

# A cellular chemical probe targeting the chromodomains of Polycomb repressive complex 1

Jacob I Stuckey<sup>1</sup>, Bradley M Dickson<sup>1</sup>, Nancy Cheng<sup>1</sup>, Yanli Liu<sup>2</sup>, Jacqueline L Norris<sup>1</sup>, Stephanie H Cholensky<sup>1</sup>, Wolfram Tempel<sup>2</sup>, Su Qin<sup>2</sup>, Katherine G Huber<sup>1</sup>, Cari Sagum<sup>3</sup>, Karynne Black<sup>3</sup>, Fengling Li<sup>2</sup>, Xi-Ping Huang<sup>4,5</sup>, Bryan L Roth<sup>4,5</sup>, Brandi M Baughman<sup>1</sup>, Guillermo Senisterra<sup>2</sup>, Samantha G Pattenden<sup>1</sup>, Masoud Vedadi<sup>2</sup>, Peter J Brown<sup>2</sup>, Mark T Bedford<sup>3</sup>, Jinrong Min<sup>2</sup>, Cheryl H Arrowsmith<sup>2</sup>, Lindsey I James<sup>1,\*</sup> & Stephen V Frye<sup>1,\*</sup>

**We report the design and characterization of UNC3866, a potent antagonist of the methyllysine (Kme) reading function of the Polycomb CBX and CDY families of chromodomains. Polycomb CBX proteins regulate gene expression by targeting Polycomb repressive complex 1 (PRC1) to sites of H3K27me3 via their chromodomains. UNC3866 binds the chromodomains of CBX4 and CBX7 most potently, with a  $K_d$  of ~100 nM for each, and is 6- to 18-fold selective as compared to seven other CBX and CDY chromodomains while being highly selective over >250 other protein targets. X-ray crystallography revealed that UNC3866's interactions with the CBX chromodomains closely mimic those of the methylated H3 tail. UNC4195, a biotinylated derivative of UNC3866, was used to demonstrate that UNC3866 engages intact PRC1 and that EED incorporation into PRC1 is isoform dependent in PC3 prostate cancer cells. Finally, UNC3866 inhibits PC3 cell proliferation, consistent with the known ability of CBX7 overexpression to confer a growth advantage, whereas UNC4219, a methylated negative control compound, has negligible effects.**

Methylation of lysine residues on histones critically contributes to chromatin regulation<sup>1,2</sup>. This post-translational modification (PTM) provides a docking site for effector proteins, termed Kme readers, that recognize this PTM through an aromatic cage that accommodates the hydrophobic lysine side chain and engages the methylated amine through cation- $\pi$ , electrostatic and van der Waals interactions<sup>1</sup>. Selectivity between different methylation states (mono-, di- and trimethylation) is conferred primarily by cage geometry and composition, while sequence specificity is achieved through interactions with the amino acids flanking the modified lysine<sup>1,3</sup>.

Trimethyllysine 27 on histone 3 (H3K27me3) is a transcriptionally repressive signal that is installed by the methyltransferases EZH1 and EZH2 of Polycomb repressive complex 2 (PRC2). The other two core components of PRC2 are SUZ12 and the H3K27me3 reader EED<sup>4</sup>. In the 'canonical' Polycomb-signaling model, H3K27me3 is deposited by PRC2 and is then recognized by the chromodomain-containing Kme reader component of the multiprotein repressor complex PRC1. The PRC1 chromodomains (CBX2, CBX4, CBX6, CBX7 and CBX8) incorporate into PRC1 in a mutually exclusive manner<sup>5</sup> and are thought to facilitate PRC1-mediated transcriptional repression by targeting the complex to H3K27me3, allowing the monoubiquitination of Lys119 on H2A (H2AK119ub) by the E3 ubiquitin ligase subunits of PRC1, RING1A, RING1B and BMI<sup>6</sup>. H2AK119ub is in turn associated with chromatin compaction, DNA methylation and repression of the underlying genes<sup>7,8</sup>.

Notwithstanding this simplistic model of sequential PRC action, it has become clear that a large number of diverse PRC1 complexes exist. However, the role of this diversity in the context of specific gene regulation events or pathology is not understood. Further highlighting

the complexity of PRC1, it was recently reported that EED, the H3K27me3 reader once believed to participate only in PRC2, is also a component of CBX-containing PRC1 complexes<sup>4</sup>, giving PRC1 two potential Kme readers for the same mark. The variable CBX component is another point of diversity that potentially affects the genomic targeting of PRC1. However, the ability of CBX proteins to target PRC1 to overlapping regions of the genome<sup>9</sup> suggests that these proteins may be partially functionally redundant. Therefore, modulation of chromatin regulation through PRC1 Kme reader antagonism may require ligands that target multiple CBX proteins.

Of the CBX chromodomains, CBX7 is the best studied and has been implicated as contributing to numerous cancer types<sup>9-19</sup>. Overexpression of CBX7 confers a proliferative advantage in prostate, gastric, leukemia and lymphoma cell lines that can be inhibited through knockdown of CBX7 (refs. 9,10,17,18), while CBX2, CBX4 and CBX8 have also been shown to enhance proliferation in various cancer types<sup>20-22</sup>. Interestingly, CBX7 also functions as a tumor suppressor in lung, pancreatic, colon and thyroid tissues<sup>11,14-16</sup>. In addition to CBX7, other developmental Polycomb factors have similarly been shown to have contradictory functions in oncogenesis, for example the H3K27 demethylase JMJD3 (ref. 23).

Modulation of H3K27me3 signaling via small molecules has recently garnered much pharmaceutical interest. Inhibitors of the PRC2 methyltransferase EZH2 are effective at targeting aberrant H3K27me3 levels in lymphomas with activating mutations and are currently in clinical trials<sup>24</sup>. Inhibitors of the H3K27me3 demethylases JMJD3 and UTX also have growth-suppressive properties in acute lymphoblastic leukemia<sup>25</sup> and robust antitumor activity in xenograft models of diffuse intrinsic pontine glioma<sup>26</sup>. Therefore, we hypothesized that antagonism of the H3K27me3 reading function of PRC1,

<sup>1</sup>Center for Integrative Chemical Biology and Drug Discovery, Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. <sup>2</sup>Structural Genomics Consortium, University of Toronto, Toronto, Ontario, Canada. <sup>3</sup>Department of Epigenetics and Molecular Carcinogenesis, University of Texas MD Anderson Cancer Center, Smithville, Texas, USA. <sup>4</sup>National Institute of Mental Health Psychoactive Drug Screening Program, University of North Carolina at Chapel Hill Medical School, Chapel Hill, North Carolina, USA. <sup>5</sup>Department of Pharmacology, University of North Carolina at Chapel Hill Medical School, Chapel Hill, North Carolina, USA.

\*e-mail: ingerman@email.unc.edu or svfrye@email.unc.edu

thought to occur immediately downstream of PRC2, may also have therapeutic potential. Although there have been other reports of ligands for PRC1 chromodomains<sup>27,28</sup>, no molecules have been reported that meet the stringent criteria<sup>29–31</sup> for classification as a high-quality, cellularly active chemical probe. Chemical probes that inhibit the Kme3-reading capacity of PRC1 will be valuable tools in explicating PRC1 biology, determining the therapeutic potential of CBX chromodomains, investigating potential pharmacologic synergy with inhibitors of EZH2 and of JMJD3 and UTX, and understanding the precise roles of CBX target genes in different cancer types. We report the design and characterization of UNC3866 as a potent and cellularly active antagonist of PRC1 chromodomains along with our initial investigation into the biological effects of antagonizing the H3K27me3 reading function of CBX chromodomains in PC3 prostate cancer cells.

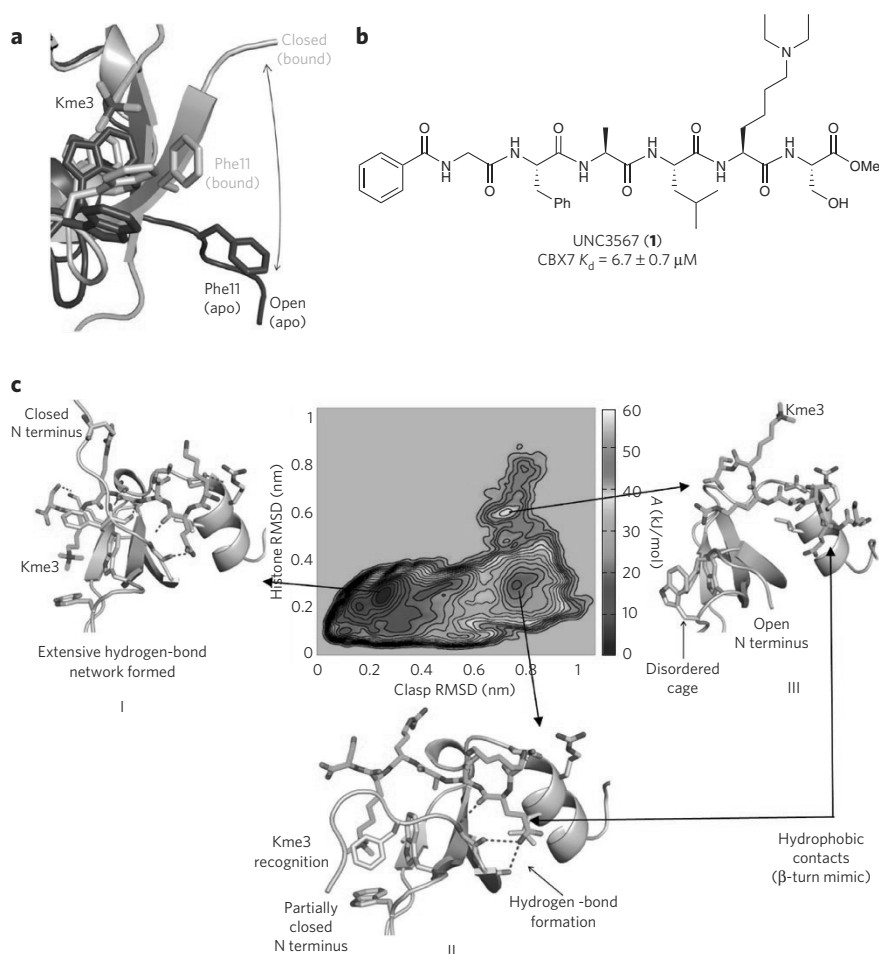
## RESULTS

### Molecular dynamics-based design of a CBX antagonist

Although the use of monospecific chemical probes is often desirable, probes with well-characterized multitarget profiles are also of great utility and may, pragmatically, be the best tools achievable within highly homologous target families. For example, the initially reported bromodomain antagonists were not selective within the BET family, and yet their impact has been extraordinary<sup>32</sup>. The high degree of similarity between the CBX chromodomains of PRC1, particularly in the regions that interact with H3, suggests that any molecule that binds one of them is likely to have affinity for other Polycomb chromodomains as well. With this in mind, we selected CBX7 as an initial target for ligand discovery because there is existing structural information<sup>3</sup>, but with an understanding that quantitative profiling versus other chromodomains would be required for full ligand characterization.

We first screened our epigenetic targeted library of small molecules designed largely for Kme reader domains and lysine methyltransferases using an established AlphaScreen competition assay<sup>33</sup>. Additionally, we screened a set of commercially available ligands known to interact with other quaternary ammonium-binding proteins at concentrations up to 30  $\mu\text{M}$  (see **Supplementary Results, Supplementary Fig. 1**, for compound names and structures). This approach proved unsuccessful in identifying tractable hits capable of inhibiting the CBX7-H3 peptide interaction. Therefore, we turned to the existing structural information to guide our hit-discovery efforts. We compared the apo CBX7 NMR structure (PDB 2K1B) with the H3K27me3-CBX7 NMR structure (PDB 2L12) and observed that the aromatic cage of CBX7 is unformed in the apo structure (**Fig. 1a**). The absence of a preformed aromatic cage makes the discovery of small-molecule Kme mimetics targeting this portion of CBX7 a significant challenge. Histone peptide substrates, in contrast, engage CBX7 via an induced-fit,  $\beta$ -sheet-like interaction that results in formation of the Kme-reading aromatic cage<sup>3</sup>.

Previous work using peptide arrays revealed that the overall CBX7 consensus binding sequence is A(R/I/L/F/Y/V)Kme3(S/T)<sup>3</sup>. In fact,



**Figure 1 | Molecular dynamics-guided inhibitor design.** (a) Comparison of the aromatic cage of CBX7 in the NMR structure of the apo (PDB 2K1B, dark gray) and H3K27me3-bound (PDB 2L1B, light gray, Kme3 shown in green) conformations of CBX7. (b) Structure of UNC3567. The  $K_d$  represents the average of three replicates  $\pm$  s.d. (c) Free energy plot from the CBX7-H3K9me3 MDS studies. Arrows are drawn from potential energy wells to representative structures associated with that well. Frames “I” to “III” represent one possible dissociative path of H3 from CBX7, and the reverse sequence represents the corresponding associative path of CBX7 and H3. The “N terminus” label refers to the N terminus of CBX7. A, free energy; RMSD, root mean square deviation.

the sequence RGFALKme3STHG binds CBX7 with 22-fold greater affinity than the corresponding H3K27me3 peptide ( $K_d$  of 5  $\mu\text{M}$  and 110  $\mu\text{M}$ , respectively)<sup>3</sup>. Therefore, we initiated optimization of peptide-based ligands with the goal of enhancing potency while decreasing charge and polarity so as to improve cell permeability, a known challenge for peptides. We first replaced the arginine preceding the Kme3 with leucine to facilitate cell permeability, and an N-terminal benzoyl cap was chosen to further increase the lipophilicity and aid in purification. Our next priority was to address the liabilities associated with Kme3: potential instability toward lysine demethylases<sup>34</sup> and limited passive membrane permeability due to the permanent positive charge. A variety of peptides were prepared and screened in which tertiary and secondary amines of varying size, hydrophobicity and basicity were incorporated as replacements for the quaternary amine. We were encouraged to find that replacing the Kme3 with a diethyllysine residue in this context resulted in a peptide, UNC3567 (1), that was approximately equipotent for CBX7 to the original consensus peptide (**Fig. 1b**;  $K_d = 6.7 \pm 0.7 \mu\text{M}$  as determined by isothermal titration calorimetry). This development represented encouraging progress toward our goal of cellularly active PRC1 antagonists and justified further optimization of UNC3567 to increase potency.

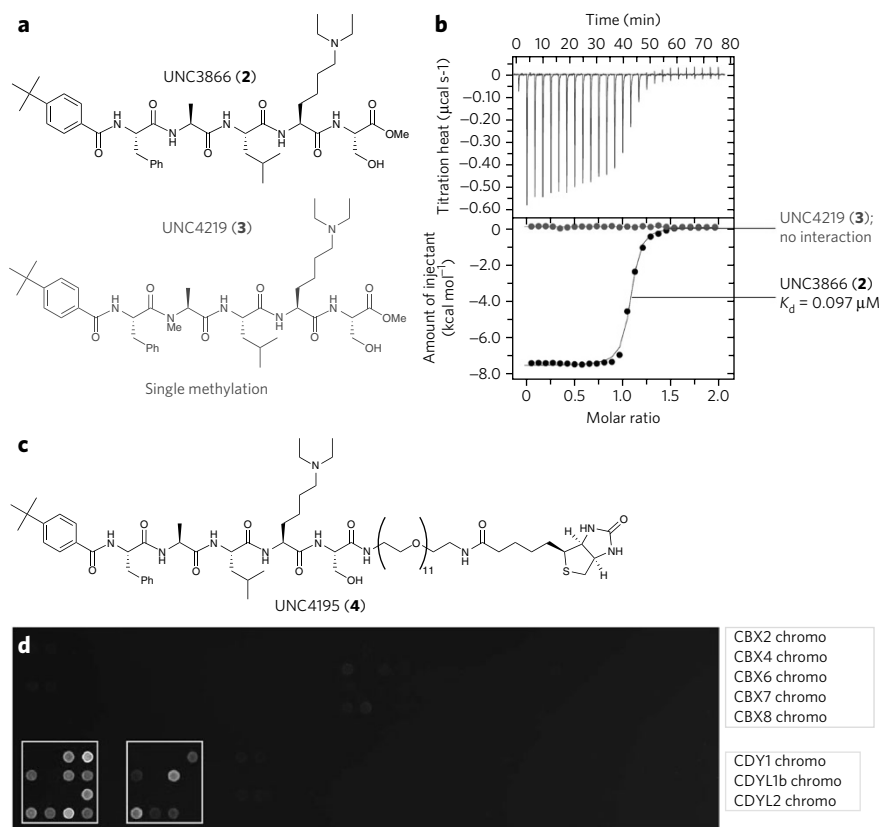
Because of the induced-fit recognition implied by prior structural studies<sup>3</sup>, we sought to better understand the critical CBX7-H3 interactions that drive the conformational rearrangement of CBX7, whereby residues 7–13 of CBX7 close around the histone and position Phe11 in close enough proximity to Trp32 and Trp35 to form the aromatic cage (Fig. 1a). We therefore turned to adaptively biased molecular dynamics simulations<sup>35</sup> (MDS) using the previously published NMR structure of H3K9me3 bound to CBX7 (PDB 2L12). We computed possible histone exit paths from the CBX7–H3K9me3 structure, and the pathway shown in Figure 1c was favored when we considered a truncated form of H3K9me3 (residues 4–10), which is most analogous to UNC3567. Starting from the bound state as shown in Figure 1c, frame I, the N terminus of CBX7 opens by losing partial contact with H3 (transition from frame I to frame II). Subsequently, the Kme3 loses contact with the aromatic cage (transition from frame II to III), leaving H3 engaged with CBX7 primarily through hydrophobic contacts made with V13, D50, R52 and L53 (frame III). Complete dissociation of H3 and CBX7 occurs following the loss of these contacts.

Association of CBX7 and H3 is simply the microscopic reversal of this exit path and is analogous to  $\beta$ -hairpin folding between two antiparallel  $\beta$ -strands. In this case, one  $\beta$ -strand is formed by the histone and the other by residues 8–13 of CBX7 (Supplementary Fig. 2). Therefore, establishing the appropriate contacts between the N-terminal portion of the H3 peptide (residue Q5; Fig. 1c, frame III) and CBX7 (namely residues V13, D50, R52 and L53) during association is analogous to forming the critical  $\beta$ -turn during hairpin folding. Once the turn is formed, backbone hydrogen bonds ‘zip’ the  $\beta$ -hairpin into the fully folded state<sup>36,37</sup>. In the case of CBX7, in addition to causing the formation of an extensive hydrogen bond network, this also orients Phe11 of CBX7 to contact Kme3 and facilitates formation of the aromatic cage. We used this pathway for H3 binding as a model for the association of CBX7 with UNC3567 and hypothesized that optimizing the N-terminal interactions of UNC3567 with CBX7 would facilitate the necessary hydrogen bonding for closure of CBX7 around UNC3567, thereby increasing ligand affinity.

We synthesized a series of replacements for the N-terminal benzoylglycine of UNC3567 that were less flexible and were predicted to interact favorably with V13, D50, R52 and L53 of CBX7. Structure-activity relationship (SAR) studies ultimately led to the synthesis of UNC3866 (2) (Fig. 2a), which is, to our knowledge, the most potent ligand reported for CBX7, with a  $K_d$  of  $97 \pm 2.4$  nM (Fig. 2b; see Supplementary Table 1 for key SAR data that led to UNC3866). We also generated a negative control compound, UNC4219 (3), for use in cellular studies through methylation of the alanine residue at the  $N_\alpha$  position (Fig. 2a). This modification disrupts a key hydrogen bond, and, consistent with our simulations, UNC4219 shows no interaction with CBX7 at concentrations up to 100  $\mu$ M (Fig. 2b).

### UNC3866 selectivity studies

Knowledge of the activity profile of a molecule is essential in order to associate its cellular effects with the modulation of a specific molecular



**Figure 2 | ITC studies of UNC3866 and UNC4219 with CBX7 and selectivity profiling of UNC3866.** (a) Structures of UNC3866 (black) and the structurally related negative control compound, UNC4219 (blue). (b) Comparison of UNC3866 ( $K_d = 97 \pm 2.4$  nM, average of three replicates  $\pm$  s.d.) and UNC4219 (no interaction up to 100  $\mu$ M) binding to CBX7 as measured by ITC. Top graph shows raw titration heat values ( $\mu$ cal/s) plotted against time after the integration baseline has been subtracted. Bottom graph shows normalized integration heat values (kcal/mol) of injectant plotted against molar ratio. (c) Structure of UNC4195, a biotinylated derivative of UNC3866. (d) UNC4195 selectively interacts with two subfamilies of chromodomains on a microarray containing 96 epigenetic reader proteins across multiple domain types.

target(s) of interest. Thus, the use and classification of a molecule as a chemical probe is dependent upon an in-depth understanding of its activity against a broad panel of potential molecular targets<sup>29</sup>. Accordingly, we evaluated the selectivity of UNC3866 against a wide range of both epigenetic and non-epigenetic targets. We first evaluated its selectivity against a small set of Kme readers encompassing proteins from the MBT, PHD and Tudor families by AlphaScreen or ITC (Supplementary Table 2). UNC3866 is a potent antagonist of the CBX7-H3 interaction, as determined by AlphaScreen ( $IC_{50} = 66 \pm 1.2$  nM), and is more than 100-fold selective for CBX7 over the other nine members of this Kme reader panel.

To facilitate further selectivity studies, we generated a biotinylated derivative of UNC3866, UNC4195 (4) (Fig. 2c), which retains high affinity for CBX7 ( $K_d = 220 \pm 22$  nM). We used this derivative to evaluate binding to 96 purified chromatin-associated effector proteins, including at least 19 known H3K9me3 or H3K27me3 binding proteins and 28 distinct chromodomains, spotted in duplicate onto a nitrocellulose-coated membrane (Fig. 2c, Supplementary Fig. 3a). Binding interactions were visualized with fluorescently tagged streptavidin, revealing that UNC4195 binds to two families of chromodomains. As expected, a positive binding interaction was observed with CBX7 and the other PRC1 orthologs, CBX2, CBX4, CBX6 and CBX8. We did not detect binding to CBX1, CBX3 or CBX5 (also known as HP1 $\beta$ , HP1- $\gamma$  and HP1- $\alpha$ , respectively), which are not associated with PRC1. The array also revealed binding to three

**Table 1 | Quantitative analysis of UNC3866 binding to CBX and CDY chromodomains by ITC**

Chromodomain	UNC3866 $K_d$ ( $\mu\text{M}$ )
CBX2	1.8 $\pm$ 0.21
CBX4	0.094 $\pm$ 0.017
CBX6	0.610 $\pm$ 0.0078
CBX7	0.097 $\pm$ 0.0024
CBX8	1.2 $\pm$ 0.021
CDY1	6.3 $\pm$ 0.92
CDYL1b	0.91 $\pm$ 0.076
CDYL2	0.85 $\pm$ 0.076

$K_d$  values are reported as the average of at least two replicates  $\pm$  s.d.

members of the CDY family of chromodomains: CDY1, CDYL1b and CDYL2. We expressed and purified all proteins that showed positive binding to UNC4195 on the array and subsequently quantified the affinity for UNC3866 by ITC (Table 1 and Supplementary Fig. 3b). In agreement with the microarray data, we detected binding to each of these chromodomains. We also found that the affinity of UNC3866 for CBX2, CBX4, CBX6 and CBX8 is surprisingly well correlated with the percent sequence identity of each chromodomain relative to that of CBX7 (Supplementary Fig. 3c). UNC3866 is equipotent for CBX4, which is most similar to CBX7, whereas it is 18-, 6- and 12-fold selective for CBX4 and CBX7 over CBX2, CBX6 and CBX8, respectively. Additionally, UNC3866 is 65-fold selective for CBX4 and CBX7 over CDY1 and 9-fold selective for CBX4 and CBX7 over CDYL1b and CDYL2. Although some work has been done to explore the biochemistry of CDY proteins<sup>38</sup>, their biological function is less well characterized than that of Polycomb CBX proteins, which makes assessing the potential consequences of antagonizing these proteins in a cellular context challenging. This is a ‘known unknown’ in the use of UNC3866.

UNC3866 was also evaluated against a panel of 48 bromodomains, 33 protein methyltransferases and 7 lysine demethylases and found to be inactive in each case (Supplementary Fig. 4 and Supplementary Tables 3 and 4). UNC3866 was tested against a general pharmacology panel (Cerep) consisting of 49 G protein-coupled receptors (GPCRs), 5 ion channels and 3 transporter proteins (Supplementary Table 5) as well as against the National Institute of Mental Health’s Psychoactive Drug Screening Program (NIMH PDSP) panel, which includes 22 additional GPCRs. Follow-up functional assays<sup>39,40</sup> for targets displaying >50% inhibition of control radioligand binding at 10  $\mu\text{M}$  UNC3866 were also performed with both UNC3866 and UNC4219 (negative control) (Supplementary Figs. 5–8). Both UNC3866 and UNC4219 showed nearly identical profiles in these follow-up assays, with both compounds having very weak antagonism for the NK2 and  $V_{1a}$  receptors (Supplementary Figs. 6 and 8a). The similar profiles of UNC3866 and UNC4219 indicate that UNC4219 will be a valuable negative control in determining the biological consequences specific to chromodomain antagonism.

### UNC3866 cocrystallization studies

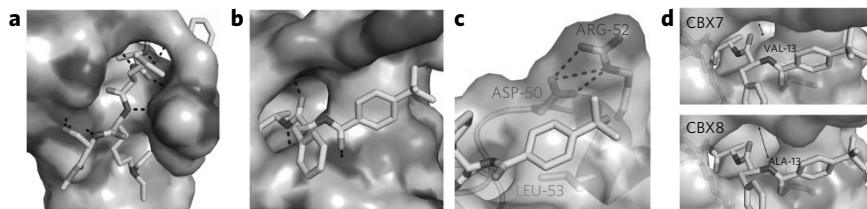
We additionally solved an X-ray cocrystal structure of UNC3866 bound to CBX7 (PDB 5EPJ) at a resolution of 1.6  $\text{\AA}$  to improve our understanding of its binding interactions. Each backbone amide of UNC3866 participates in at least one hydrogen bond with the backbone of CBX7 (Fig. 3a,b), while the N-terminal *tert*-butylbenzoyl cap of UNC3866 primarily contacts the side chains of D50, R52 and L53

(Fig. 3c). The guanidinium moiety of R52 forms a salt bridge with D50, and this rigidified R52 side chain forms a hydrophobic groove with L53 within which the *tert*-butylphenyl group of UNC3866 is embedded. On the basis of our MDS studies, we suspect that the 70-fold affinity increase observed upon substitution of the benzoylglycyl cap of UNC3567 for the *tert*-butylbenzoyl cap of UNC3866 is the result of an improvement in the ability of this moiety to nucleate closure of CBX7 around the molecule through interactions with D50, R52 and L53.

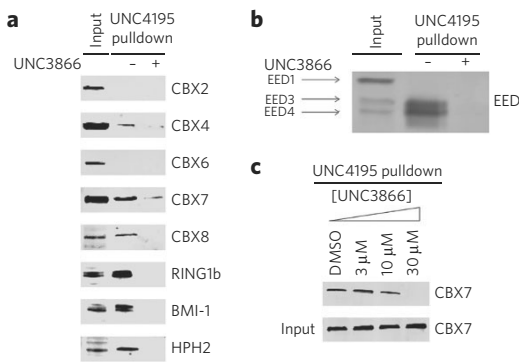
Additionally, we solved X-ray crystal structures of UNC3866 bound to CBX2 (PDB 5EPK), CBX4 (PDB 5EPL) and CBX8 (PDB code 5EQ0) in order to better understand the selectivity profile of UNC3866 (Fig. 3d and Supplementary Fig. 9). Alignment of these structures clearly illustrates that UNC3866 binds in a similar fashion to each chromodomain (Supplementary Fig. 9a). A closer inspection of the binding pocket for the alanine of UNC3866 in each protein revealed that the improved potency of UNC3866 for CBX4 and CBX7 over CBX2 and CBX8 might be the result of a subtle, yet significant, difference in this pocket. The alanine pocket in CBX2 and CBX8 consists of a valine, leucine and alanine, whose side chains form a relatively open hydrophobic pocket that binds the UNC3866 alanine methyl group. In contrast, in CBX4 and CBX7 this pocket is made up of two valines and a leucine (Fig. 3d, Supplementary Fig. 9b–f), which mold more tightly to the alanine, enhancing the van der Waals interactions with UNC3866. This single residue is likely to be the predominant factor that determines the higher affinity of UNC3866 for CBX4 and CBX7, as there are no other significant variations within the amino acids that interact directly with UNC3866.

In the absence of cocrystal structures, it is more challenging to explain the selectivity of UNC3866 for CBX4 and CBX7 over the CDY proteins. Alignment of the sequences (Supplementary Fig. 9f) of the CBX and CDY proteins that bind UNC4195 on the microarray (Fig. 3a) indicates that the key motif in the CBX family that engages the *tert*-butyl of UNC3866 (D-X-R) is replaced by an (N/H)-C-E motif in the CDY proteins. Given the demonstrated importance of this interaction in the binding of UNC3866 to CBX7, this sequence variation may contribute to the modest selectivity of UNC3866 for CBX4 and CBX7. Further structural and in-depth SAR studies are needed to obtain a more complete understanding to aid in the design of selective ligands.

We were also interested in the interactions between the Polycomb CBX chromodomains and the diethyllysine of UNC3866 revealed by our structures. Whereas the H3K27me3 side-chain has been modeled as adopting a fully extended, all-*anti* conformation<sup>3</sup>, the lysine side chain of UNC3866, when bound to CBX proteins in our crystal structures, is modeled as adopting a *gauche* conformation about the  $\text{C}\gamma\text{-C}\delta$  bond in Figure 3a. Although the corresponding electron density is comparatively weak for CBX7 (Supplementary Fig. 10a), we model the side chain in the *gauche* conformation based on the



**Figure 3 | Structural studies of UNC3866 with CBX7 and CBX8.** (a) Interaction of CBX7 (PDB 5EPJ; surface shown) with the C-terminal portion of UNC3866 (green; serine methyl ester only partially resolved by electron density). (b) Interaction of CBX7 with the N-terminal portion of UNC3866. (c) Interactions of the *tert*-butylbenzoyl cap of UNC3866 with the hydrophobic groove formed by CBX7 residues D50, R52 and L53. (d) Comparison of the UNC3866 alanine-binding pocket in CBX7 and CBX8 (PDB 5EQ0) (top and bottom panels, respectively).



**Figure 4 | UNC4195 pulldown studies in PC3 cells.** (a) UNC4195 selectively chemiprecipitates CBX4, CBX7 and CBX8 from PC3 cell lysates and chemiprecipitates additional PRC1 components. (b) UNC4195 chemiprecipitates EED3 and EED4 from PC3 cell lysates. (c) Incubation of intact PC3 cells with UNC3866 for 24 h inhibits pulldown of CBX7 with UNC4195.  $n = 3$  for all samples in **a-c**. Full-sized western blots are provided in **Supplementary Figure 19**.

higher-resolution CBX8-UNC3866 cocrystal structure (1.2 Å as compared to 1.6 Å for the CBX7-UNC3866 structure). Although the electron density of the CBX8-UNC3866 structure more strongly supports the *gauche* conformation of this side chain in the bound state (**Supplementary Fig. 10b**), the ambiguity in the position of these atoms may reflect that the side chain in UNC3866 remains somewhat flexible when bound to CBX7. Further, this apparent flexibility in the bound state may explain the ability of CBX7 and the other Polycomb chromodomains to accommodate the diethyl functional group, which is more sterically demanding than the endogenous Kme3 recognized by these domains.

### UNC4195 engages 'canonical' PRC1

Chemiprecipitation experiments in PC3 prostate cancer cell lysates were performed with UNC4195 in order to investigate its utility as a CBX pulldown reagent. PC3 cells were chosen for their relatively high abundance of each CBX Polycomb protein and their prior utilization to study CBX7 function<sup>10</sup>. Upon chemiprecipitation with UNC4195, we probed for the presence of CBX2, CBX4, CBX6, CBX7 and CBX8. We detected pulldown of CBX4, CBX7 and CBX8, which was inhibited by the presence of excess UNC3866 in the lysate (**Fig. 4a**), confirming that successful pulldown is dependent on the specific binding of the chromodomains to UNC4195. We failed to pull down CBX2 and CBX6 in these experiments, suggesting that, in contrast to CBX4, CBX7 and CBX8, CBX2 and CBX6 participate in PRC1 complexes in such a way that their chromodomains become inaccessible to pulldown with UNC4195. This is supported by the fact that the purified, isolated chromodomains of CBX2 and CBX6 interact with UNC4195 on the microarrays (**Fig. 2c**). We also investigated the ability of UNC4195 to chemiprecipitate PRC1 as an intact complex by probing for representative proteins of the remaining three 'canonical' PRC1 subunits (**Fig. 4b**). We detected RING1b, BMI-1 and HPH2, indicating that UNC4195, and by extension UNC3866, engages 'canonical' PRC1 without disrupting the complex's core composition.

### Isoform-specific incorporation of EED into PRC1

The recent finding that EED can participate in CBX-containing PRC1 complexes intrigued us<sup>4</sup>; therefore, we investigated the ability of UNC4195 to pull down EED in PC3 cells. Three of the four mammalian isoforms of EED (EED1, EED3 and EED4)<sup>41</sup> were detected in the PC3 lysate input; however, UNC4195 pulled down only EED3 and EED4, both of which could be competed away through the addition of UNC3866 to the lysate (**Fig. 4b**). Additionally, because EED binds

H3K27me3 via its WD40 Kme reading domain, we confirmed that UNC3866 does not bind directly to EED (**Supplementary Fig. 11**), indicating that the pulldown is mediated through incorporation of EED into PRC1. We next investigated the presence of EED isoforms in PRC2 using the EZH1 and EZH2 biotinylated pulldown reagent UNC2399 (ref. 24). Interestingly, we detected only EED4 (**Supplementary Fig. 12**). It has been shown that the incorporation of different isoforms of EED into PRC2 can dictate the *in vitro* histone substrate of PRC2 (ref. 42), and hence our discovery that EED incorporates into PRC1 in an isoform-dependent fashion provides the basis for future work to examine the effects of these isoforms on PRC1 function. Furthermore, the absence of EED1 in both PRC1 and PRC2 raises questions about the binding partners and function of this isoform in PC3 cells.

### UNC3866 antagonizes PRC1 chromodomains in cells

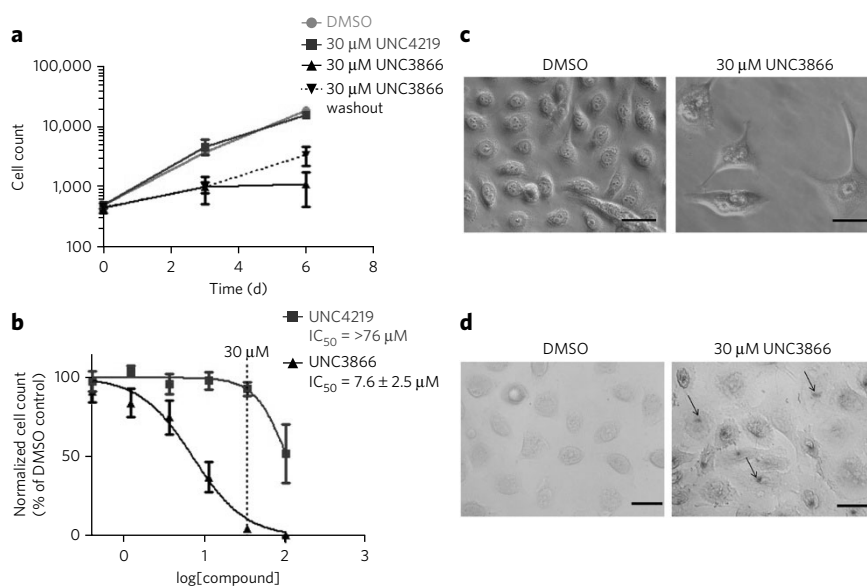
We next assessed the utility of UNC3866 in a cellular context. As peptide inhibitors often have poor cellular activity, we evaluated the permeability of UNC3866 using a Caco-2 assay and found it to be low (**Supplementary Fig. 13**), indicating a need for relatively high compound concentrations for cellular studies relative to the *in vitro*  $K_d$ . Additionally, we noted the potential for UNC3866 to serve as a prodrug if the C-terminal methyl ester is susceptible to hydrolysis by intracellular esterases; therefore, we confirmed through a combination of pulldown and ITC experiments that the corresponding C-terminal acid, UNC4007 (5), is still capable of engaging PRC1 chromodomains (**Supplementary Fig. 14 and 20**). Next, we quantified the intracellular concentration of both UNC3866 and UNC4007 by LC-MS/MS (**Supplementary Table 6**). Owing to the low Caco-2 permeability, we incubated PC3 cells with 30 μM UNC3866 to ensure sufficient accumulation of intracellular UNC3866. Treatment for 24 h resulted in an intracellular concentration of  $1.4 \pm 0.3$  μM, indicating that the intracellular concentration of UNC3866 is ~5% of the extracellular concentration. The UNC3866 intracellular concentration was approximately seven-fold higher than the intracellular concentration of UNC4007, indicating that hydrolysis of UNC3866 is limited under these conditions. Overall, UNC3866 is sufficiently cell permeable and stable that it is possible to evaluate its ability to engage and antagonize the functions of its chromodomain targets in cells.

The nonenzymatic nature of chromodomains makes demonstrating cellular engagement of these proteins challenging. Although PRC1 has E3 ligase activity, the presence of noncanonical PRC1 complexes that do not contain CBX proteins<sup>6</sup> makes monoubiquitination of H2AK119 an unreliable readout. Further, neither mutation nor deletion of the chromodomains of CBX proteins has significant effects on the cellular mobility of these proteins as determined by fluorescence recovery after photobleaching (FRAP) experiments<sup>43</sup>, ruling out this technique, which has been previously applied to chromatin readers<sup>32,33</sup>. We failed to detect thermostabilization of CBX7 by UNC3866 in cell lysates (**Supplementary Fig. 15**), likely because the chromodomain comprises a relatively small portion (~25%) of the full-length protein, which precludes evaluation of cellular target engagement via the cellular thermal shift assay (CETSA)<sup>44</sup>. Therefore, we hypothesized that competition-based chemiprecipitation experiments could be used to demonstrate cellular target engagement, with the expectation that pretreatment with UNC3866 would block chemiprecipitation. Whole PC3 cells were incubated with varying concentrations of UNC3866 for 24 h, after which excess medium and compound were removed and the cells were washed, harvested with trypsin and lysed. Encouragingly, CBX7 chemiprecipitation with UNC4195 was completely inhibited in lysates from cells pretreated with 30 μM UNC3866 (intracellular concentration =  $1.4 \pm 0.3$  μM, **Fig. 4c**), indicating that UNC3866 effectively engages CBX7 in cells at this concentration.

We next focused on establishing a biological consequence of treating cells with UNC3866. The high levels of CBX proteins in PC3 cells, coupled with the demonstrated ability of CBX7 overexpression to confer a growth advantage in these cells<sup>10,45</sup>, encouraged us to investigate the effects of UNC3866 on PC3 cell proliferation. As a prelude to these studies, we confirmed that treatment with UNC3866 up to 100  $\mu$ M did not induce any toxic effects, as assessed by CellTiter-Glo in PC3 and HEK293T cells (**Supplementary Fig. 16**). Similarly, UNC4219 had no observable toxicity within the same concentration range. HEK293T cells were used to evaluate UNC3866 toxicity in a noncancerous cell line. To assess effects on proliferation, PC3 cells were treated with UNC3866 and the negative control, UNC4219, at 30  $\mu$ M, as this concentration was shown to effectively engage CBX7 in cells (**Fig. 4c**). Consistent with the known ability of CBX7 to enhance proliferative capacity in PC3 cells<sup>14,45</sup>, we observed a dramatic decrease in cell proliferation upon UNC3866 treatment, resulting in almost no change in cell count after 3 and 6 d, while cells treated with UNC4219 or DMSO exhibited exponential cell growth (**Fig. 5a**). Further, we observed that cells resumed proliferating when washed and not replenished with UNC3866 after 3 d (**Fig. 5a**, dotted line), indicating that the compound's effects are at least partially reversible. We next examined dose dependency for UNC3866 and UNC4219, and we observed a significant reduction in PC3 cell count only for UNC3866-treated cells, resulting in an  $EC_{50}$  of  $7.6 \pm 2.5$   $\mu$ M (**Fig. 5b**). The intracellular concentration of UNC3866 at its  $EC_{50}$  in PC3 cells is therefore estimated to be 340 nM, which is adequate to fully engage CBX4 and CBX7 given the measured  $\sim 100$  nM  $K_d$  values against these targets (**Fig. 3b**). Although some contribution to this antiproliferative phenotype from other chromodomains bound by UNC3866 cannot be ruled out, the concordance of this result with previously published data from CBX7 genetic manipulation studies is striking<sup>17,18</sup>.

We also observed that PC3 cells treated with UNC3866 exhibited an enlarged and flattened morphology (**Fig. 5c**), suggesting that they had become senescent<sup>46</sup>. We checked for the presence of senescence-associated  $\beta$ -galactosidase (SA- $\beta$ -gal) and observed a dramatic increase in the number of cells positive for SA- $\beta$ -gal following treatment with 30  $\mu$ M UNC3866 (**Fig. 5d**). This was an intriguing observation because CBX7 was originally identified in a genetic screen for genes capable of bypassing replicative senescence in human prostate epithelial cells (HPrEC)<sup>18</sup>. Furthermore, the same study observed that CBX7 knockdown is capable of inducing senescence.

Although it has been suggested that CBX7 controls cellular proliferation through regulation of the *Ink4a/ARF* (*CDKN2A*) locus<sup>18</sup>, its actual role in regulating this locus is controversial<sup>9,11,18</sup>. For example, CBX7 has been shown to enhance proliferative capacity without negatively regulating *Ink4a/ARF*<sup>9</sup>, and negative regulation of this locus does not necessarily enhance proliferative capacity<sup>14</sup>. Nonetheless, we analyzed the expression of *Ink4a/ARF* in response to knockdown of CBX7 by shRNA (>60% mRNA reduction) in PC3 cells and did not observe a statistically significant change in expression. Further, evaluation of the effects of UNC3866 on expression of



**Figure 5 | Cellular effects of UNC3866.** (a) Treatment of PC3 cells with 30  $\mu$ M UNC3866 inhibits cell proliferation. UNC4219 treatment has no effect on PC3 cell proliferation. For washout experiments (dotted line), UNC3866 was not replenished at day 3. Error bars represent the s.d. of each data point ( $n = 3$ , single biological replicate with three technical replicates). The data are representative of at least four biological replicates with varying plating densities. (b) UNC3866 inhibits PC3 cell proliferation in a dose-dependent manner after exposure for 6 d ( $n = 8$ , two biological replicates with three technical replicates and one biological replicate with two technical replicates). The  $EC_{50}$  is reported as the 95% confidence interval. UNC4219 was used to control for off-target effects, which were not evident at concentrations below 100  $\mu$ M. The  $EC_{50}$  for UNC4219 is  $>76$   $\mu$ M, which is the lower bound of the 95% confidence interval ( $n = 6$ , three biological replicates with two technical replicates each). Error bars represent the s.e.m. of each data point. (c,d) UNC3866 induces a senescent-like morphology in PC3 cells (c) and expression of SA- $\beta$ -gal (d). Arrows indicate examples of positive SA- $\beta$ -gal expression. Cell morphology was assessed using phase-contrast microscopy (40 $\times$ ). SA- $\beta$ -gal expression was assessed using bright-field microscopy (40 $\times$ ). The images are representative of three biological replicates, and the black scale bars are equal to 60  $\mu$ m.

this locus showed no change at 30  $\mu$ M, ruling out *Ink4a/ARF* regulation in the antiproliferative effects of UNC3866 (**Supplementary Fig. 17**). This was not entirely surprising as there is considerable evidence indicating that this locus is deeply repressed in PC3 cells through DNA hypermethylation<sup>47–50</sup>. Further, methylation of *Ink4a/ARF* is unaffected by both DNMT knockdown and inhibition with 5-AzaC, suggesting that this gene may be irreversibly repressed in PC3 cells<sup>47–49</sup>.

Finally, we evaluated the stability of UNC3866 following intraperitoneal injection (10 mg/kg) in male Swiss albino mice and found that UNC3866 is the predominant species in plasma at all time points tested, relative to UNC4007 (**Supplementary Tables 7–9** and **Supplementary Fig. 18**), and shows 25% bioavailability and moderate clearance. Although these pharmacokinetic results are promising for a peptidic compound, the use of UNC3866 *in vivo* may be limited by the high circulating levels required for intracellular target engagement because of its poor cell permeability. The potential utility of UNC3866 at higher doses for *in vivo* experiments is currently under investigation.

## DISCUSSION

Our investigations into the molecular dynamics of H3 recognition by CBX7 enabled the design of UNC3866, the most potent CBX7 antagonist reported to date<sup>27,28</sup>. UNC3866 is an equipotent nanomolar inhibitor of CBX4 and CBX7 Kme recognition *in vitro*, with a thoroughly characterized selectivity profile. Further, through the use of UNC4195, we demonstrated that UNC3866 binds in

the context of intact, 'canonical' PRC1. Finally, UNC3866 exhibits micromolar cellular potency in competition pulldown and cellular proliferation assays with no associated toxicity. The replacement of the quaternary amine of the native peptide ligand with an unnatural tertiary amine mimetic was a key achievement in the development of a cellularly active ligand, and, more broadly, this signifies that a quaternary amine is not required for inhibition of Kme3 readers.

Knowledge of the activity profile of a molecule is essential in order to associate its cellular effects with modulation of a specific molecular target(s) of interest. Although UNC3866 is most potent for CBX4 and CBX7, it possesses affinity for other Polycomb CBX proteins and the CDY family of chromodomains. Its use as a chemical probe will require consideration of these activities and evaluation of dose dependency to assess possible effects from these other targets. This is not unusual for first-in-class chemical probes: for example, the initially reported bromodomain antagonists were not selective within the BET family and yet their utility is well established<sup>32</sup>. The thorough characterization of the *in vitro* selectivity of UNC3866 versus >250 proteins, the demonstrated cellular target engagement of UNC3866, and the use of UNC4219 as a negative control validate UNC3866 as the first high-quality chemical probe for both the Polycomb CBX family of Kme readers and, to our knowledge, any Kme3 reader.

Our work toward the development of chemical tools for Polycomb CBX Kme readers has revealed that EED is incorporated into PRC1 and PRC2 in an isoform-dependent manner in PC3 cells, providing the basis for future investigation into the biochemical role of the different EED isoforms in Polycomb-mediated signaling. The ability of UNC3866 to inhibit PC3 cell proliferation supports further investigation of the use of PRC1 chromodomain antagonists as therapeutics in oncology, alone or in combination with EZH2 inhibitors. In addition to their role in different cancers, PRC1 chromodomains play a pivotal role in regulating cellular differentiation<sup>9</sup>, and UNC3866 and its related derivatives provide a new set of tools that will assist in further probing the role of PRC1 chromodomains in cell fate determination and, more generally, in understanding Polycomb signaling.

Received 1 May 2015; accepted 25 November 2015;  
published online 25 January 2016

## METHODS

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

**Accession codes.** Protein Data Bank: CBX2 with UNC3866 bound, 5EPK; CBX4 with UNC3866 bound, 5EPL; CBX7 with UNC3866 bound, 5EPJ; CBX8 with UNC3866 bound, 5EQ0.

## References

- Wagner, T., Robaa, D., Sippl, W. & Jung, M. Mind the methyl: methyllysine binding proteins in epigenetic regulation. *ChemMedChem* **9**, 466–483 (2014).
- Arrowsmith, C.H., Bountra, C., Fish, P.V., Lee, K. & Schapira, M. Epigenetic protein families: a new frontier for drug discovery. *Nat. Rev. Drug Discov.* **11**, 384–400 (2012).
- Kaustov, L. *et al.* Recognition and specificity determinants of the human cbx chromodomains. *J. Biol. Chem.* **286**, 521–529 (2011).
- Cao, Q. *et al.* The central role of EED in the orchestration of polycomb group complexes. *Nat. Commun.* **5**, 3127 (2014).
- Vandamme, J., Volkel, P., Rosnoble, C., Le Faou, P. & Angrand, P.O. Interaction proteomics analysis of polycomb proteins defines distinct PRC1 complexes in mammalian cells. *Mol. Cell Proteomics* **10**, M110.002642 (2011).
- Morey, L., Aloia, L., Cozzuto, L., Benitah, S.A. & Di Croce, L. RYBP and Cbx7 define specific biological functions of polycomb complexes in mouse embryonic stem cells. *Cell Reports* **3**, 60–69 (2013).
- Jin, B. *et al.* Linking DNA methyltransferases to epigenetic marks and nucleosome structure genome-wide in human tumor cells. *Cell Rep.* **2**, 1411–1424 (2012).
- Eskeland, R. *et al.* Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol. Cell* **38**, 452–464 (2010).
- Klauke, K. *et al.* Polycomb Cbx family members mediate the balance between haematopoietic stem cell self-renewal and differentiation. *Nat. Cell Biol.* **15**, 353–362 (2013).
- Bernard, D. *et al.* CBX7 controls the growth of normal and tumor-derived prostate cells by repressing the Ink4a/Arf locus. *Oncogene* **24**, 5543–5551 (2005).
- Forzati, F. *et al.* CBX7 is a tumor suppressor in mice and humans. *J. Clin. Invest.* **122**, 612–623 (2012).
- Shinjo, K. *et al.* Expression of chromobox homolog 7 (CBX7) is associated with poor prognosis in ovarian clear cell adenocarcinoma via TRAIL-induced apoptotic pathway regulation. *Int. J. Cancer* **135**, 308–318 (2014).
- Mansueti, G. *et al.* Identification of a new pathway for tumor progression: microRNA-181b up-regulation and CBX7 down-regulation by HMGA1 protein. *Genes Cancer* **1**, 210–224 (2010).
- Pallante, P. *et al.* Loss of the CBX7 gene expression correlates with a highly malignant phenotype in thyroid cancer. *Cancer Res.* **68**, 6770–6778 (2008).
- Pallante, P. *et al.* The loss of the CBX7 gene expression represents an adverse prognostic marker for survival of colon carcinoma patients. *Eur. J. Cancer* **46**, 2304–2313 (2010).
- Karamitopoulou, E. *et al.* Loss of the CBX7 protein expression correlates with a more aggressive phenotype in pancreatic cancer. *Eur. J. Cancer* **46**, 1438–1444 (2010).
- Zhang, X.W. *et al.* Oncogenic role of the chromobox protein CBX7 in gastric cancer. *J. Exp. Clin. Cancer Res.* **29**, 114 (2010).
- Gil, J., Bernard, D., Martínez, D. & Beach, D. Polycomb CBX7 has a unifying role in cellular lifespan. *Nat. Cell Biol.* **6**, 67–72 (2004).
- Scott, C.L. *et al.* Role of the chromobox protein CBX7 in lymphomagenesis. *Proc. Natl. Acad. Sci. USA* **104**, 5389–5394 (2007).
- Tan, J. *et al.* CBX8, a polycomb group protein, is essential for MLL-AF9-induced leukemogenesis. *Cancer Cell* **20**, 563–575 (2011).
- Wang, B. *et al.* Chromobox homolog 4 is correlated with prognosis and tumor cell growth in hepatocellular carcinoma. *Ann. Surg. Oncol.* **20** (suppl. 3): S684–S692 (2013).
- Clermont, P.L. *et al.* Genotranscriptomic meta-analysis of the Polycomb gene CBX2 in human cancers: initial evidence of an oncogenic role. *Br. J. Cancer* **111**, 1663–1672 (2014).
- Sauvageau, M. & Sauvageau, G. Polycomb group genes: keeping stem cell activity in balance. *PLoS Biol.* **6**, e113 (2008).
- Konze, K.D. *et al.* An orally bioavailable chemical probe of the Lysine Methyltransferases EZH2 and EZH1. *ACS Chem. Biol.* **8**, 1324–1334 (2013).
- Ntziachristos, P. *et al.* Contrasting roles of histone 3 lysine 27 demethylases in acute lymphoblastic leukaemia. *Nature* **514**, 513–517 (2014).
- Hashizume, R. *et al.* Pharmacologic inhibition of histone demethylation as a therapy for pediatric brainstem glioma. *Nat. Med.* **20**, 1394–1396 (2014).
- Simhadri, C. *et al.* Chromodomain antagonists that target the polycomb-group methyllysine reader protein chromobox homolog 7 (CBX7). *J. Med. Chem.* **57**, 2874–2883 (2014).
- Ren, C. *et al.* Small-molecule modulators of methyl-lysine binding for the CBX7 chromodomain. *Chem. Biol.* **22**, 161–168 (2015).
- Frye, S.V. The art of the chemical probe. *Nat. Chem. Biol.* **6**, 159–161 (2010).
- Bunnage, M.E., Chekler, E.L. & Jones, L.H. Target validation using chemical probes. *Nat. Chem. Biol.* **9**, 195–199 (2013).
- Workman, P. & Collins, I. Probing the probes: fitness factors for small molecule tools. *Chem. Biol.* **17**, 561–577 (2010).
- Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073 (2010).
- James, L.I. *et al.* Discovery of a chemical probe for the L3MBTL3 methyllysine reader domain. *Nat. Chem. Biol.* **9**, 184–191 (2013).
- Hopkinson, R.J. *et al.* Is JmjC oxygenase catalysis limited to demethylation? *Angew. Chem. Int. Edn. Engl.* **52**, 7709–7713 (2013).
- Dickson, B.M. Approaching a parameter-free metadynamics. *Phys. Rev. E* **84**, 037701 (2011).
- Muñoz, V., Thompson, P.A., Hofrichter, J. & Eaton, W.A. Folding dynamics and mechanism of  $\beta$ -hairpin formation. *Nature* **390**, 196–199 (1997).
- Xiao, Y., Chen, C. & He, Y. Folding mechanism of  $\beta$ -hairpin trpzip2: heterogeneity, transition state and folding pathways. *Int. J. Mol. Sci.* **10**, 2838–2848 (2009).
- Fischle, W., Franz, H., Jacobs, S.A., Allis, C.D. & Khorasanizadeh, S. Specificity of the chromodomain Y chromosome family of chromodomains for lysine-methylated ARK(S/T) motifs. *J. Biol. Chem.* **283**, 19626–19635 (2008).
- Besnard, J. *et al.* Automated design of ligands to polypharmacological profiles. *Nature* **492**, 215–220 (2012).
- Kroeze, W.K. *et al.* PRESTO-Tango as an open-source resource for interrogation of the druggable human GPCRome. *Nat. Struct. Mol. Biol.* **22**, 362–369 (2015).

41. Montgomery, N.D., Yee, D., Montgomery, S.A. & Magnuson, T. Molecular and functional mapping of EED motifs required for PRC2-dependent histone methylation. *J. Mol. Biol.* **374**, 1145–1157 (2007).
42. Kuzmichev, A., Jenuwein, T., Tempst, P. & Reinberg, D. Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. *Mol. Cell* **14**, 183–193 (2004).
43. Ren, X., Vincenz, C. & Kerppola, T.K. Changes in the distributions and dynamics of polycomb repressive complexes during embryonic stem cell differentiation. *Mol. Cell Biol.* **28**, 2884–2895 (2008).
44. Martinez Molina, D. *et al.* Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* **341**, 84–87 (2013).
45. Li, Q. *et al.* Polycomb CBX7 directly controls trimethylation of histone H3 at lysine 9 at the p16 locus. *PLoS One* **5**, e13732 (2010).
46. Kong, Y., Cui, H., Ramkumar, C. & Zhang, H. Regulation of senescence in cancer and aging. *J. Aging Res.* **2011**, 963172 (2011).
47. Jarrard, D.F. *et al.* Deletional, mutational, and methylation analyses of CDKN2 (p16/MTS1) in primary and metastatic prostate cancer. *Genes Chromosom. Cancer* **19**, 90–96 (1997).
48. Yaqinuddin, A., Qureshi, S.A., Qazi, R. & Abbas, F. Down-regulation of DNMT3b in PC3 cells effects locus-specific DNA methylation, and represses cellular growth and migration. *Cancer Cell Int.* **8**, 13 (2008).
49. Yaqinuddin, A., Qureshi, S.A., Qazi, R., Farooq, S. & Abbas, F. DNMT1 silencing affects locus specific DNA methylation and increases prostate cancer derived PC3 cell invasiveness. *J. Urol.* **182**, 756–761 (2009).
50. Mirochnik, Y. *et al.* Androgen receptor drives cellular senescence. *PLoS One* **7**, e31052 (2012).

## Acknowledgments

We thank K. Pearce (UNC Chapel Hill) for useful discussions and careful reading of the manuscript, O. Fedorov (SGC Oxford) for support with the bromodomain selectivity screening, A. Tumber (SGC Oxford) for support with the lysine demethylase selectivity screening and J.R. Walker for the review of the crystal structures. We thank G. Wang (UNC) for providing PHF1, PHF19, PHF23 and JARID1A protein constructs. We thank the labs of T. Magnuson (UNC) and D. Margolis (UNC) for providing EED and BMI-1 antibodies, respectively. We also thank N. Sciaky (UNC) for help with cell counting using the high content imaging microscope. Results shown in this report are derived from work performed at Argonne National Laboratory, Structural Biology Center at the

Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. The research described here was supported by the National Institute of General Medical Sciences, US National Institutes of Health (NIH, grant R01GM100919), the Carolina Partnership and the University Cancer Research Fund, University of North Carolina at Chapel Hill, and the Welch Foundation (G-1847). The SGC is a registered charity (no. 1097737) that receives funds from AbbVie, Boehringer Ingelheim, the Canada Foundation for Innovation (CFI), the Canadian Institutes of Health Research (CIHR), Genome Canada, Ontario Genomics Institute Grant OGI-055, GlaxoSmithKline, Janssen, Lilly Canada, the Novartis Research Foundation, the Ontario Ministry of Economic Development and Innovation, Pfizer, Takeda and Wellcome Trust Grant 092809/Z/10/Z. M.T.B. directs the Protein Array Core, which is supported by the Cancer Prevention Research Institute of Texas (RP130432) and by the Centre for Environmental and Molecular Carcinogenesis at the MD Anderson Cancer Center.

## Author contributions

J.I.S. designed and synthesized all compounds and related analogs, performed ITC studies and pulldown studies, rendered structural images, assisted with AlphaScreen studies and performed all cell culture; B.M.D. performed adaptively biased molecular dynamics studies and assisted in compound design; N.C. performed CellTiter-Glo assays and cellular proliferation assays; Y.L., W.T. and S.Q. solved the X-ray crystal structures of UNC3866 with the various CBX proteins; J.L.N. and S.H.C. expressed purified proteins; K.G.H. performed pulldown studies; C.S. and K.B. performed protein array experiments; E.L. performed methyltransferase assays; X.-P.H. performed GPCR functional assays. B.M.B. performed and analyzed AlphaScreen assays; G.S. performed EED ITC experiments; J.I.S., B.D.M., S.G.P., J.M., B.L.R., M.V., P.J.B., M.T.B., C.H.A., L.I.J. and S.V.F. designed studies and discussed results; J.I.S., L.I.J. and S.V.F. wrote the paper.

## Competing financial interests

The authors declare no competing financial interests.

## Additional Information

Any supplementary information, chemical compound information and source data are available in the online version of the paper. Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Correspondence and requests for materials should be addressed to L.I.J. and S.V.F.



## ONLINE METHODS

**Experimental methods.** *Adaptively biased molecular dynamics simulations.* Adaptive biasing potential<sup>55</sup> was implemented in GROMACS 4.5.5 (ref. 51). The bias parameters were  $b = 0.8$ ,  $c = 0.0005/\delta t$ , where  $\delta t = 2$  fs is the molecular dynamics time step. The Gaussian width was  $1/4$  Å. This choice of bias parameters floods stable states extremely slowly, to minimize adaptation of the bias near transition states. Thus, one may argue for approximate state-to-state dynamics as has recently been done for metadynamics<sup>52</sup> following the principles of hyperdynamics<sup>53</sup>. Using the first frame of the CBX7+H3K9me3 NMR structure, two collective variables were defined: r.m.s. deviation of clasp, and r.m.s. deviation of peptide. We used the Kabsch algorithm to align the reference and trajectory during the simulation. All simulations were performed in a cubic box with a minimum of 12 Å between the protein or peptide and the nearest cube face, resulting in 23,710 atoms after adding water and salt. System construction included 5 ns of isothermal-isobaric equilibration at 1 bar. Production simulations were carried out in the canonical ensemble with stochastic velocity rescaling<sup>54</sup>. Van der Waals and direct electrostatic interactions used a 10 Å cutoff. Long-range electrostatics were treated with the particle mesh Ewald approach with a grid spacing of 1.6 Å. The neighbor list was updated every five steps. We ran five 128 ns simulations to get some intuition that could inform ligand design. All images from the simulations were rendered using the PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.

*ITC experiments.* All ITC measurements were recorded at 25 °C with an AutoITC200 microcalorimeter (MicroCal Inc., MA). All protein and compound stock samples were in the target buffer (25 mM Tris-HCl, pH 8, 150 mM NaCl, and 2 mM  $\beta$ -mercaptoethanol) and then diluted in the same buffer to achieve the desired concentrations: 50  $\mu$ M protein and 0.5 mM compound. The concentration of the protein stock solution was established using the Edelhoch method, whereas compound stock solutions were prepared based on mass. A typical experiment included a single 0.2  $\mu$ l compound injection into a 200  $\mu$ l cell filled with protein, followed by 26 subsequent 1.5  $\mu$ l injections of compound. Injections were performed with a spacing of 180 s and a reference power of 8  $\mu$ cal/s. Control experiments were performed titrating each compound into buffer under identical conditions to determine the heat signals, if any, that arise from diluting the compound. If applicable, the heats of dilution generated were then subtracted from the protein-compound binding curves. The initial data point was routinely deleted. The titration data was analyzed using Origin Software (MicroCal Inc., USA) by nonlinear least-squares, fitting the heats of binding as a function of the compound:protein ratio to a one site binding model.

*Kme reader AlphaScreen assays.* The AlphaScreen assay (Perkin Elmer) was generally performed as previously described<sup>55</sup>. In brief, compound plates (1  $\mu$ l at 10 mM highest concentration; 3-fold, 10-point dilutions in DMSO) were diluted in 1 $\times$  assay buffer (20 mM TRIS pH 8.0, 25 mM NaCl, 2 mM DTT and 0.05% Tween-20) to 1 mM using a Multitek robotic pipettor (Nanoscreen) and 1  $\mu$ l was spotted into the wells of 384-well low-volume Proxiplates (Perkin Elmer). To these plates 9  $\mu$ l of protein-peptide mix (**Supplementary Table 10**) in 1 $\times$  assay buffer was added by Multidrop (Thermo) to bring the final compound concentration to 100  $\mu$ M and incubated for 30 min at room temperature. Next, 2  $\mu$ l of a 1:1 mixture of streptavidin-conjugate donor and nickel-chelate or  $\alpha$ -GST acceptor beads (45  $\mu$ g/ml in 1 $\times$  assay buffer) were added and the plates were allowed to incubate for an additional 30 min in the dark at room temperature. After incubation, the plates were read on an EnVision multi-label reader equipped with an HTS AlphaScreen laser (Perkin Elmer). The expression and purification of the constructs used in this assay was described previously<sup>56</sup>.

*Methyltransferase selectivity assays.* The effect of UNC3866 on the methyltransferase activity of G9a, EHMT1, SUV39H1, SUV39H2, SETDB1, SETD8, SUV420H1, SUV420H2, SETD7, MLL1 trimeric complex, MLL3 tetrameric complex, EZH2 trimeric complex, PRMT1, PRMT3, PRMT5-MEP50 complex, PRMT6, PRMT7, PRMT8, PRDM9, SETD2, SMYD2, SMYD3, BCDIN3D and DNMT1 was assessed by monitoring the incorporation of tritium-labeled methyl group to lysine or arginine residues of peptide substrates using Scintillation Proximity Assay (SPA) as previously described<sup>57</sup>. Assays were performed in a 20  $\mu$ l reaction mixture containing 3H-SAM (Cat. # NET155V250UC; Perkin Elmer; www.perkinelmer.com) at substrate concentrations close to the  $K_m$  values for each enzyme. Three concentrations (1  $\mu$ M, 10  $\mu$ M and 50  $\mu$ M) of UNC3866 were used in all selectivity assays.

To stop the enzymatic reactions, 7.5 M guanidine hydrochloride was added, followed by 180  $\mu$ l of buffer (20 mM Tris, pH 8.0). The reactions were mixed and then transferred to a 96-well FlashPlate (Cat. # SMP103; Perkin Elmer; http://www.perkinelmer.com). The reaction mixtures in Flash plates were incubated for 1 h and the CPM were measured using a TopCount plate reader (Perkin Elmer). The CPM counts in the absence of compound for each data set were defined as 100% activity. In the absence of the enzyme, the CPM counts in each data set were defined as background (0%).

For DOT1L, NSD1, NSD2, NSD3, ASH1L, METTL21A, METTL21D, DNMT3A/3L, and DNMT3B/3L, a filter-based assay was used. In this assay, 20  $\mu$ l of reaction mixtures were incubated at 22 °C for 1 h, 100  $\mu$ l of 10% trichloroacetic acid (TCA) was added, and the samples were mixed and transferred to filter-plates (Millipore; cat.# MSFBN6B10; http://www.millipore.com). Plates were centrifuged at 2000 rpm (Allegra X-15R - Beckman Coulter, Inc.) for 2 min, followed by two additional 10% TCA washes, one ethanol wash (180  $\mu$ l), and centrifugation. Plates were dried and 100  $\mu$ l of MicroO (MicroScint-O; Cat. # 6013611, Perkin Elmer; www.perkinelmer.com) was added to each well, centrifuged and removed. Another 70  $\mu$ l of MicroO was added and the CPM was measured using a TopCount plate reader.

*Bromodomain selectivity assays.* Thermal melting experiments were carried out using an Mx3005p Real Time PCR machine (Stratagene). Proteins were buffered in 10 mM HEPES pH 7.5, 500 mM NaCl and assayed in a 96-well plate at a final concentration of 2  $\mu$ M in 20  $\mu$ L volume. Compounds were added at a final concentration of 25  $\mu$ M. SYPRO Orange (Molecular Probes) was added as a fluorescence probe at a dilution of 1:1,000. Excitation and emission filters for the SYPRO-Orange dye were set to 465 and 590 nm, respectively. The temperature was raised with a step of 3 °C per minute from 25 °C to 96 °C and fluorescence readings were taken at each interval.

*Demethylase AlphaScreen assays.* The demethylase AlphaScreen assay was performed in 384-well plate format using white proxiplates (Perkin Elmer) and compound transfers (100 nl) were performed using an ECHO 550 Acoustic Dispenser (Labcyte). All subsequent steps were carried out in assay buffer (50 mM HEPES pH 7.5, 0.1% (w/v) Bovine Serum Albumin and 0.01% (v/v) Tween-20). In brief, 5 ml of assay buffer containing demethylase enzyme at 2 $\times$  final assay concentration (see **Supplementary Table 11–13** for assay specifics) was pre-incubated for 15 min with dilutions of compound. The enzyme reaction was initiated by addition of substrate (5 ml) consisting of L-ascorbic acid (100 mM), 2-oxoglutarate (2 $\times$  final assay concentration), ferrous ammonium sulfate (2 $\times$  final assay concentration) and histone H3 substrate peptide (2 $\times$  final assay concentration) and the enzyme reaction was allowed to proceed for the required time. The final assay DMSO concentration was 1%. The enzyme reaction was stopped by addition of 5 ml assay buffer containing EDTA (30 mM) and NaCl (800 mM). Streptavidin Donor beads (0.08 mg/ml) and Protein-A conjugated acceptor beads (0.08 mg/ml) were pre-incubated for 1 h with anti-methyl mark antibody (4 $\times$  final assay concentration) and the presence of histone H3 product methyl mark was detected by addition of the pre-incubated AlphaScreen beads (5 ml). Detection was allowed to proceed for 2 h at room temperature and the assay plates were read in a BMG Pherastar FS plate reader (Excitation 680 nm / Emission 570 nm). Data were normalized to the no enzyme control and the IC<sub>50</sub> determined from the nonlinear regression curve fit using GraphPad Prism 5.

*NIMH PDSP functional assays.* The NK1 and NK2 receptors;  $\Delta$ -,  $\kappa$ -, and  $\mu$ -opioid receptors; and V<sub>1A</sub>, V<sub>1B</sub> and V<sub>2</sub> receptor functional assays were performed according to the NIMH PDSP Assay Protocol Book Version II found at <https://pdspdb.unc.edu/pdspWeb/?site=binding>.

*Cell culture and lysis.* PC3 cells were obtained from ATCC (CRL-1435) through the UNC Lineberger Tissue Culture Facility. Cells were cultured using GIBCO DMEM/F12 (Ham) supplemented with L-glutamine and 15 mM HEPES. Cells were trypsinized using 0.25% trypsin. Lysis was performed using Cytobuster protein extraction reagent supplemented with protease inhibitors and Benzonase (used at 25 U/ml). Samples were incubated at 37 °C for 10 min, followed by incubation at RT on a rotator for 20 min. The samples were spun down and the supernatant collected and transferred to a clean Eppendorf tube. Protein concentrations were quantified using the Bradford protein assay.

*Pull-down studies.* Cells were cultured on a T175 tissue culture flask until reaching ~80–90% confluency. Following trypsinization, the pellet was washed twice with PBS and lysed with 500  $\mu$ l of lysis buffer. When applicable, vehicle control or 100  $\mu$ M UNC3866 was added to an aliquot of 1,000  $\mu$ g of total protein and diluted to 500  $\mu$ l with 20 mM Tris (pH 8)/150 mM NaCl/0.1% Tween-20

(TBST). Pulldown reagents were bound to Streptavidin M-270 Dynabeads by rotating 30  $\mu$ l of beads with a 20-fold excess of pulldown reagent for 30 min at room temperature in. Unbound pulldown reagent was then removed and the beads were washed 3 $\times$  with 200  $\mu$ l of TBST. The lysate was then transferred to an Eppendorf tube containing 30  $\mu$ l of Streptavidin M-270 Dynabeads that had been pre-bound to UNC4195 or 2399. The mixture was rotated overnight at 4  $^{\circ}$ C. The following morning, the depleted lysate was removed and the beads were washed 3 $\times$  with 300–500  $\mu$ l of TBST. The beads were re-suspended with 30  $\mu$ l of 1 $\times$  Laemmli sample buffer and heated at 95  $^{\circ}$ C for 3 min. Fifteen microliters of the sample was then loaded into a gel and analyzed via western blotting using the appropriate primary antibody (**Supplementary Table 14**) and detected using a LI-COR Odyssey instrument and the appropriate fluorescent secondary antibodies (diluted 1:10,000 in PBST). One percent of input was used for western blotting of all proteins except CBX8 and HPH2. For both of these proteins, 3% of sample input was used.

**Cell proliferation assay.** PC3 cells were seeded at 200 cells/well into 24-well plates (Costar #3524). Cells were allowed to adhere overnight. The media (DMEM supplemented with 10% FBS) was then exchanged with fresh media containing DMSO, UNC3866 or UNC4219. On day 3, the media were exchanged with fresh media containing DMSO, UNC3866 or UNC4219. For dose-response studies, the EC<sub>50</sub> was derived from a six-point titration ranging from 100  $\mu$ M to 0.4  $\mu$ M of UNC3866 or UNC4219. At day 0, 3 or 6, cells were fixed with ice-cold methanol for 30 s. and rehydrated with PBS. Nuclei of the cells were stained with DAPI (0.05  $\mu$ g/ml) and numerated using High Content Microscopy (Array Scan High Content Analysis, Thermo Fisher #NX10002L). For dose-response studies, the cell count of UNC3866- or UNC4219-treated cells was normalized to the average cell count of DMSO-treated cells. The EC<sub>50</sub> was calculated using the “log[inhibitor] vs. the normalized response – variable slope” equation in GraphPad Prism 5.

**CellTiter-Glo luminescent cell viability assay.** The effect of UNC3866 and UNC4219 on cell viability was determined using a CellTiter-Glo ATP detection system (Promega #7573). Ten point, 1:3 dilution curves of compounds starting at 100  $\mu$ M final concentration were diluted to 5 $\times$  final concentration in PBS (vehicle control) and then 5  $\mu$ l were added to 384-well white, clear bottom tissue culture plates (Corning #3707) with a Multitek automated liquid handling device (Nanoscreen, Charleston, SC). Twenty microliters of low passage, subconfluent HEK293T/PC3 cells grown in Dulbecco's Modified Eagle's Medium without phenol red (GIBCO #31053) and supplemented with 10% Fetal Bovine Serum (GIBCO #26140) were immediately added at a density of 5,000 cells per well using a Multidrop 384 (Titertek). Cell plates were incubated for 48 h at 37  $^{\circ}$ C and 5% CO<sub>2</sub> and then lysed with 25 microliters of CellTiter-Glo reagent. Luminescence was read on an Envision plate reader (Perkin Elmer) after 15 min at room temperature in dim light.

**Cellular thermal shift assay (CETSA).** A PC3 cell pellet consisting of 10<sup>7</sup> cells was re-suspended in 1,200  $\mu$ l of PBS complete with protease inhibitors. The suspension was subjected to 3 consecutive freeze-thaw cycles performed with liquid N<sub>2</sub> in order to lyse the cells. Following the final freeze-thaw cycle, the suspension was centrifuged at 20,000g for 20 min at 4  $^{\circ}$ C. The supernatant was collected and the resulting pellet was discarded. The lysate was then divided into two tubes (600  $\mu$ l) each. The volume of each tube was brought up to 650  $\mu$ l with PBS + protease inhibitors. DMSO or 30  $\mu$ M UNC3866 was added to each tube. The tubes were incubated at RT for 30 min. The lysates were aliquoted into 12 thin-walled PCR tubes (50  $\mu$ l/tube). The tubes were incubated for 3 min at temperatures ranging from 37  $^{\circ}$ C to 73  $^{\circ}$ C in 3 $^{\circ}$  increments in a Mastercycle nexus gradient qPCR machine. Tubes were then incubated at room temperature for 3 min followed by centrifugation at 20,000g for 20 min at 4  $^{\circ}$ C. The supernatant was collected. Ten microliters of 6 $\times$  Laemmli buffer was added to each tube. The tubes were then heated to 95  $^{\circ}$ C for 3 min. Twenty microliters was then loaded into an SDS page gel and analyzed by western blotting.

**CBX7 knockdown and quantitative PCR.** PC3 cells were seeded at a density of 10<sup>5</sup> cells/well in a 6-well plate and allowed to adhere overnight. The media was then exchanged for 2 ml of media containing polybrene (8  $\mu$ g/ml) and pre-packaged shRNA (200,000 Transducing Units/well) targeting CBX7 (Mission CBX7, SHCLNV (Sigma) TRCN0000019144RNA) or Non-targeting control (Mission pLKO.1-puro Non-Target shRNA, SHC016V (Sigma)). After 20 h, the transduction mixture was exchanged with fresh media (GIBCO DMEM/F12 (Ham) supplemented with L-glutamine and 15 mM HEPES) and allowed to incubate for an additional 24 h. The media was then replaced with media containing 2  $\mu$ g/ml of puromycin and allowed to incubate until selection was complete as

evidence by the death of all control cells that were not transduced. RNA was extracted using the RNeasy Plus Mini Kit (Qiagen) and RNA was quantified on the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific). The reverse transcription reaction was performed with 500 ng RNA and the Superscript III First Strand Synthesis Supermix (Invitrogen #1172-050) according to manufacturer instructions. The resulting cDNA was diluted 1:10, and 2  $\mu$ l was used for each PCR reaction. Quantitative PCR was performed using Roche FastStart Universal SYBR Green Master Mix (Rox) in a 384-well plate format on the ABI 2900HT instrument. Oligo sequences are available upon request.

**Phase contrast microscopy images.** PC3 cells were seeded at 200 cells/well into 24-well plates (Costar #3524). Cells were allowed to adhere overnight. The media (DMEM supplemented with 10% FBS) was then exchanged with fresh media containing DMSO or 30  $\mu$ M UNC3866. On day 3, the media were exchanged with fresh media containing DMSO or 30  $\mu$ M UNC3866. At day 6, cells were fixed with ice-cold methanol for 30 s. and rehydrated with PBS. Samples were imaged using Volocity software and an Olympus IX81 microscope equipped with an Orca ER digital camera.

**Senescence-associated  $\beta$ -Gal staining.** PC3 cells were seeded in a 96-well plate at a density of 100–200 cells/well and allowed to adhere overnight. The media was then replaced with fresh media containing 30  $\mu$ M UNC3866 or DMSO. After 3 days, the media was removed and the cells were fixed and stained using the Senescence  $\beta$ -Galactosidase Kit (#9860) from Cell Signaling Technology according to the manufacturer's recommendations. Wells were imaged using bright-field microscopy with an Olympus IX81 microscope equipped with an Orca ER digital camera.

**Protein expression and purification.** *Expression constructs.* The chromodomains of CBX2 (residues 9–66 of NP\_005180), CBX4 (residues 8–65 of NP\_003646), CBX6 (residues 8–65 of NP\_055107) and CBX8 (residues 8–61 of NP\_065700) were expressed with N-terminal His-tags in pET28 expression vectors. The chromodomain of CBX7 (residues 8–62 of NP\_783640) was expressed with a C-terminal His-tag in a pET30 expression vector. The chromodomains of CDY1 (residues 1–101 of NP\_733841), CDYL1b (residues 1–78 of NP\_004815) and CDYL2 (residues 1–75 of NP\_689555) were expressed with N-terminal GST-tags in pGEX4T expression vector.

*Protein expression and purification.* All expression constructs were transformed into Rosetta BL21(DE3)pLysS competent cells (Novagen, EMD Chemicals, San Diego, CA). Protein expression was induced by growing cells at 37  $^{\circ}$ C with shaking until the OD<sub>600</sub> reached ~0.6–0.8 at which time the temperature was lowered to 18  $^{\circ}$ C and expression was induced by adding 0.5 mM IPTG and continuing shaking overnight. Cells were harvested by centrifugation and pellets were stored at –80  $^{\circ}$ C.

His-tagged proteins were purified by re-suspending thawed cell pellets in 30 ml of lysis buffer (50 mM sodium phosphate pH 7.2, 50 mM NaCl, 30 mM imidazole, 1 $\times$  EDTA free protease inhibitor cocktail (Roche Diagnostics, Indianapolis, IN)) per liter of culture. Cells were lysed on ice by sonication with a Branson Digital 450 Sonifier (Branson Ultrasonics, Danbury, CT) at 40% amplitude for 12 cycles with each cycle consisting of a 20 s pulse followed by a 40 s rest. The cell lysate was clarified by centrifugation and loaded onto a HisTrap FF column (GE Healthcare, Piscataway, NJ) that had been preequilibrated with 10 column volumes of binding buffer (50 mM sodium phosphate, pH 7.2, 500 mM NaCl, 30mM imidazole) using an AKTA FPLC (GE Healthcare, Piscataway, NJ). The column was washed with 15 column volumes of binding buffer and protein was eluted in a linear gradient to 100% elution buffer (50 mM sodium phosphate, pH 7.2, 500 mM NaCl, 500 mM imidazole) over 20 column volumes. Peak fractions containing the desired protein were pooled and concentrated to 2 ml in Amicon Ultra-15 concentrators 3,000 molecular weight cut-off (Merck Millipore, Carrigtwohill Co. Cork IRL). Concentrated protein was loaded onto a HiLoad 26/60 Superdex 75 prep grade column (GE Healthcare, Piscataway, NJ) that had been preequilibrated with 1.2 column volumes of sizing buffer (25 mM Tris, pH 7.5, 250 mM NaCl, 2 mM DTT, 5% glycerol) using an ATKA Purifier (GE Healthcare, Piscataway, NJ). Protein was eluted isocratically in sizing buffer over 1.3 column volumes at a flow rate of 2 ml/min collecting 3-ml fractions. Peak fractions were analyzed for purity by SDS-PAGE and those containing pure protein were pooled and concentrated using Amicon Ultra-15 concentrators 3,000 molecular weight cut-off (Merck Millipore, Carrigtwohill Co. Cork IRL).

GST-tagged proteins were purified by re-suspending thawed cell pellets in 30 ml of lysis buffer (1 $\times$  PBS, 5 mM DTT, 1 $\times$  EDTA free protease inhibitor

cocktail (Roche Diagnostics, Indianapolis, IN) per liter of culture. Cells were lysed on ice by sonication as described for His-tagged proteins. Clarified cell lysate was loaded onto a GStrap FF column (GE Healthcare, Piscataway, NJ) that had been pre-equilibrated with 10 column volumes of binding buffer (1x PBS, 5mM DTT) using a AKTA FPLC (GE Healthcare, Piscataway, NJ). The column was washed with 10 column volumes of binding buffer and protein was eluted in 100% elution buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 10 mM reduced glutathione) over 10 column volumes. Peak fractions containing the desired protein were pooled and concentrated to 2 ml in Amicon Ultra-15 concentrators, 10,000 molecular weight cut-off (Merck Millipore, Carrigtwohill Co. Cork IRL). Concentrated protein was loaded onto a HiLoad 26/60 Superdex 200 prep grade column (GE Healthcare, Piscataway, NJ) that had been pre-equilibrated with 1.2 column volumes of sizing buffer (25 mM Tris, pH 7.5, 250 mM NaCl, 2 mM DTT, 5% glycerol) using an ATKA FPLC (GE Healthcare, Piscataway, NJ). Protein was eluted isocratically in sizing buffer over 1.3 column volumes at a flow rate of 2 ml/min collecting 3-ml fractions. Peak fractions were analyzed for purity by SDS-PAGE and those containing pure protein were pooled and concentrated using Amicon Ultra-15 concentrators 10,000 molecular weight cut-off (Merck Millipore, Carrigtwohill Co. Cork IRL).

**Affinity tag removal.** The N-terminal affinity tag was removed from CBX2, CBX6, CDY1, CDYL1b and CDYL2 by thrombin cleavage according to manufacturer's recommendations (Novagen, EMD Chemicals, San Diego, CA). Briefly, purified protein was incubated with biotinylated thrombin at a final concentration of 1 unit thrombin per milligram tagged protein for 16 h at 4 °C. The cleavage reaction was then passed over either a HisTrap FF column (His-tagged proteins) or a GStrap FF column (GST-tagged proteins) to remove any protein that still retained the tag. The column flow through was collected and incubated with streptavidin agarose to remove thrombin from the sample. Streptavidin agarose was removed from the sample by centrifugation and the protein was concentrated using Amico Ultra-15 concentrators, 3,000 molecular weight cut-off (Merck Millipore, Carrigtwohill Co. Cork IRL). Protein was exchanged into a buffer containing 25 mM Tris, pH 7.5, 150 mM NaCl, 2 mM  $\beta$ -mercaptoethanol before use in ITC.

**X-ray crystallography. Preparation of proteins for crystallization.** Expression and purification followed a previous description<sup>3</sup>. The coding DNA fragments of single chromodomains of CBX2 (residues 9–62), CBX4 (residues 8–65), CBX7 (residues 8–62) and CBX8 (residues 8–61) were subcloned into modified pET28-MHL or pET28a-Lic vectors to encode N-terminal His-tagged fusion proteins. Recombinant proteins were overexpressed in *E. coli* BL21 (DE3) Codon plus RIL cells (Stratagene) at 15 °C, purified by affinity chromatography on Ni-nitrilotriacetate resin (Qiagen), and treated with TEV or thrombin protease to remove the tag. The proteins were further purified by gel filtration over a Superdex75 column (GE Healthcare, Piscataway, NJ). For crystallization, purified proteins were concentrated to 10 mg/ml in a buffer containing 20 mM Tris, pH 7.5, 150 mM NaCl and 1 mM DTT.

**Crystallization.** Purified protein was mixed with compound at a 1:2 molar ratio and crystallized using the sitting drop vapor diffusion method at 18–20 °C by mixing 0.5  $\mu$ L of the protein with 0.5  $\mu$ L of the reservoir solution. Reservoir solutions were 20% PEG 3350, 0.2 M KSCN for CBX2, 30% PEG 3350, 0.2 M NaCl, 0.1 M sodium cacodylate, pH 5.5 for CBX4, 20% PEG 3350, 0.2 M ammonium formate for CBX7, and 1.4 M sodium citrate, 0.1 M HEPES, pH 7.5, 5% glycerol for CBX8. Crystals were soaked in a cryoprotectant consisting of 85% reservoir solution and 15% glycerol, and flash-frozen in liquid nitrogen.

**Data collection/structure determination and refinement.** Due to space restrictions, literature references to selected computer programs are provided in the Supplementary Information of refs. 4–16). Details on diffraction data and model refinement are given in **Supplementary Table 15**. Diffraction data for model refinement were collected at a copper rotating anode (CBX2, CBX7) or APS beamline 19ID (CBX4, CBX8; wavelength 0.979 Å). Diffraction intensities were integrated with XDS<sup>58</sup> and merged with AIMLESS<sup>59</sup>. Molecular replacement was performed with the program PHASER<sup>60</sup>. Geometry restraints for the modified lysyl residue as well as the 4-*tert*-butylbenzoyl protecting group were prepared with eLBOW/Mogul, restraints for selected peptide links with jLigand. Current crystallographic models underwent iterative rebuilding, refinement and validation in COOT, REFMAC and MOLPROBITY, respectively. Favored regions of the respective model's MOLPROBITY Ramachandran plot included at least 98% of the residues. Anisotropic displacement parameters were analyzed on the PARVATI server. PDB\_EXTRACT as well as PHENIX, IOTBX and CCP4 software were used to compile data for model deposition in the Protein Data Bank and for publication.

Specifically: CBX2. Coordinates from PDB entry 3H91 (ref. 3) were used for molecular replacement.

CBX4. The structure of an isomorphous crystal was solved by molecular replacement with CBX2 coordinates. ARP/WARP was used in “map improvement” and “automated model building” modes. The model refinement was further refined against data from the current crystal.

CBX7. The structure was first solved in the lower-symmetry P41 setting by molecular replacement with coordinates from PDB entry 4MN3 (ref. 27). After merging of data in space group P41212, molecular replacement was performed with preliminarily refined CBX7 coordinates.

CBX8. The structure of an isomorphous crystal was solved by molecular replacement with CBX7 coordinates. ARP/WARP was used in “map improvement” and “automated model building” modes. The model was further refined against data from an additional crystal and, subsequently, the current crystal.

51. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435–447 (2008).
52. Tiwary, P. & Parrinello, M. From metadynamics to dynamics. *Phys. Rev. Lett.* **111**, 230602 (2013).
53. Voter, A.F. Hyperdynamics: accelerated molecular dynamics of infrequent events. *Phys. Rev. Lett.* **78**, 3908–3911 (1997).
54. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
55. Wigle, T.J. *et al.* Screening for inhibitors of low-affinity epigenetic peptide-protein interactions: an AlphaScreen-based assay for antagonists of methyl-lysine binding proteins. *J. Biomol. Screen.* **15**, 62–71 (2010).
56. Perfetti, M.T. *et al.* Identification of a fragment-like small molecule ligand for the methyl-lysine binding protein, 53BP1. *ACS Chem. Biol.* **10**, 1072–1081 (2015).
57. Barsyte-Lovejoy, D. *et al.* (R)-PFI-2 is a potent and selective inhibitor of SETD7 methyltransferase activity in cells. *Proc. Natl. Acad. Sci. USA* **111**, 12853–12858 (2014).
58. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
59. Evans, P.R. & Murshudov, G.N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
60. McCoy, A.J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

## Addendum: A cellular chemical probe targeting the chromodomains of Polycomb repressive complex 1

Jacob I Stuckey, Bradley M Dickson, Nancy Cheng, Yanli Liu, Jacqueline L Norris, Stephanie H Cholensky, Wolfram Tempel, Su Qin, Katherine G Huber, Cari Sagum, Karynne Black, Fengling Li, Xi-Ping Huang, Bryan L Roth, Brandi M Baughman, Guillermo Senisterra, Samantha G Pattenden, Masoud Vedadi, Peter J Brown, Mark T Bedford, Jinrong Min, Cheryl H Arrowsmith, Lindsey I James and Stephen V Frye

Addendum to: *Nat. Chem. Biol.* <https://doi.org/10.1038/nchembio.2007>, published online 25 January 2016.

In our original Article, we reported the discovery of a chemical probe for the CBX domains of Polycomb repressive complex 1 (PRC1). In subsequent studies undertaken in our laboratories by Kelsey Lamb (Center for Integrative Chemical Biology and Drug Discovery, Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA), we have observed that our modular synthetic approach detailed in this article leads to epimerization at the alanine position. In the reported synthesis in the original paper, the amide coupling of tripeptide intermediate **13** to tripeptide intermediate **7** for the formation of final product UNC3866 (or for the formation of UNC4219, the amide coupling of tripeptide intermediate **13** to tripeptide intermediate **10**) resulted in approximately 40% epimerization at the alanine side chain by NMR. The linear approach outlined in the revised protocol in the Supplementary Note avoids this problem by utilizing a synthetic route in which each amide coupling is performed on the N-Boc-protected amino acid, a method known to diminish epimerization, as validated in our studies by NMR spectra. We note that UNC3866 was originally synthesized using the linear route, despite the modular synthesis reported in the original Supplementary Note, and therefore biological data reported for this compound were not affected. While UNC4219, UNC4007 and UNC4195 were synthesized using the modular route, resynthesis of compounds by the linear route validated that all data obtained using these compounds were unaffected by alanine epimerization. Based on these observations, we provide an alternative synthetic route in this Addendum that avoids alanine epimerization. The protocols have been updated in the Supplementary Note to this Addendum.

Published online: 06 March 2019  
<https://doi.org/10.1038/s41589-019-0242-5>