Reciprocal interactions of human C10orf12 and C17orf96 with PRC2 revealed by BioTAP-XL cross-linking and affinity purification

Artyom A. Alekseyenko[a,b], Andrey A. Gorchakov[a,b,c], Peter V. Kharchenko[d,e,1], and Mitzi I. Kuroda[a,b,1]

[a]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115; [b]Department of Genetics, Harvard Medical School, Boston, MA 02115; [c]Institute of Molecular and Cellular Biology, Novosibirsk 630090, Russia; [d]Center for Biomedical Informatics, Harvard Medical School, Boston, MA 02115; and [e]Hematology/Oncology Program, Children's Hospital, Boston, MA 02115

Contributed by Mitzi I. Kuroda, January 13, 2014 (sent for review December 20, 2013)

Understanding the composition of epigenetic regulators remains an important challenge in chromatin biology. Traditional biochemical analysis of chromatin-associated complexes requires their release from DNA under conditions that can also disrupt key interactions. Here we develop a complementary approach (BioTAP-XL), in which cross-linking (XL) enhances the preservation of protein interactions and also allows the analysis of DNA targets under the same tandem affinity purification (BioTAP) regimen. We demonstrate the power of BioTAP-XL through analysis of human EZH2, a core subunit of polycomb repressive complex 2 (PRC2). We identify and validate two strong interactors, C10orf12 and C17orf96, which display enrichment with EZH2-BioTAP at levels similar to canonical PRC2 components (SUZ12, EED, MTF2, JARID2, PHF1, and AEBP2). ChIP-seq analysis of BioTAP-tagged C10orf12 or C17orf96 revealed the similarity of each binding pattern with the location of EZH2 and the H3K27me3-silencing mark, validating their physical interaction with PRC2 components. Interestingly, analysis by mass spectrometry of C10orf12 and C17orf96 interactions revealed that these proteins may be mutually exclusive PRC2 subunits that fail to interact with each other or with JARID2 and AEBP2. C10orf12, in addition, shows a strong and unexpected association with components of the EHMT1/2 complex, thus potentially connecting PRC2 to another histone methyltransferase. Similarly, results from CBX4-BioTAP protein pulldowns are consistent with reports of a diversity of PRC1 complexes. Our results highlight the importance of reciprocal analyses of multiple subunits and suggest that iterative use of BioTAP-XL has strong potential to reveal networks of chromatin-based interactions in higher organisms.

chromatin IP | formaldehyde cross-linking | LC-MS/MS | protein–protein interactions

The organization of the genome into active and silent domains is integral to the fidelity of gene regulation in higher organisms. Since their discovery in *Drosophila*, the genetic factors known collectively as the Polycomb Group (PcG) (1) have emerged as the prototypical epigenetic factors, required for the critical maintenance of gene silencing during development in higher organisms. Polycomb group proteins are known to form large multicomponent complexes that vary in their composition, with a broadly conserved distinction between PRC1 and PRC2 complexes (reviewed in ref. 2). How these classes of key epigenetic factors are targeted to their sites of action, and interact with appropriate partners within their chromatin context, remains an important question.

Biochemical analyses of PRC1 and PRC2 have been invaluable for the discoveries of enzymatic activities, specific binding properties, and strong subunit interactions (reviewed in refs. 2, 3). Recent seminal work on PRC1 has led to an emerging recognition of the diversity of PRC1 subcomplexes in mammals and their possible individual roles in transcriptional repression (4). These discoveries have highlighted the importance of dissecting chromatin-based complexes using multiple strategies. For example, it would be ideal to recover information about both strong and weak interactions of a particular subunit within its chromatin context. However, a biochemical approach typically requires the release of the complex from the DNA to solubilize it, and such parameters (salt and/or detergent) might simultaneously compromise complex integrity. Furthermore, each chromatin complex requires specific conditions for release from the DNA, and it is hard to predict this property in advance.

Given these considerations, we reasoned that establishing cross-links before the first step of affinity purification might enhance the preservation of protein composition of chromatin-associated complexes and provide the additional advantage of characterizing their DNA targets under the same purification regimen. We present our progress toward this goal by analyzing human EZH2, CBX4, and two candidate EZH2 interactors in cultured cells. Our results provide strong validation for interactions of canonical PRC2 with two uncharacterized proteins encoded by the C10orf12 and C17orf96 genes. We propose that, if extended, our approach will reveal chromatin-associated networks composed of distinct subcomplexes of epigenetic factors and their genomic targets.

Results

BioTAP-XL Strategy and Its Application for Human EZH2: Identification of Canonical Subunits and C10orf12 and C17orf96 in PRC2 Complexes. To preserve DNA–protein and protein–protein interactions within their chromatin context, we developed a cross-linking/tandem

Significance

The fidelity of gene expression is regulated by chromosome-associated protein complexes. A traditional approach to characterizing complexes bound to chromosomes requires their release from the DNA to solubilize them. Here we develop an alternative approach, BioTAP-XL, that allows identification of protein–protein interactions while complexes remain linked to the DNA. We focus on protein interactions and genome localization of human EZH2 and two of its relatively uncharacterized interactors, C10orf12 and C17orf96. Our results provide strong evidence for diversity in human Polycomb repressive complexes, which are composed of factors essential for gene silencing during development in higher organisms. We propose that BioTAP-XL is an effective general approach for investigating the composition and subunit diversity of chromosome-associated complexes.

Author contributions: A.A.A. and M.I.K. designed research; A.A.A. and A.A.G. performed research; A.A.A. and P.V.K. analyzed data; and A.A.A., P.V.K., and M.I.K. wrote the paper.

The authors declare no conflict of interest.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE53495).

[1]To whom correspondence may be addressed. E-mail: mkuroda@genetics.med.harvard.edu or peter.kharchenko@post.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1400648111/-/DCSupplemental.

affinity purification approach (BioTAP-XL). First, we fixed crude nuclear extracts with formaldehyde to introduce covalent bonds between DNA and proteins. Second, we used a dual tag for high-affinity two-step purification (Fig. 1). The BioTAP tag includes two epitopes: Protein A (5) and a 75-amino-acid biotinylation targeting sequence that is recognized by endogenous biotin ligases in both prokaryotic and eukaryotic cells (6).

To assess the BioTAP-XL approach in mammalian cells, we performed pulldowns of both N- and C-terminal–tagged EZH2 and C-terminal–tagged CBX4 in human HEK 293T cells. EZH2 is a core component of the mammalian PRC2 complex, has been shown to associate with SUZ12 and EED, and is responsible for the deposition of the H3K27me3 histone modification (reviewed in ref. 3). CBX4 belongs to a family of Polycomb-related CBX proteins known to be involved in a subset of PRC1, but not PRC2, complexes, providing an important contrast with which to compare EZH2-BioTAP results. We used lentiviral transduction to generate stable HEK-293 T-REx cell lines carrying EZH2 constructs with the BioTAP tag at either the N or the C terminus and CBX4 with a C-terminal BioTAP tag. To induce transcription of these clones from the CMV/2xtetO promoter, we incubated cells in the presence of doxycycline for 4 d. We confirmed the expected size and presence of the biotin mark on the tagged proteins (Fig. S1 *A–C*), and proceeded with the BioTAP pulldown using 1.5×10^9 cells.

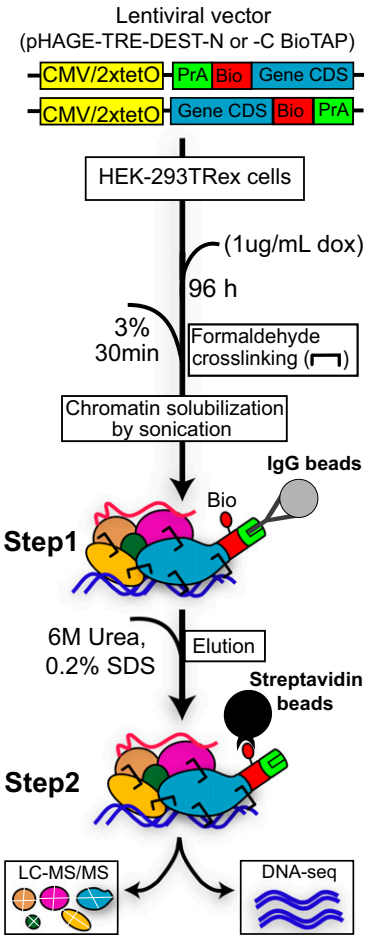

Fig. 1. Overview of the BioTAP-XL purification strategy. The BioTAP tag includes two epitopes: Protein A and Bio, a 75-amino-acid sequence that is biotinylated in vivo. Lentiviral vectors were used to make stable 293T-REx cells expressing N- and C-terminal BioTAP-tagged human proteins. Expression was induced by adding doxycycline (1 μg/mL) to the medium and incubating for 4 d. Crude nuclear extracts were cross-linked using formaldehyde, sonicated, and subjected to tandem affinity purification, first with rabbit IgG–agarose beads eluted under denaturing conditions and subsequently using streptavidin–agarose beads. The resulting DNA was analyzed by high-throughput sequencing. Peptides from the protein fraction were released by direct on-bead trypsin digestion and then identified by LC-MS/MS.

Sequencing of the DNA fraction from the EZH2 BioTAP pulldown closely recapitulated the genome-wide ChIP-seq enrichment profile of H3K27me3, previously published for the 293T cell line (4) (Fig. 2*A*). Specifically, the log-fold enrichment profiles of both N- and C-terminal EZH2 pulldowns showed high correlation with H3K27me3 (Pearson $r = 0.70, 0.72$, respectively; P value $< 10^{-16}$) with a substantial overlap of the genomic regions covered by the respective enrichment domains (Fig. 2*B*). In contrast, the CBX4-BioTAP pattern was distinct and correlated well with previous FLAG-CBX2 ChIP-seq results from the Reinberg laboratory (51.2% of CBX4 sites are cobound by CBX2, and 42.3% of CBX2 sites are bound by CBX4) (4).

Affinity purification after cross-linking may recover abundant proteins based solely on their general proximity on chromatin. Therefore, enrichment over input can be a valuable parameter to identify candidate interactions with the highest specificity. Table 1 reveals the liquid chromatography–mass spectrometry/MS (LC-MS/MS) analysis of the results from the N- and C-terminal EZH2-BioTAP pulldowns presented in two categories: first, the most enriched components, only rarely recovered in the input, and second, the factors that are recovered efficiently in both input and purification. The first group contains the known core members of PRC2 (reviewed in ref. 3) as SUZ12, EED, and MTF2 comprised the top three EZH2-interacting proteins in both pulldowns. Similarly, enriched interactions of RBBP4 and RBBP7 histone-binding members of PRC2 (reviewed in ref. 2), the AEBP2 cofactor (7, 8), PHF1 known to stimulate enzymatic conversion of H3K27me2 to H3K27me3 (9, 10), and the JARID2 modulator of histone methyltransferase activity (11) were confirmed in both experiments (Table 1, with full data sets in Table S1 and Datasets S1 and S2). We also noted the absence of EZH1 from EZH2 pulldowns, in agreement with previous results in human and mouse that EZH1 and EZH2 may form mutually exclusive complexes (12–14). Enriched proteins identified in the EZH2 purifications were mostly missing from the CBX4 pulldown (Table 1 and Table S1), thereby demonstrating specificity for PRC2.

Most importantly, among the enriched EZH2 interactions we found two relatively uncharacterized human proteins: C10orf12, reported to be ubiquitously expressed (http://biogps.org/gene/26148/), and C17orf96, a potential homolog of mouse E13 (E130012A19Rik) protein implicated in epigenetic regulation of neuronal differentiation (15). C10orf12 and C17orf96 lack apparent homologs in *Drosophila*, but are conserved from bony fishes to humans, without discernible protein domains suggestive of their possible functions in the complex. Both proteins were recently identified as potential members of the PRC2 complex via their interaction with the EED protein in HeLa cells (16). That each protein was discovered in two different human cell lines and in association with two distinct PRC2 subunits makes a strong case for an authentic interaction with PRC2. Furthermore, the mouse homolog E13 (E130012A19Rik), also known as esPRC2p48, has been implicated in PRC2 function in mouse ES cells (14).

Reciprocal Mass Spectrometry and ChIP Localization of C10orf12 and C17orf96 Validate Their Strong Interactions with PRC2. To rigorously validate and extend these associations, we performed reciprocal BioTAP pulldowns on tagged versions of C10orf12 and C17orf96 proteins (Fig. S1 *D* and *E*). In both cases, C10orf12 and C17orf96 pulldowns recovered core members of the PRC2 complex (EZH2, EED, SUZ12) among the top significantly interacting proteins (Table 2). Furthermore, the ChIP-seq analyses of tagged C10orf12 and C17orf96 show genome-wide patterns of large enrichment domains closely matching those of EZH2 and H3K27me3 (Fig. 3, Pearson $r = 0.72$ for C17orf96, 0.68 for C10orf12). These results strongly support a model in which C10orf12 and C17orf96 are authentic interactors with the PRC2 complex.

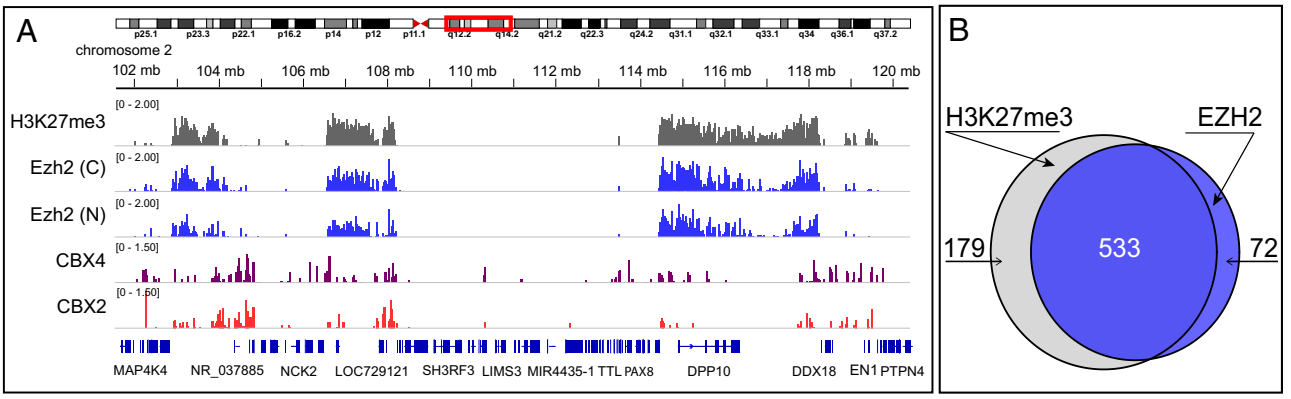

Fig. 2. PRC1 and PRC2 show distinct distributions in 293T-REx cells. (*A*) Representative ChIP-seq profiles in a region of chromosome 2 demonstrate the similar binding patterns of PRC2-associated EZH2-CBioTAP, EZH2-NBioTAP, and H3K27me3. In contrast, PRC1-associated CBX4-CBioTAP and Flag-His–tagged CBX2 display similar binding patterns that are distinct from the PRC2 profiles. H3K27me3 and Flag-His–tagged CBX2 profiles are from ref. 4. (*B*) Venn diagram illustrates the substantial overlap (measured in Mbp) of EZH2 and H3K27me3 large enrichment domains.

The magnitude of the C10orf12 and C17orf96 binding data was not significant enough to clearly evaluate localization of these factors at a resolution beyond the pronounced large enrichment domains. It is possible that incorporation of tagged C10orf12 and C17orf96 proteins into chromatin was relatively inefficient because they were expressed transiently after induction or could not compete fully with native versions. Analyses using antibodies to the endogenous proteins will be important to extend these studies in the future. Therefore, it remains possible that C10orf12 and C17orf96 might occupy additional, distinct locations in the genome, which would be indicative of the formation of PRC2 independent complexes (see below).

Table 1. Peptide counts for top interactions from EZH2 pulldowns

	EZH2		Input		Mock		
Group 1	C	N	R1	R2	R1	R2	CBX4-C
SUZ12	124 (26)	102 (26)	0	0	0	0	2 (2)
EED	81 (14)	69 (18)	0	0	0	0	0
MTF2	75 (22)	85 (28)	0	0	0	0	0
EZH2-bait	103 (13)	105 (20)	0	1	0	0	0
C10ORF12	66 (27)	64 (30)	0	0	0	0	0
JARID2	40 (18)	52 (27)	0	0	0	0	0
PHF1	21 (10)	24 (14)	0	0	0	0	0
C17ORF96	15 (7)	18 (10)	0	0	0	0	0
AEBP2	9 (6)	16 (9)	0	0	0	0	0
PHF19	8 (6)	12 (10)	0	0	0	0	0
RBBP7	16 (4)	16 (8)	0	1	0	0	5 (2)
SKIDA1	4 (3)	4 (4)	0	0	0	0	0
LCOR	3 (3)	4 (3)	0	0	0	0	2 (2)
RBBP4	10 (7)	12 (5)	2 (1)	3 (2)	0	0	2 (2)
SCML2	2 (2)	3 (3)	0	0	0	0	57 (18)
Others							
EZH1	0	1	0	0	0	0	0
Group 2							
PARP1	31 (18)	51 (30)	46 (26)	43 (21)	0	0	66 (34)
TOP2A	11 (8)	16 (14)	16 (12)	20 (13)	0	0	34 (26)
SMC1A	4 (4)	13 (12)	8 (7)	8 (8)	0	0	18 (17)
SSRP1	8 (8)	8 (7)	10 (8)	10 (7)	0	0	14 (11)
H2AFY	14 (7)	19 (1)	24 (9)	16 (8)	0	0	9 (6)
TOP1	5 (4)	10 (10)	10 (5)	8 (6)	0	0	8 (5)
MKI67	13 (12)	25 (24)	38 (31)	30 (24)	0	0	33 (30)

Total peptides are listed in each column, with the number of unique peptides recovered from the EZH2-CBioTAP (Ez-C), EZH2-NBioTAP (Ez-N), CBX4-CBioTAP (CBX4-C), and Mock (untagged) 293T-Rex cells in parentheses. Group 1 proteins are the most highly enriched over input, whereas group 2 proteins are abundant in both pulldown and input.

C10orf12 and C17orf96 May Define Distinct PRC2 Subcomplexes. Interestingly, despite its similar genomic distributions and association with all core PRC2 proteins, C10orf12 did not appear in the tagged C17orf96 reciprocal pulldown, and vice versa, indicating that the two may represent different states of the PRC2 complex. Such a distinction is further supported by the difference in other top-interacting proteins. For example, C10orf12 shows strong association with the components of the H3K9me1/2 HMTase complex (EHMT1/EHMT2/WIZ/CDYL) (17). Interestingly, CDYL has been implicated in recognition of H3K27me3 and also interacts functionally with PRC2 (18). It is not clear at this point if C10orf12 is a subunit of both PRC2 and EHMT complexes or if it might orchestrate interplay between H3K9me1/2 and H3K27me3-mediated gene repression. An example of this coordination has recently been demonstrated in the context of the inactive X chromosome in mouse ES cells (19). Alternatively, this finding could support a previously reported in vitro and in vivo ability of EHMT1 and EHMT2 to contribute to H3K27 methylation (20). Recently, Mozzetta et al. (21) reported that EHMT1 and EMHT2 (GLP and G9a) interact physically and functionally with PRC2 in vivo and more specifically with the core PRC2 components, including EZH2 in vitro. However, we did not find significant recovery of EHMT1 and EHMT2 in our EZH2-BioTAP pulldowns. We demonstrate that, if this connection exists, it is more likely to be mediated through C10orf12 than directly by the PRC2 core complex.

Although both C10orf12 and C17orf96 show strong association with MTF2—a Tudor domain protein that binds H3K36me3 and is proposed to guide PRC2 to transcribed regions—C17orf96 shows significant association with another *Drosophila* PCL homolog, PHF19, also implicated in guiding PRC2 to H3K36me3-containing nucleosomes (22–24).

Ubiquitin-specific proteases USP7 and USP11 were shown to interact with PRC1 proteins and have been implicated in the regulation of PRC1 activity (25, 26). We and others confirm the interaction of USP7, but not that of USP11, with CBX4-containing PRC1 (Table S1). Neither USP7 nor USP11 appear to interact in the context of EZH2-BioTAP-XL (Table 1), but strikingly, we observe pronounced USP7 and USP11 peptide counts in C10orf12 pulldowns and only USP7 among C17orf96 interactors (Table 2).

One of the strong interaction partners found in C10orf12 pulldowns is an uncharacterized ZNF518B protein, not known to be related to PRC1 or PRC2 complexes (and not found in our EZH2 or CBX4 pulldowns) (Table 1 and Table S1). Its function and the nature of its interaction with C10orf12 thus remain to be

Table 2. Peptide counts for top interactions from the C10ORF12 and C17ORF96 pulldowns

Protein name	C10ORF12-N		C17ORF96		EZH2		Input	Mock
	1	2	N	C	C	N	R1	R1
Common hits between C10ORF12 and C17ORF96								
SUZ12	43 (16)	49 (18)	121 (22)	141 (28)	124 (26)	102 (26)	0	0
MTF2	32 (16)	30 (16)	71 (24)	94 (25)	75 (22)	85 (28)	0	0
EED	31 (13)	37 (16)	64 (13)	64 (14)	81 (14)	69 (18)	0	0
EZH2	26 (8)	36 (13)	88 (16)	79 (17)	103 (13)	105 (20)	0	0
RBBP7	7 (4)	5 (2)	21 (6)	29 (7)	16 (4)	16 (8)	0	0
RBBP4	4 (3)	5 (3)	14 (5)	11 (4)	10 (7)	12 (5)	2 (1)	0
PHF1	6 (5)	6 (4)	0	8 (6)	21 (10)	24 (14)	0	0
PHF19	1	1	10 (5)	16 (9)	8 (6)	12 (10)	0	0
EZH1	3 (3)	4 (3)	5 (3)	11 (6)	0	1	0	0
HDAC2	7 (4)	7 (4)	7 (5)	5 (3)	0	0	0	0
PKM	7 (6)	9 (8)	42 (16)	14 (13)	0	2 (2)	0	0
USP7	10 (9)	16 (13)	9 (8)	6 (5)	0	2 (2)	0	0
Other reported PRC2 subunits								
JARID2	0	0	0	0	40 (18)	52 (27)	0	0
AEBP2	0	0	0	0	9 (6)	16 (9)	0	0
C10ORF12 specific hits								
C10ORF12	207 (41)	243 (42)	0	1	66 (27)	64 (30)	0	0
EHMT1	60 (26)	76 (24)	0	0	0	3 (2)	0	0
EHMT2	54 (21)	68 (22)	1	1	1	3 (3)	0	0
WIZ	40 (22)	45 (22)	0	1	0	3 (3)	0	0
CDYL	7 (6)	5 (4)	0	0	0	0	0	0
ZNF518B	30 (22)	34 (22)	0	0	0	0	0	0
USP11	21 (17)	24 (16)	0	0	0	0	0	0
ZNF644	10 (9)	12 (10)	0	0	0	0	0	0
CHAMP1	6 (5)	6 (6)	0	0	0	1	0	0
C17ORF96 specific hits								
C17ORF96	0	1	65 (11)	77 (11)	15 (7)	18 (10)	0	0
SYNE2	0	0	92 (73)	33 (30)	0	0	1	0

Total peptides are listed in each column, with the number of unique peptides recovered from two biologically independent replicates of C10ORF12-NBioTAP (C10ORF12-N-1 and C10ORF12-N-2), C17ORF96-NBioTAP (C17ORF96-N), C17ORF96-CBioTAP (C17ORF96-C), EZH2-CBioTAP (Ezh2-C), EZH2-NBioTAP (Ezh2-N), and Mock (untagged) 293T-Rex cells in parentheses.

explored. Another zinc-finger protein, CHAMP1, was present only in C10orf12 and CBX4 pulldowns (Table 2 and Table S1).

Most importantly, neither C10orf12 nor C17orf96 showed substantial interaction with JARID2 or AEBP2. Together, these observations suggest that the PRC2 core (SUZ12, EED, MTF2, and EZH2) interacts with subsets of additional chromatin-associated proteins such as C10orf12, C17orf96, or JARID2 and/or AEBP2, presumably for the purpose of targeting or functional diversity.

Discussion

We have developed BioTAP-XL to enable a comprehensive view of chromatin-associated complexes and have used this approach to explore newly recognized candidate components of human PRC2 complexes in cell culture. This approach was inspired by the realization that intact chromatin complexes might be difficult to characterize using biochemical methods that require their initial removal from the DNA template. We believe that a key strength of BioTAP-XL is its ability to capture such complexes without their initial dissociation from chromatin.

BioTAP-XL employs a two-step purification of cross-linked chromatin, which allows enrichment from complex samples without the complication of isolation of endogenous biotinylated proteins recovered in a one-step purification (27). This also allows highly enriched sequencing of associated DNA fragments starting from the same affinity-purified cross-linked chromatin. Therefore, this is a very practical approach, requiring no prior knowledge of the biochemical properties of a given complex. Here we show that transient expression of an EZH2 cDNA tagged with BioTAP at either its N or C terminus was sufficient for incorporation of the tagged protein into the known locations of PRC2-modified H3K27me3 chromatin in 293T cells. Affinity purification of the tagged EZH2 was likewise able to enrich for C10orf12 and C17orf96, which exhibited similar behavior once tagged and affinity-purified. Therefore, the expression of transgenic cDNAs may be sufficient for success in many cases and could be complemented by tagging of endogenous gene copies for recapitulation of normal regulation when necessary, using recently developed TALEN or CRISPR technologies (28, 29).

Enabling a comprehensive view of nuclear interactions of a given chromatin protein, simultaneous with identification of its genomic location, is a key strength of BioTAP-XL, but may also be considered a drawback compared with classical biochemistry, as no distinction can be made initially to delineate functional subcomplexes. One interesting way to sort this out is to continue characterization of the interaction network by BioTAP-XL of candidate partners of the initial bait protein, looking for distinct reciprocal interaction and genomic distribution patterns and creating a chromatin-specific interaction network. For example, follow-up analysis by BioTAP-XL tagging demonstrated that the EZH2 interactors C10orf12 and C17orf96 are preferentially associated with distinct partners beyond the canonical PRC2 complex identified in their individual protein interaction mass spectrometry lists. A promising avenue of investigation of subcomplexes would be to ask whether post-translational modifications might govern subsets of interactions. Indeed, EZH2 is known to have multiple sites of phosphorylation in vivo (30). Site-specific mutagenesis, eliminating subsets of these sites in a EZH2-BioTAP bait protein, might remove distinct

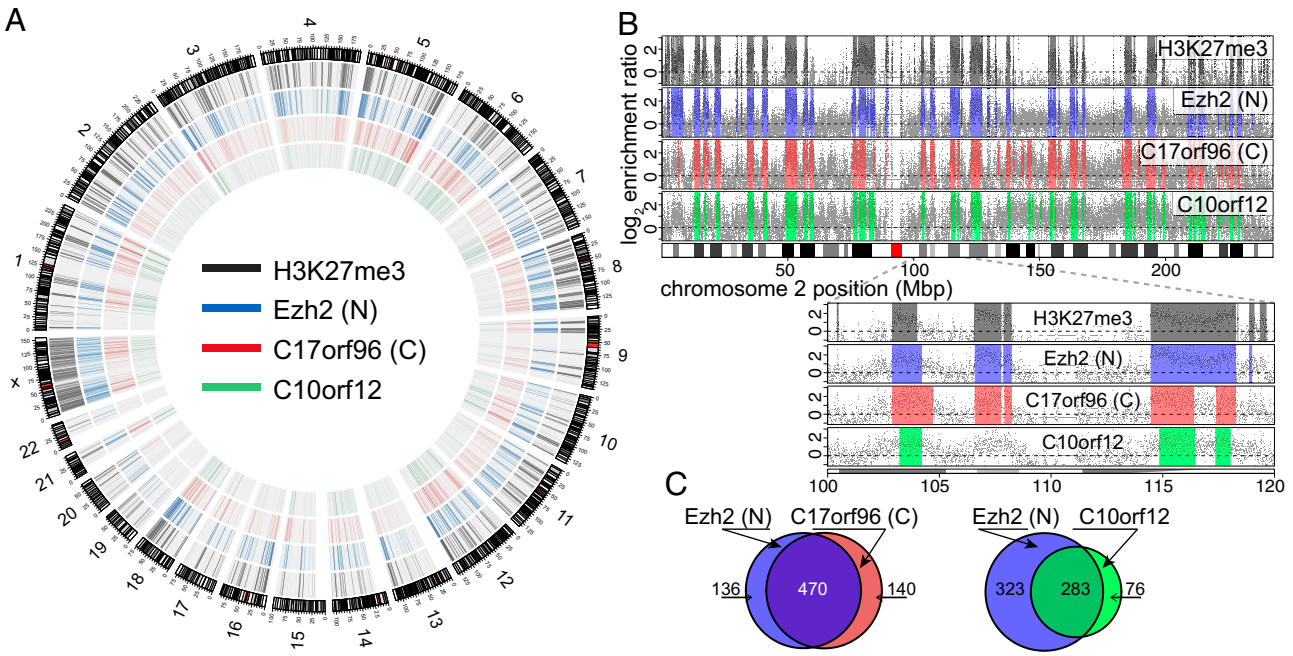

Fig. 3. C17orf96 and C10orf12 protein distributions closely match those of EZH2 and H3K27me3 modification. (*A*) A circos plot showing enrichment (log_2 scale) of C17orf96-CBioTAP and C10orf12-NBioTAP proteins, along with H3K27me3 (4) and EZH2-CBioTAP in the human genome. (*B*) The enrichment estimates (gray dots) and large enrichment domains (color shading) detected by the Hidden Markov model are shown for the entire human chromosome 2, illustrating similarity of the enrichment patterns, with an enlarged view of a 20-Mb region. (*C*) Venn diagrams illustrate the substantial overlap (measured in Mbp) of the large enrichment domains.

subcomplexes from a given affinity purification, thereby revealing how functional interactions may be regulated by posttranslational modifications. In another variation of BioTAP-XL, the tandem tag could be split between putative subunits and tested pairwise, resulting in specific enrichment of only subcomplexes containing both tagged factors. Replacement of the constitutive biotinylation target sequence with the recognition site for *Escherichia coli* BirA biotin ligase expressed under tissue-specific or temporal control (31) could add an additional level of versatility and specificity to our approach in model organisms. In addition to recovery of protein–protein interactions, another advantage of the method may be its potential ability to retrieve chromatin factors bound in proximity rather than through direct interaction. In the future, this could be assessed by comparing results before and after a DNA digestion step.

In summary, we envision that BioTAP-XL can be a powerful and effective tool to complement traditional biochemical and genomic analyses of chromatin-associated protein complexes. Promising initial results using human EZH2, two of its previously uncharacterized interactors, and CBX4 suggest that many links for key epigenetic regulators remain to be explored.

Experimental Procedures

Expression of BioTAP-tagged human ORFs, ChIP-seq, and LC-MS/MS analyses are described in *SI Experimental Procedures*.

BioTAP-XL for Human 293 T-REx Cells. Cells were grown in 150- × 25-mm dishes in DMEM (Invitrogen, catalog #11965), supplemented with 10% (vol/vol) FBS, 1% penicillin/streptomycin. Cells (1.5×10^9) were harvested by spinning for 5 min at 300 × *g*, +4 °C and washed two to three times in 250 mL PBS.

Formaldehyde cross-linking: Harvested 293 T-REx cells were homogenized by using a 100-mL Dounce homogenizer (Bellco, Glass Inc.), with five strokes of each of A and B pestles. For every 4–5 mL of cell pellet volume, 100 mL of NEB buffer + 0.1 mM PMSF prechilled on ice were added. Without delay, 100 mL of cell/nuclear homogenate was poured into a T-225 flask containing a room-temperature mixture of 360 mL of PBS and 40 mL of 37% formaldehyde and incubated for 30 min at 25 °C on an orbital shaker platform with vigorous shaking (100 rpm). Fixed nuclei were pelleted by spinning for 10 min at 4,000 × *g*, +4 °C. The supernatant was carefully decanted, and the nuclear pellet was washed four times with 100 mL of ice-cold PBS with 0.1 mM PMSF, and once with N-sucrose buffer. Nuclei were pelleted between washes for 10 min at 4,000 × *g*, +4 °C. Nuclei were resuspended in N-glycerol buffer and snap-frozen in liquid N_2 before further processing.

Chromatin preparation: Frozen nuclear extracts were thawed and spun-down for 10 min at 4,000 × *g*, +4 °C. Pellets were washed with 10–20 volumes of TE buffer with 0.1 mM PMSF and spun-down for 10 min at 4,000 × *g*, +4 °C. Pellets were resuspended with 10 vol of TE buffer with 0.1 mM PMSF by pipetting up and down. SDS was added to the mixture to a final concentration of 1%. The mixture was inverted in the tube 10 times and spun-down for 10 min at 4,000 × *g*, +4 °C. The supernatant was carefully removed (note: the pellet may be quite loose), and the pellet was resuspended with 10 vol of TE buffer with 0.1 mM PMSF by pipetting and further spun-down for 10 min at 4,000 × *g*, +4 °C. This washing step was repeated twice. The pellet was resuspended with 1.5 vol of TE buffer with 0.1 mM PMSF by pipetting up and down. SDS was added to a mixture to a final concentration of 0.1%. The resulting viscous mixture was sonicated in 4.5 mL aliquots using a Misonix Sonicator 3000 with Microtip power output level 7 and total sonication processing time of 5 min, 15 s pulse "on" and 45 s "off" time, to generate DNA fragments in the range of 300–3,000 bp. Triton X-100 (1% final) and NaCl (140 mM final) were added to the sonicated samples. Samples were mixed on a rotating wheel for 5 min at +4C and spun-down for 10 min at 10,000 × *g*, +4 °C. The supernatant containing soluble chromatin was collected. [Note: To control for chromatin input composition and quality (for protein, DNA and RNA), a 1-mL aliquot is reserved before proceeding with the next step.]

Protein A-IgG affinity purification and elution: Soluble chromatin was incubated with IgG agarose beads (Sigma, catalog #A2909). For every 10 mL of sonicated chromatin, 0.5–1 mL beads were added. The mixture was rotated end-over-end in the 15-mL Falcon tubes for 12–16 h at +4 °C. Beads were then washed three times for 10 min at +4 °C with 15 bead volumes of RIPA buffer and spun-down for 5 min at 1,000 × *g*, +4 °C between washes. Beads were then washed for 10 min with 15 bead volumes of TEN 140 buffer at +25 °C and spun-down for 5 min at 1,000 × *g*, +25 °C. To elute the complexes (protein–DNA), 12–15 bead volumes of IgG elution buffer were added to the beads. The slurry was mixed by inverting for 1 h at +25 °C. This elution step was repeated one more time. To eliminate urea from the samples, the eluates were first concentrated in 10,000 Amicon Ultra-15 columns (3,000 × *g*, 15 min at +25 °C) and then washed/buffer-exchanged three times with

15 mL of TEN 140 buffer with 0.1% Triton X-100 (3,000 × *g*, 5–15 min at +25 °C) in a fresh Amicon column. The resulting 0.5 mL of concentrate was diluted with 2,500 μL of RIPA buffer and transferred into two Eppendorf tubes.

Biotin–streptavidin affinity purification: A total of 150–300 μL of streptavidin–agarose beads (Thermo Scientific, cat # 20349) were added to each tube. The mixture was rotated end-over-end for 12–16 h at +16 °C. Beads from both tubes were pooled into one 15-mL Falcon tube, washed once with RIPA buffer, once with TEN 140 buffer with 0.1% Triton ×100, twice with IgG elution buffer, and twice with IgG elution buffer without SDS. For each washing step, beads were mixed with 12 mL of a given buffer on a rotating wheel for 10 min at +25 °C and spun-down for 5 min at 1,000 × *g*, +25 °C. Beads were resuspended in 10 mL of TEN 140 buffer and divided into two tubes: 7 mL for protein work and 3 mL for DNA analysis.

On-Bead Trypsin Digestion, C18 Column Peptide Purification, and LC-MS/MS. A total of 7 mL of the bead suspension was spun-down for 5 min at 1,000 × *g* and washed seven times with 12 mL of 50 mM ammonium bicarbonate. (Note: To check the protein composition and quality of the pulldown, a small aliquot, about 1/20 of bead volume, can be taken for Western blot analysis. Before loading on the gel, the beads were incubated with 1.5 bead volumes of reverse cross-linking buffer for 25 min at +100 °C). The remaining beads were resuspended in 800 μL of 50 mM ammonium bicarbonate with 20 μL of sequencing-grade trypsin (Promega, catalog #V5111), and the mixture was inverted for 12 h at +37 °C on a rotating wheel. The reaction was stopped by adding 1 μL of 100% formic acid. The mixture was spun-down for 5 min at 1,000 × g, and supernatant was collected. Beads were washed three times with 75 μL of 25% acetonitrile with 0.1% formic acid, and all supernatants were combined in one tube. Samples were dried using a Speedvac to a final volume of 10 μL. A total of 50 μL of 0.1% trifluoroacetic acid was added to the samples, and 60 μL of the resulting mixture was loaded on the Pierce C18 Spin Tips (Thermo Scientific, catalog #84850) followed by the clean-up protocol provided by the manufacturer. Peptides were eluted from the column once with 25 μL of 50% acetonitrile and once with 25 μL of 100% acetonitrile. eluates were combined and dried using a Speedvac and submitted for LS-MS/MS.

Buffers for BioTAP-XL. The following buffers were used for BioTAP-XL: NEB buffer (10% sucrose, 20 mM Hepes, pH 7.6, 10 mM NaCl, 3 mM $MgCl_2$, 0.2% Triton); N-sucrose buffer (300 mM sucrose, 10 mM Hepes NaOH, pH 7.9, 1% Triton X-100, 2 mM MgOAc); N-glycerol buffer (25% glycerol, 10 mM Hepes NaOH, pH 7.9, 0.1 mM EGTA, 5 mM MgOAc); RIPA buffer (140 mM NaCl, 10 mM Tris·HCl, pH 8.0, 1 mM EDTA, pH 8.0, 1% Triton X-100, 0.1% SDS); TEN 140 buffer (10 mM Tris·HCl, pH 8.0, 1 mM EDTA, pH 8.0, 140 mM NaCl); 1× PBS, pH 7.4 (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.47 mM KH_2PO_4); TE buffer (10 mM Tris·HCl, pH 8.0, 1 mM EDTA, pH 8.0); IgG elution buffer (100 mM Tris·HCl, pH 8.0, 200 mM NaCl, 6 M urea, 0.2% SDS); and reverse cross-linking buffer (250 mM Tris·HCl, pH 8.8, 2% SDS, 0.5 M 2-mercaptoethanol).

Data Accessibility. DNA sequencing data for this article have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus public repository with the accession no. GSE53495. Peptide counts are provided in Datasets S1 and S2.

ACKNOWLEDGMENTS. We thank Ross Tomaino (Taplin Mass Spectrometry Facility, Harvard Medical School) for helpful advice regarding mass spectrometry; Stephen Elledge, Alberto Ciccia, and Natasha Pavlova for sharing reagents and advice regarding human tissue culture; and Peter Kaiser for reagents. We also thank Myles Brown, Sarah Elgin, Bob Kingston, and members of the M.I.K. laboratory for critically reading an earlier version of the manuscript. This work was supported by National Institutes of Health Grant K25AG037596 (to P.V.K.) and Grant GM101958 (to M.I.K.).

1. Lewis EB (1978) A gene complex controlling segmentation in Drosophila. *Nature* 276(5688):565–570.
2. Simon JA, Kingston RE (2013) Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol Cell* 49(5):808–824.
3. Margueron R, Reinberg D (2011) The Polycomb complex PRC2 and its mark in life. *Nature* 469(7330):343–349.
4. Gao Z, et al. (2012) PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol Cell* 45(3):344–356.
5. Rigaut G, et al. (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* 17(10):1030–1032.
6. Guerrero C, Tagwerker C, Kaiser P, Huang L (2006) An integrated mass spectrometry-based proteomic approach: Quantitative analysis of tandem affinity-purified in vivo cross-linked protein complexes (QTAX) to decipher the 26 S proteasome-interacting network. *Mol Cell Proteomics* 5(2):366–378.
7. Cao R, Zhang Y (2004) SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Mol Cell* 15(1):57–67.
8. Kim H, Kang K, Kim J (2009) AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. *Nucleic Acids Res* 37(9):2940–2950.
9. Nekrasov M, et al. (2007) Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. *EMBO J* 26(18):4078–4088.
10. Sarma K, Margueron R, Ivanov A, Pirrotta V, Reinberg D (2008) Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. *Mol Cell Biol* 28(8):2718–2731.
11. Herz HM, Shilatifard A (2010) The JARID2-PRC2 duality. *Genes Dev* 24(9):857–861.
12. Shen X, et al. (2008) EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol Cell* 32(4):491–502.
13. Margueron R, et al. (2008) Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol Cell* 32(4):503–518.
14. Zhang Z, et al. (2011) PRC2 complexes with JARID2, MTF2, and esPRC2p48 in ES cells to modulate ES cell pluripotency and somatic cell reprogramming. *Stem Cells* 29(2): 229–240.
15. De Cegli R, et al. (2013) Reverse engineering a mouse embryonic stem cell-specific transcriptional network reveals a new modulator of neuronal differentiation. *Nucleic Acids Res* 41(2):711–726.
16. Smits AH, Jansen PW, Poser I, Hyman AA, Vermeulen M (2013) Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Res* 41(1):e28.
17. Kuppuswamy M, et al. (2008) Role of the PLDLS-binding cleft region of CtBP1 in recruitment of core and auxiliary components of the corepressor complex. *Mol Cell Biol* 28(1):269–281.
18. Zhang Y, et al. (2011) Corepressor protein CDYL functions as a molecular bridge between polycomb repressor complex 2 and repressive chromatin mark trimethylated histone lysine 27. *J Biol Chem* 286(49):42414–42425.
19. Escamilla-Del-Arenal M, et al. (2013) Cdyl, a new partner of the inactive X chromosome and potential reader of H3K27me3 and H3K9me2. *Mol Cell Biol* 33(24):5005–5020.
20. Wu H, et al. (2011) Histone methyltransferase G9a contributes to H3K27 methylation in vivo. *Cell Res* 21(2):365–367.
21. Mozzetta C, et al. (2013) The histone H3 lysine 9 methyltransferases G9a and GLP regulate Polycomb Repressive Complex 2-mediated gene silencing. *Mol Cell*, 10.1016/j.molcel.2013.12.005.
22. Musselman CA, et al. (2012) Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. *Nat Struct Mol Biol* 19(12):1266–1272.
23. Walker E, et al. (2010) Polycomb-like 2 associates with PRC2 and regulates transcriptional networks during mouse embryonic stem cell self-renewal and differentiation. *Cell Stem Cell* 6(2):153–166.
24. Qin S, et al. (2013) Tudor domains of the PRC2 components PHF1 and PHF19 selectively bind to histone H3K36me3. *Biochem Biophys Res Commun* 430(2):547–553.
25. Maertens GN, El Messaoudi-Aubert S, Elderkin S, Hiom K, Peters G (2010) Ubiquitin-specific proteases 7 and 11 modulate Polycomb regulation of the INK4a tumour suppressor. *EMBO J* 29(15):2553–2565.
26. de Bie P, Zaaroor-Regev D, Ciechanover A (2010) Regulation of the Polycomb protein RING1B ubiquitination by USP7. *Biochem Biophys Res Commun* 400(3):389–395.
27. Wang CI, et al. (2013) Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in Drosophila. *Nat Struct Mol Biol* 20(2):202–209.
28. Miller JC, et al. (2011) A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* 29(2):143–148.
29. Jinek M, et al. (2013) RNA-programmed genome editing in human cells. *Elife* 2: e00471.
30. Kaneko S, et al. (2010) Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. *Genes Dev* 24(23):2615–2620.
31. de Boer E, et al. (2003) Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc Natl Acad Sci USA* 100(13):7480–7485.