought to result in the generation of large regions of H3K27me3- and H2AK119Ub-marked transcriptionally silent chromatin.

During the past decade, genetic, biochemical, and MS-based approaches from multiple laboratories have elucidated the subunits and interactors of Polycomb complexes in a variety of species. However, systematic, quantitative, and integrative studies are required to understand the dynamics and function of Polycomb protein complexes during cellular differentiation. Here, we set out to characterize the subunit composition, stoichiometry, and architecture of Polycomb complexes in mouse ESCs and neural progenitor cells by using label-free quantitative MS-based proteomics. We complemented this work by profiling the genome-wide binding of PRC1 and PRC2 in the same cell types. Our work revealed striking dynamics of Polycomb complexes during differentiation, at both the proteomic and genomic levels. For PRC2, we identified both stem-cell and neural-specific interactors, and we observed a global downregulation of PRC2 during differentiation despite the maintenance of global H3K27me3 levels. We also observed a switch between variant PRC1 and canonical PRC1 during differentiation, which was accompanied by a drastic change in genome-wide Ring1b binding. In addition to ‘classical’ H3K27me3-marked repressed Polycomb targets, we observed PRC1 binding at active promoters and enhancers enriched for neural functions. Strikingly, these genes were not bound by PRC2, thus revealing a partial decoupling between PRC1 and PRC2 during cellular differentiation. Finally, overexpression of NPC-specific Ring1b interactors in ESCs affected the genomic localization of Ring1b and the expression

1Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, Nijmegen, the Netherlands. 2Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology, Barcelona, Spain. 3Department of Biomedical Genetics, Universitat Pompeu Fabra (UPF), Barcelona, Spain. 4Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain. 5Cancer GenomiCs.nl (CGC.nl) Consortium, Center for Molecular Medicine, UMC Utrecht, Utrecht, the Netherlands. 6These authors contributed equally to this work. Correspondence should be addressed to M.V. (m.vermeulen@ncmls.ru.nl).

Received 30 October 2015; accepted 17 May 2016; published online 13 June 2016; doi:10.1038/nsmb.3248
of Ring1b-target genes, thus indicating that the observed subunit dynamics at least partially drives Polycomb target-gene binding.

RESULTS
Identification of dynamically incorporated PRC2 subunits
To explore the subunit composition and dynamics of PcG complexes during mouse ESC differentiation, we used stably integrated bacterial artificial chromosome (BAC) transgenes containing the endogenous promoter of the gene of interest as well as a C-terminal GFP tag, as described previously\(^\text{23}\). For PRC2, we generated an Eed BAC line in mouse ESCs, which we then differentiated into NPCs. Western blotting revealed nearly endogenous expression levels of GFP-Eed in both cell types (Fig. 1a). We also observed extreme downregulation of core PRC2 members during differentiation to NPCs, at both the protein and transcript levels (Fig. 1a and Supplementary Fig. 1a). The western blot data (Fig. 1a) were in agreement with previously obtained total proteome data\(^\text{24}\), which also show a strong reduction in the abundance of PRC2 core subunits (Fig. 1a, table).

Next, we performed label-free GFP pulldowns followed by LC-MS/MS on nuclear extracts from ESCs and NPCs to assess the subunit composition of PRC2 in both cell types. GFP-Eed pulldowns identified all known core PRC2 subunits (Ezh1 or Ezh2, Suz12, Rbbp4 or Rbbp7, Mtf2, Jarid2, and Aebp2) in both ESCs and NPCs (Fig. 1b,c and Supplementary Fig. 1b). PRC2 was stable in our pulldown conditions, because Suz12 successfully immunoprecipitated Jarid2 (Supplementary Fig. 1b). We also identified C17orf96, a recently identified PRC2 interactor in HeLa cells\(^\text{12,25}\), as an Eed interactor in ESCs and NPCs (Fig. 1b,c and Supplementary Fig. 1b). Additionally, the uncharacterized Gm340 protein (also known as C10orf12 (refs. 12,25)) was present in both ESC and NPC pulldowns. Two PRC2-associated proteins—the F-box protein Fbxw11 and the uncharacterized protein AU022751, which has recently been reported to be a PRC2 interactor in ESCs\(^\text{26}\)—were present in only the ESC pulldowns. We then used the intensity-based absolute quantification (iBAQ) algorithm\(^\text{27}\) to calculate the abundance (stoichiometry) of each interactor relative to Ezh1 and Ezh2. Whereas the stoichiometry of the core subunits remained stable during differentiation, the abundance of substoichiometric interactors varied substantially between the two

Figure 1 PRC2 interactors and architecture during stem-cell differentiation. (a) Western blot (top) of core PRC2 members on nuclear extracts from ESCs and NPCs. Hdac1 is a loading control. Uncropped blots are shown in Supplementary Data Set 1. The bottom table lists the absolute abundance of Eed, Ezh2, and Suz12 in nuclear extracts from each cell type\(^\text{24}\). (b,c) Volcano plots from label-free GFP pulldowns on Eed-GFP ESC (b) and NPC (c) nuclear extracts. Statistically enriched proteins in the Eed-GFP pulldowns (n = 3 pulldowns) are identified by a permutation-based false discovery rate (FDR)-corrected two-sided t test. The label-free quantification (LFQ) intensity of the GFP pulldown relative to the control (fold change (FC), x axis) is plotted against the −log\(_10\) transformed P value of the t test (y axis). Dotted gray lines represent statistical cutoffs. The proteins in the upper-right corner represent the bait (Eed, green) and its interactors. Snrpn, Snrpd2, and Fhl3 are known GFP contaminants. (d) Stoichiometry of Eed-GFP interactors in ESCs and NPCs. The iBAQ value of each protein group is divided by the iBAQ value of the Eed-GFP interactors in ESCs, based on cross-links identified from single affinity-purified Eed-GFP in ESCs. Ambiguous cross-links between paralogous subunits (Rbbp4 and Rbbp7, Ezh1 and Ezh2) are combined in this visualization. (g) Summary of PRC2 architecture in mouse ESCs, based on cross-links from f.
cell types (Fig. 1d). The Eed-GFP interactors C17orf96, Mtf2, Tceb1, and Tceb2 were highly enriched (approximately ten-fold or higher) in ESCs but not NPCs (Fig. 1e). In contrast, Gm340 and Phf19 were nearly ten-fold enriched in NPCs. We also purified Eed-GFP from the ESC nuclear pellet (NP) fraction remaining after nuclear extraction (Supplementary Fig. 1c). These experiments revealed that the core PRC2 complex was similar in the nuclear extract (NE) and the NP pulldown, whereas certain substoichiometric interactors were slightly more abundant in either the NE or NP fraction (Supplementary Fig. 1d). In summary, these results revealed a global downregulation of PRC2 abundance during stem-cell differentiation toward a neural lineage. Furthermore, whereas the core of PRC2 was stable during differentiation, substoichiometric interactors were highly dynamic.

Cross-linking MS reveals the PRC2 architecture in mouse ESCs

To determine the architecture of PRC2 in ESCs, we performed single affinity purification cross-linking MS experiments on the endogenous complex after GFP-Eed purification. We incubated GFP pulldown fractions from GFP-Eed ESC nuclear extracts on-bead with BS3, an amine-to-amine cross-linker, and identified cross-linked peptides with pLink28 (Fig. 1f). Our data were consistent with a PRC2 architecture in which Ezh1 or Ezh2, Suz12, and Eed represent a well-connected core complex, and auxiliary subunits bind at the periphery29 (Fig. 1g). In addition, we observed cross-links between multiple partially overlapping regions on Mtf2, Jarid2, or C17orf96, and Ezh1 or Ezh2. Mtf2 and Jarid2 use overlapping binding surfaces in the Ezh1 and Ezh2 SET domains, whereas Mtf2 and C17orf96 bind to a similar N-terminal region of Ezh1 or Ezh2. This finding suggests that these proteins may bind PRC2 in a mutually exclusive manner, as previously hypothesized30. Altogether, these findings extend knowledge of PRC2 architecture, particularly for ESC-enriched substoichiometric interactors.

Most PRC2 is lost from chromatin during differentiation

To determine the functional consequences of PRC2 downregulation during differentiation, we performed chromatin immunoprecipitation followed by massively parallel DNA sequencing (ChIP–seq) experiments on core PRC2 complex members in ESCs and NPCs. We used endogenous antibodies raised against the Suz12 and Ezh2 subunits. More than 80% of Ezh2 peaks overlapped with Suz12 in both ESCs and NPCs (Fig. 2a). We identified more peaks for Suz12 than for Ezh2, possibly because of the presence of PRC2 complexes containing Ezh1. The gene ontology (GO) terms enriched among PRC2-bound genes were related to developmental processes or cell-type specificity (Supplementary Fig. 2a). Unexpectedly, more than 85% of Suz12 and Ezh2 peaks were lost during differentiation to NPCs, and very few NPC-specific peaks appeared (Fig. 2a,b and Supplementary Fig. 2b,c). This result suggests that lower PRC2 protein levels in NPCs results in a drastic loss of PRC2 binding to chromatin during differentiation. Heat maps centered on ESC Suz12 peaks

Figure 2 ChIP–seq of core PRC2 subunits during stem-cell differentiation. (a) Venn diagrams summarizing the number of peaks called from Suz12 and Ezh2 ChIP–seq in ESCs and NPCs. (b) UCSC genome browser views of binding profiles at three classes of genes with cell-type-independent, ESC-specific, or NPC-specific binding. Chr, chromosome. (c) Suz12, Ezh2, and H3K27me3 occupancy (ChIP–seq read density) in ESCs and NPCs, centered on Suz12 ESC peaks. (d) Average binding profile of Suz12 and Ezh2 in ESCs and NPCs at H3K27me3 peaks in ESCs. (e) Western blot analysis of acid-extracted histones from ESCs and NPCs, detected with the indicated antibodies. Uncropped blots are shown in Supplementary Data Set 1. (f) Average binding profile of Suz12, Ezh2, and H3K27me3 in NPCs at Suz12 NPC peaks.
confirmed the loss of Suz12 and Ezh2 binding in NPCs (Fig. 2c).

Next, we asked what happens to H3K27me3 levels during differentiation, in light of the PRC2 downregulation. Strikingly, we observed that H3K27me3-marked nucleosomes in ESCs retained the mark during differentiation to NPCs (Fig. 2c,d), whereas Suz12 and Ezh2 were lost at these sites. A western blot detecting bulk H3K27me3 levels showed that the abundance of the histone mark also decreased during differentiation but did not decrease nearly as drastically as the PRC2-subunit protein levels (Fig. 2e). H3K27me3 was still deposited at Suz12 and Ezh2 NPC peaks (Fig. 2f), thus indicating that the loss of PRC2 at ESC-marked sites was not due to loss of a functional complex in NPCs. The loss of PRC2 during differentiation did not lead to global changes in the expression of PRC2-target genes, and a large proportion of genes were in fact downregulated after loss of PRC2 (Supplementary Fig. 2d–h).

Together, these findings suggest that downregulation of PRC2 protein levels during differentiation leads to a loss of binding at most Polycomb-target genes during differentiation, whereas, notably, H3K27me3 deposition at most of these sites is maintained.

Identification of dynamically incorporated PRC2 subunits

To identify interactors of the PR-DUB complex, we used a BAC line containing the core subunit Bap1 and followed the same MS workflow as that used for PRC2. We identified the core of the complex (Bap1, Asxl1 and Asxl2, Foxk1 and Foxk2, and Kdm1b) in both cell types (Fig. 3a,b). As previously reported in HeLa cells, the methyl-binding-domain proteins Mbd5 and Mbd6 were identified as PR-DUB interactors in both ESCs and NPCs. Using the iBAQ algorithm, we determined that the stoichiometry of the core complex changed very little during differentiation (Fig. 3c). Bap1 and the Asxl proteins were present in a ratio of ~2:1, thus indicating that Bap1 was dimeric within the PR-DUB complex. Similarly to PRC2, some of the substoichiometric interactors appeared to be cell-type specific (Fig. 3d). The glycosylase Ogt and histone demethylase Kdm1b were enriched in ESC PR-DUB, whereas Mbd5 was enriched in NPCs, in agreement with a role of this protein in brain function. We also purified Asxl2-GFP from ESCs, thus revealing that this protein interacts only with known PR-DUB subunits (Supplementary Fig. 3). Furthermore, this analysis showed that Asxl1 and Asxl2 define mutually exclusive PR-DUB subcomplexes. Because the core of PR-DUB remained relatively stable during differentiation, our results suggest a conserved function of this complex in different cell types.

Identification of dynamically incorporated PRC1 subunits

We used a similar label-free proteomics strategy for PRC1. PRC1 interacts with six different Pcgf proteins in a mutually exclusive manner (Fig. 4a). Each of these Pcgf proteins has a shared (Pcgf2 and Pcgf4; Pcgf3 and Pcgf5) or unique set of interactors. We identified nearly the same set of core subunits and substoichiometric interactors in both ESCs (Fig. 4b) and NPCs (Fig. 4c). However, a dramatic switch occurred with respect to the binding of Pcgf and Cbx family members during differentiation (Fig. 4d–f). iBAQ-derived stoichiometry values revealed that Pcgf6 occupied 60% of Ring1b complexes in ESCs (Supplementary Fig. 4a and 4c). Pcgf6 also remained the predominant Ring1b interactor in the NP fraction (Supplementary Fig. 4c,d). Differentiation to NPCs mainly resulted in an exchange of Pcgf6 for Pcgf4 (Bmi1)-containing PRC1 (54% of NPC complexes, Fig. 4e and Supplementary Fig. 4a,b). We observed a similar switch for the Cbx proteins. Notably, the total Cbx occupancy in Ring1b complexes increased during differentiation (3.5% to 12%) but remained substoichiometric. This result was not an overexpression artifact, because Ring1b-GFP levels remained lower than those of the endogenous Ring1b in both cell types (Fig. 4g). Purification of Flag-tagged Pcgf2 from ESCs also revealed substoichiometric Cbx proteins (Supplementary Fig. 4a). We found Cbx7 only in association with Ring1b in ESCs, as previously reported. Purification of Flag-tagged Pcgf2 from ESCs also revealed substoichiometric Cbx proteins (Supplementary Fig. 4c). We found Cbx7 only in association with Ring1b in ESCs, as previously reported. Purification of Flag-tagged Pcgf2 from ESCs also revealed substoichiometric Cbx proteins (Supplementary Fig. 4). We found Cbx7 only in association with Ring1b in ESCs, as previously reported. Purification of Flag-tagged Pcgf2 from ESCs also revealed substoichiometric Cbx proteins (Supplementary Fig. 4c). We found Cbx7 only in association with Ring1b in ESCs, as previously reported. Purification of Flag-tagged Pcgf2 from ESCs also revealed substoichiometric Cbx proteins (Supplementary Fig. 4a,b). Cbx7 was replaced by Cbx6 in NPCs (Fig. 4f and Supplementary Fig. 4a,b). However, Cbx4 and Cbx8 had the highest stoichiometry values in both ESCs and NPCs. In general, Pcgf6-specific interactors showed the most dynamic changes between ESCs and NPCs, owing to the loss of Pcgf6 from the complex during differentiation to a neural lineage (Fig. 4d). Conversely, Pcgf2 and Pcgf4 interactors were more abundant relative to Ring1b in NPCs, most probably because of the strong
upregulation of Pcgf4 in NPC cells. Interestingly, Phc1, an interactor with Pcgf2 and Pcgf4, was more abundant in ESCs than NPCs. From these results, we conclude that the composition of the core PRC1 complex is highly dynamic during differentiation.

Cross-linking MS reveals the PRC1 architecture in mouse ESCs
We also performed cross-linking MS after Ring1b-GFP affinity purification from ESC nuclear lysates. Our data suggested that Ring1a or Ring1b and Rybp or Yaf2 formed a stable core for PRC1 (Fig. 4h). The Pcgf6 interactor Mga cross-linked directly with Ring1a or Ring1b and Rybp or Yaf2, a result suggesting that it stably interacts with a subset of PRC1 complexes. We also identified cross-links between Pcgf6 and Rybp or Yaf2, and between Pcgf6 and Ring1a or Ring1b. This finding suggests that Pcgf6 binds to Ring1a or Ring1b complexes through an interaction surface mediated by Rybp or Yaf2 and Ring1a or Ring1b (Fig. 4i). We did not identify any cross-links between Ring1b and other Pcgf proteins, probably because Pcgf6 is the predominant Pcgf interactor in ESCs.

PRC1 is retained at H3K27me3-marked sites in ESCs and NPCs
Next, we determined the genomic localization of the PRC1 complex in both cell types by using endogenous antibodies raised against Ring1b and Pcgf2. Nearly all Pcgf2 peaks overlapped with Ring1b peaks in both ESCs and NPCs, thus indicating that Pcgf2 binds chromatin only in the context of PRC1 (Fig. 5a). However, we observed many Ring1b peaks independent of Pcgf2 in both cell types, probably because of the presence of other Pcgf proteins bound to Ring1b. In contrast to the results for PRC2, ~50% of Ring1b peaks (n = 9,968 peaks) were shared between ESCs and NPCs. There were also large groups of peaks that were more than three-fold enriched in only ESCs (n = 6,540 peaks) or NPCs (n = 2,754 peaks) (Fig. 5b, c and Supplementary Fig. 5a–c). GO-term analysis of PRC1-associated genes revealed that ESC-enriched targets were enriched for developmental and morphogenic functions (Supplementary Fig. 5a). These functions were also the top functions in the shared gene set and the NPC-enriched set, but many more biosynthetic and metabolic processes were identified for the genes shared between cell types, thus suggesting a conserved role for PRC1 at housekeeping genes. Mining RNA-seq data revealed no global up- or downregulation of the three classes of PRC1-bound genes (Supplementary Fig. 5d). However, individual transcripts did show up- or downregulation after gain or loss of PRC1 (Supplementary Fig. 5e–g).
During differentiation to NPCs, whereas PRC2 was not (Figs. 5c, d and 2d). Indeed, clustering of NPC-enriched Ring1b sites identified a cluster of Ring1b that did not overlap with Suz12 in NPCs but still colocalized with H3K27me3 (Fig. 5f, cluster 2). These results suggested that PRC1 binds H3K27me3-marked chromatin independently of PRC2.

---

**Figure 5** ChIP–seq of core PRC1 subunits during stem-cell differentiation. (a) Venn diagrams summarizing the number of peaks called from Ring1b and Pcgf2 ChIP–seq in ESCs and NPCs. (b) UCSC genome browser views of binding profiles at three classes of genes with shared, ESC-enriched, or NPC-enriched binding. (c) Ring1b, Pcgf2, and H3K27me3 occupancy (ChIP–seq read density) in ESCs and NPCs at three clusters of genes centered on the union of all Ring1b peaks. ESC-enriched peaks have more than three-fold more reads at the Ring1b-binding site relative to NPCs, and NPC-enriched peaks have more than three-fold more reads at the Ring1b-binding site relative to ESCs. (d) Average binding profile of Ring1b and Pcgf2 in ESCs and NPCs at H3K27me3 peaks in ESCs. (e) K-means clustering of Ring1b ESC-enriched peaks on Ring1b, GFP-Ring1b, H3K27me3, H3K4me1, and H3K4me3 ChIP–seq results. (f) K-means clustering of Ring1b NPC-enriched peaks on Ring1b, Suz12, H3K27me3, H3K4me1, and H3K4me3 ChIP–seq results. (g) Reads per kilobase per million mapped reads (RPKM) values for genes located near Ring1b peaks from each of the four clusters identified in (f). For all box plots: midline, median; box limits, 25th percentile (first quartile) and 75th percentile (third quartile); upper whisker, min(max(x)), third quartile +1.5× interquartile range (IQR; third-quartile minus first-quartile values); lower whisker, max(min(x)), first quartile −1.5× IQR.
Figure 6 Overexpression of NPC-enriched PRC1 subunits affects genomic localization of Ring1b. (a) Stoichiometry values for Cbx proteins bound to Ring1b in mock- or Cbx4-transfected Ring1b-GFP BAC ESCs. Data are shown as mean ± s.d. (n = 3 pulldowns). (b) Western blots of Ring1b coimmunoprecipitations in mock- or Pcgf4-transfected ESCs. Uncropped blots are shown in Supplementary Data Set 1. In, input. (c,d) ChIP–qPCR of Ring1b, IgG, and histone H3 at two NPC-enriched loci (c) and at two ESC-enriched loci (d) in mock-, Cbx4- or Pcgf4-transfected ESCs. Data shown are from one representative out of three ChIP experiments and are plotted as mean ± s.d. (n = 3 technical replicates). (e) qRT–PCR analysis of gene expression levels in mock-, Cbx4- or Pcgf4-transfected ESCs. Data shown are from one representative transfection out of three and are plotted as mean ± s.d. (n = 3 technical replicates).

Ring1b binds enhancers independently of H3K27me3

k-means clustering revealed three discrete classes of ESC-enriched Ring1b-binding sites (Fig. 5e). The first cluster represented H3K4me3- and H3K27me3-marked bivalent genes enriched for developmental GO terms (Fig. 5e and Supplementary Fig. 5f). Cluster two represented H3K4me1-marked enhancers, which were also enriched for developmental GO terms (Fig. 5e and Supplementary Fig. 5f). A third cluster contained low amounts of H3K27me3, H3K4me1, and H3K4me3. We also performed ChIP–seq in the GFP-Ring1b BAC line by using an anti-GFP antibody, which showed a very good overlap with the endogenous Ring1b ChIP–seq reads (Fig. 5e). NPC-enriched Ring1b sites also colocalized with histone marks for enhancers (H3K4me1, Fig. 5f, cluster 3, 58% of peaks) or active genes (H3K4me3, Fig. 5f, cluster 4, 10% of peaks). RNA-seq analysis (Fig. 5g) of genes from the H3K27me3-marked clusters showed that these loci were repressed independently of PRC2 binding (cluster 1 versus cluster 2). There was an appreciable amount of transcription from genes near the H3K4me1-marked chromatin, and transcript levels from the H3K4me3-marked chromatin were high, as expected. Interestingly, genes near H3K4me1-marked chromatin were almost exclusively involved in neural function (Supplementary Fig. 5j), thus suggesting that these putative enhancer regions were required for NPC propagation or terminal differentiation.

Altering PRC1 composition affects genomic PRC1 binding

Finally, we sought to determine whether the subunit switching that we observed for PRC1 during differentiation affects genomic localization of Ring1b and consequently gene expression. To address this, we overexpressed two NPC-enriched PRC1 subunits, Cbx4 and Pcgf4, in ESCs. Cbx4 overexpression and subsequent label-free GFP-Ring1b purification resulted in a nearly three-fold increase in Cbx4 stoichiometry relative to Ring1b in the Ring1b BAC-GFP line compared with a mock-transfected control (Fig. 6a). No changes were apparent in the Cbx7 stoichiometry, thus suggesting that Cbx4 overexpression increases the pool of Ring1b containing Cbx4, rather than replacing Cbx7 within the complex. Furthermore, coimmunoprecipitation analysis revealed that the transfected Pcgf4 interacted with Ring1b (Fig. 6b). Next, we performed ChIP–qPCR to determine whether Cbx4 and Pcgf4 overexpression results in increased Ring1b binding at genomic sites that are predominantly bound in NPCs. We chose two target genes for this analysis (Nodal and Orc2) that are highly expressed in ESCs relative to NPCs (Supplementary Fig. 5g) and that gain Ring1b binding during differentiation (Supplementary Table 1). Cbx4 and Pcgf4 overexpression in ESCs, compared with a mock-transfected control, led to increased Ring1b binding at these NPC-enriched sites (Fig. 6c). ChIP–qPCR at ESC-enriched loci (Fig. 6d) revealed no changes in Ring1b recruitment after Cbx4 overexpression. Increased Ring1b binding to Nodal and Orc2 correlated with decreased expression of these genes, whereas the ESC-enriched loci were not affected after Cbx4 and Pcgf4 overexpression (Fig. 6e). These experiments serve as a proof of principle demonstrating that perturbation of Ring1b interactors affects the genomic localization of the PRC1 complex and the expression of PRC1-target genes.

DISCUSSION

We performed both proteomic and genomic characterization of Polycomb group protein complexes during ESC differentiation to a neural lineage. This work revealed several key observations regarding Polycomb biology. First, we found that the interactomes of PRC1 and PRC2 are highly cell-type specific, whereas the subunits of PR-DUB are largely cell-type invariant. The total abundance of Ring1b-containing complexes is mildly downregulated during differentiation, whereas PRC2 is drastically downregulated (at least 30-fold) in NPCs relative to ESCs. Genome-wide analyses confirmed a loss of PRC2 from chromatin, despite bulk H3K27me3 levels being largely maintained. Strikingly, only a fraction (32%) of genome-wide Ring1b sites in neural progenitor cells represent ‘canonical’ Polycomb targets, which are characterized by broad H3K27me3 domains. The remaining 68% of Ring1b sites in NPCs
are at enhancers and active promoters. Finally, we integrated our proteomic and genomic data to show that expression of cell-type-enriched Polycomb interactors affect target-gene specificity and expression.

Our quantitative MS-based proteomics workflows allowed us to identify known and new components of the Polycomb machinery during stem-cell differentiation. Furthermore, the application of the iBAQ algorithm to our data added an additional layer of information, stoichiometry, which was essential for the identification of dynamic interactors. For example, we identified C17orf96 and Mtf2 as high-confidence Eed interactors in ESCs and NPCs, but only after calculating the iBAQ values were we able to determine that both of these proteins were highly enriched in the ESC PRC2 complex. The workflow presented here may be a powerful tool to characterize dynamic protein-protein interactions in a variety of cellular contexts and after cellular perturbation (such as cell-cycle stages, growth-factor stimulation, and DNA damage)\(^1\).

Our data revealed a number of substoichiometric PRC2 interactors including Mtf2, C17orf96, Gm340, and Tceb1 and Tceb2. From our data, we cannot conclude whether these proteins assemble together in a subset of PRC2 complexes or whether all of these proteins interact stoichiometrically with PRC2 in cells and are partially lost during affinity purification. Alternatively, some of the substoichiometric PRC2 interactors may define mutually exclusive PRC2 subcomplexes with distinct functions. Our cross-linking data support the latter hypothesis, because we found that some of the substoichiometric interactors share a binding surface on Ezh1 or Ezh2. Clearly, the tagging, purification and cross-linking of substoichiometric PRC2 interactors is needed to substantiate these findings.

One intriguing observation was the near-complete loss of PRC2 from chromatin in NPCs despite the continued presence of H3K27me3. Previously, it has been shown that PRC2 binds to and colocalizes with H3K27me3 in the G1 phase of the cell cycle and remains at sites of DNA replication in S phase, thus ensuring propagation of the modification during cell division\(^1\). The cell-cycle duration of NPCs is nearly three times longer than that of ESCs (30 h versus ~10 h)\(^37,38\). We hypothesize that a longer cell cycle requires less PRC2 to propagate the mark during cell division. Alternatively, once cells are lineage committed, less de novo H3K27me3, and consequently less PRC2, is required to maintain H3K27me3 domains.

Finally, in the NPCs we identified thousands of Ring1b-binding sites devoid of H3K27me3. These binding sites are typically very narrow (<1 kb), especially relative to the classical Polycomb domains. These binding sites may represent examples of permissive chromatin ‘sampling’ by PRC1 (ref. 39). However, the functional relevance of Ring1b binding at these sites remains unclear. Interestingly, several recent studies have also reported a link between PRC1 and gene activation\(^1\)–\(^1\^). Clearly, further work is required to determine whether distinct PRC1 subcomplexes also play a role in gene activation during neural lineage commitment.

**METHODS**

Methods and any associated references are available in the online version of the paper.

**Accession codes.** ChiP–seq reads and peak files have been deposited in the GEO repository under accession code GSE74330. Proteomics data have been deposited in the ProteomeXchange via the PRIDE partner repository under identifier PXD003758.

**Note.** Any Supplementary Information and Source Data files are available in the online version of the paper.
25. Alekseyenko, A.A., Gorchakov, A.A., Kharchenko, P.V. & Kuroda, M.I. Reciprocal interactions of human C10orf12 and C17orf96 with PRC2 revealed by BioTAP-XL cross-linking and affinity purification. *Proc. Natl. Acad. Sci. USA* **111**, 2488–2493 (2014).

26. Maier, V.K. et al. Functional proteomic analysis of repressive histone methyltransferase complexes reveals ZNF518B as a G9A regulator. *Mol. Cell. Proteomics* **14**, 1435–1446 (2015).

27. Schwahnhausser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).

28. Yang, B. et al. Identification of cross-linked peptides from complex samples. *Nat. Methods* **9**, 904–906 (2012).

29. Ciferri, C. et al. Molecular architecture of human polycomb repressive complex 2. *eLife* **1**, e00005 (2012).

30. Shen, X. et al. Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. *Cell* **139**, 1303–1314 (2009).

31. Baymaz, H.I. et al. MBD5 and MBD6 interact with the human PR-DUB complex through their methyl-CpG-binding domain. *Proteomics* **14**, 2179–2189 (2014).

32. Maussion, G. et al. Investigation of genes important in neurodevelopment disorders in adult human brain. *Hum. Genet.* **134**, 1037–1053 (2015).

33. Morey, L. et al. Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. *Cell Stem Cell* **10**, 47–62 (2012).

34. O’Loghlen, A. et al. MicroRNA regulation of Cbx7 mediates a switch of Polycomb orthologs during ESC differentiation. *Cell Stem Cell* **10**, 33–46 (2012).

35. Smits, A.H. & Vermeulen, M. Characterizing protein-protein interactions using mass spectrometry: challenges and opportunities. *Trends Biotechnol.* [http://dx.doi.org/10.1016/j.tibtech.2016.02.014](http://dx.doi.org/10.1016/j.tibtech.2016.02.014) (2016).

36. Hansen, K.H. et al. A model for transmission of the H3K27me3 epigenetic mark. *Nat. Cell Biol.* **10**, 1291–1300 (2008).

37. Savatier, P., Huang, S., Szekely, L., Wiman, K.G. & Samarut, J. Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts. *Oncogene* **9**, 809–818 (1994).

38. Conti, L. et al. Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol.* **3**, e283 (2005).

39. Klose, R.J., Cooper, S., Farcas, A.M., Blackledge, N.P. & Brockdorff, N. Chromatin sampling: an emerging perspective on targeting polycomb repressor proteins. *PLoS Genet.* **9**, e1003717 (2013).
ONLINE METHODS

Cell culture and BAC-line generation. R1 mESCs were obtained from the ATCC and cultured as described in Spruijt et al.23. NPCs were differentiated and propagated according to a protocol from Conti et al.38. All cell lines were tested for mycoplasma contamination. BACs were tagged according to a protocol from Poser et al.23. Tagged BAC lines were prepared on NucleoBond BAC 100 columns (Macherey-Nagel) and transfected into ESCs with Lipofectamine LTX Plus (Invitrogen), then subjected to G418 selection for 10–12 d. Individual colonies were picked, expanded, and screened for GFP expression.

Nuclear extracts and nuclear-pellet solubilization. Nuclear extracts were prepared essentially according to Dignam et al.40. Cells were harvested with trypsin, washed twice with PBS, and centrifuged at 400 g for 5 min at 4 °C. Cells were resuspended in two volumes of buffer A (10 mM HEPES-KOH, pH 7.9, 15 mM MgCl₂, and 100 mM KCl), and then pelleted at 400 g for 5 min at 4 °C. Cells were resuspended in two volumes of buffer A plus protease inhibitors and 0.15% NP-40 and transferred to a Dounce homogenizer. After homogenization through 30–40 strokes with a type B pestle, the lysates were spun at 3,200g for 15 min at 4 °C. The nuclear pellet was washed once with PBS and spun at 3,200g for 5 min at 4 °C. The nuclear pellet was resuspended in two volumes of buffer C (420 mM NaCl, 20 mM HEPES-KOH, pH 7.9, 20% (v/v) glycerol, 2 mM MgCl₂, and 0.2 mM EDTA) with 0.1% NP-40, protease inhibitors, and 0.5 mM dithiothreitol. The suspension was incubated with rotation for 1 h at 4 °C and then spun at 18,000g for 15 min at 4 °C. The supernatant was divided into aliquots and stored at −80 °C until further use.

The nuclear pellets remaining after nuclear extraction were resolubilized by resuspension in four volumes of RIPA buffer (150 mM NaCl, 50 mM Tris, pH 8.0, 1% NP-40, 5 mM MgCl₂, and 10% glycerol) plus benzonase (Millipore) at 1,000 U per 100 µl nuclear pellet. Samples were incubated at 37 °C with shaking until solubilized, then spun at 18,000g for 15 min at 4 °C. The supernatant was divided into aliquots and stored at −80 °C until further use.

Label-free pulldowns and cross-linking MS. Label-free GFP pulldowns were performed in triplicate as previously described12 with the following modifications. For GFP pulldowns, 2 mg of nuclear extract was incubated with 7.5 µl GFP Nano trap beads (Chromotek) and 50 µg/mL ethidium bromide in buffer C (300 mM NaCl, 20 mM HEPES-KOH, pH 7.9, 20% (v/v) glycerol, 2 mM MgCl₂, and 0.2 mM EDTA) with 0.1% NP-40, protease inhibitors, and 0.5 mM dithiothreitol in a total volume of 400 µl. Six washes were performed: two with buffer C and 0.5% NP-40, two with PBS and 0.5% NP-40, and two with PBS. Flag pulldowns were performed as above but with 10 µl anti-Flag M2 affinity gel (Sigma). Affinity-purified proteins were subjected to on-bead trypsin digestion, as previously described12. Tryptic peptides were acidified and desalted with (Sigma). Affinity-purified proteins were subjected to on-bead trypsin digestion, pulldowns were performed as above but with 10 µl anti-Flag M2 affinity gel (Sigma). Affinity-purified proteins were subjected to on-bead trypsin digestion, and 0.2 mm EDTA) with 0.1% NP-40, protease inhibitors, and 0.5 mM dithiothreitol. The suspension was incubated with rotation for 1 h at 4 °C and then spun at 18,000g for 15 min at 4 °C. The supernatant was divided into aliquots and stored at −80 °C until further use.

Label-free quantitation (LFQ) LC-MS/MS analysis. The tryptic peptides from Ring1b-GFP ESC and NPC pulldowns, Bap1-GFP ESC pulldowns, and Flag-PgR2 pulldowns were separated with a 94-min gradient from 9–32% buffer B followed by washes at 50% and then 95% buffer B for 120 min of total data-collection time. Mass spectra were recorded on an LTQ-Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific), and the top ten most intense precursor ions were selected for fragmentation. The tryptic peptides from Bap1-GFP NPC pulldowns, Axd2-GFP ESC pulldowns, and all NP pulldowns were measured by developing a gradient from 9–32% buffer B for 114 min before washes at 50% then 95% buffer B were performed, for 140 min of total data-collection time. Mass spectra were recorded on an LTQ-Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific). Scans were collected in data-dependent top-speed mode with dynamic exclusion set at 60 s.

LFP peptide identification and analysis. Thermo RAW files from LFP AP MS/MS were analyzed with MaxQuant version 1.5.1.0 with default settings and searching against the UniProt mouse proteome, release 2014_09. Additional options for ‘match between runs’, LFP, and iBAQ were selected. Stoichiometry calculations and volcano plots were produced essentially as previously described12 with Perseus version 1.4.0.8 and in-house R scripts. Proteins used to generate the volcano plots are listed in Supplementary Table 2. Statistical cutoffs were chosen such that no proteins were present as outliers on the control, non-GFP side of the volcano plot.

Single affinity cross-linking LC-MS/MS and data analysis. Samples were purified by GFP affinity purification from nuclear extracts with high-stringency washing conditions as described above. Cross-linked and digested peptides were measured on an LTQ-Orbitrap QExactive as described previously28. Thermo RAW files were converted to mgf format with MSeConvert29 and analyzed with pLink28. Identifications were accepted with a 0.05 FDR. To increase confidence in reported identifications, all spectral matches were additionally filtered to include only matches of peptide length 5 and 40, a precursor error of <10 p.p.m., and an e value <0.01. Filtered cross-links are listed in Supplementary Table 3. Cross-link maps were produced with x!NET44.

Chromatin preparation. Attached mESCs and NPCs were cross-linked with 1% formaldehyde for 10 min at room temperature with gentle shaking. Cross-linking was quenched with the addition of 1/10 volume 1.25 M glycine. Cells were washed with PBS, then harvested by scraping in buffer B (20 mM HEPES, 0.25% Triton X-100, 10 mM EDTA, and 0.5 mM EGTA). Cells were pelleted by centrifugation at 600 g for 5 min at 4 °C. Cell pellets were resuspended in buffer C (150 mM NaCl, 50 mM HEPES, 1 mM EDTA, and 0.5 mM EGTA) and rotated for 10 min at 4 °C. Cells were pelleted by centrifugation at 600 g for 5 min at 4 °C. The cell pellet was then resuspended in 1× incubation buffer (0.15% SDS, 1% Triton X-100, 150 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, and 20 mM HEPES) at 15 million cells/ml. Cells were sheared in a Bioruptor Pico sonicator (Diagenode) at 4 °C with 5 or 7 cycles of 30 s on, 30 s off for mESCs and NPCs, respectively. Sonicated material was spun at 18,000g for 10 min at 4 °C, then divided into aliquots and stored at −80 °C.

Chromatin immunoprecipitation. 10 million cells were used as input for library prep, and 5 million cells were used as input for ChIP–qPCR experiments. Chromatin was incubated overnight at 4 °C in 1× incubation buffer (0.15% SDS, 1% Triton X-100, 150 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, and 20 mM HEPES) supplemented with protease inhibitors and 0.1% BSA. Antibody amounts, catalog numbers, and validation information are listed in Supplementary Table 4. A 50:50 mix of Protein A and G Dynabeads (Invitrogen) were added the next day and incubated for 90 min. The beads were washed twice with wash buffer 1 (0.1% SDS, 0.1% sodium deoxycholate, 1% Triton X-100, 150 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, and 20 mM HEPES) once with wash buffer 2 (wash buffer 1 with 500 mM NaCl), once with wash buffer 3 (250 mM LiCl, 0.5% sodium deoxycholate, 0.5% NP-50, 1 mM EDTA, 0.5 mM EGTA, and 20 mM HEPES), and twice with wash buffer 4 (1 mM EDTA, 0.5 mM EGTA, and 20 mM HEPES). After the washing steps, beads were rotated for 20 min at room temperature in elution buffer (1% SDS and 0.1 M NaHCO₃). The supernatant was de-cross-linked with 200 µM NaCl and 100 µg/mL protease K for 4 h at 65 °C. De-cross-linked DNA was purified with MinElute PCR Purification columns (Qiagen). DNA amounts were determined with Qubit fluorometric quantification (Thermo Fisher Scientific). qPCR analysis of ChIP DNA was performed with iQ SYBR Green Supermix (Bio-Rad) on a CFX96 Real-Time System C1000 Thermal Cycler (Bio-Rad). Primers used for qPCR analysis are listed in Supplementary Table 4.
Illumina high-throughput sequencing and data analysis. Libraries were prepared with a Kapa Hyper Prep Kit for Illumina sequencing (Kapa Biosystems) according to the manufacturer’s protocol with the following modifications. 5 ng DNA was used as input, with NEXTflex adapters (Bioo Scientific) and ten cycles of PCR amplification. Post-amplification cleanup was performed with QIAquick MinElute columns (Qiagen), and size selection was performed with an E-gel (300-bp fragments) (ThermoFisher Scientific). Size-selected samples were analyzed for purity with a High Sensitivity DNA Chip on a Bioanalyzer 2100 system (Agilent). Samples were sequenced on an Illumina HiSeq2000 or NextSeq genome analyzer. The 43-bp tags were mapped to the reference mouse genome mm9 (NCBI build 37) with the Burrows–Wheeler Alignment tool (BWA), allowing one mismatch. Only uniquely mapped reads were used for data analysis and visualization. External files used for analysis are listed in Supplementary Table 4.

Mapped reads were filtered for quality, and duplicates were removed. Peak-calling was performed with the MACS 2.0 tool against a reference input sample from the same cell line with a q value of 0.001 and with the ‘–broad’ option enabled for the H3K27me3 ESC sample. Heat mapping and k-means clustering were carried out with a Python package available at http://simonvh.github.io/fluff/. Average binding profiles were generated with an in-house Perl package. All average profiles were obtained by counting tags per 100-bp window. ChIP–seq data sets used for generating heat maps and average profiles were normalized to the total number of uniquely mapped reads. For gene expression analysis, RPKM values were calculated with RNA-seq data from Guttman et al. GREAT was used for GO-term analysis, and P values were computed with a hypergeometric distribution with FDR correction. R was used to generate most of the graphs.

Quantitative reverse-transcription PCR (qRT–PCR). RNA was isolated with an RNeasy Mini Kit (Qiagen), and 1 µg of RNA was used for cDNA synthesis with an iScript CDNA Synthesis Kit (Bio-Rad). qRT–PCR was performed with iQ SYBR Green Supermix (Bio-Rad) on a CFX96 Real-Time System C1000 Thermal Cycler (Bio-Rad). Primers used for qRT–PCR analysis are listed in Supplementary Table 4. Gapdh was used as the reference gene.

40. Dignam, J.D., Lebowitz, R.M. & Roeder, R.G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* 11, 1475–1489 (1983).
41. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* 2, 1896–1906 (2007).
42. Makowski, M.M., Willems, E., Jansen, P.W. & Vermeulen, M. Cross-linking immunoprecipitation-MS (xIP-MS): topological analysis of chromatin-associated protein complexes using single affinity purification. *Mol. Cell. Proteomics* 15, 854–865 (2016).
43. Chambers, M.C. et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* 30, 918–920 (2012).
44. Combe, C.W., Fischer, L. & Rappsilber, J. xiNET: cross-link network maps with residue resolution. *Mol. Cell. Proteomics* 14, 1137–1147 (2015).
45. Mikkelsen, T.S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553–560 (2007).
46. Tiwari, V.K. et al. A chromatin-modifying function of JNK during stem cell differentiation. *Nat. Genet.* 44, 94–100 (2012).
47. Rissing, E.M. et al. Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Mol. Cell* 55, 347–360 (2014).
48. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74 (2012).
49. Guttman, M. et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* 28, 503–510 (2010).
50. McLean, C.Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* 28, 499–501 (2010).