CBX1 mutations cause a neurodevelopmental syndrome due to heterochromatin organizational alterations

Aiko Iwata-Otsubo[1], Suzanna Lindsay-Temple[2,3], Kerith-Rae Dias[2,4], Ying Zhu[2], Sarah K Fiordaliso[1], Chun Su[5], Alyssa Ritter[1], Samuel W Baker[6], Yukiko Kuroda[1], Beth Keena[1], Struan F.A. Grant[1,5,7] , Elaine Zackai[1,7], Matt Edwards[8], Carey-Anne Evans[2,4], Matthew Dulik[6], Michael F Buckley[2], Tony Roscioli[2,3,4], Kosuke Izumi[1,7]

[1] Division of Human Genetics, Department of Pediatrics, The Children's Hospital of Philadelphia, Philadelphia, USA

[2] Randwick Genomics Laboratory, NSW Health Pathology, Prince of Wales Hospital, Randwick, NSW, Australia

[3] Centre for Clinical Genetics, Sydney Children's Hospital, Randwick, NSW, Australia

[4] Neuroscience Research Australia and Prince of Wales Clinical School, University of New South Wales, Kensington, NSW, Australia

[5] Center for Spatial and Functional Genomics, The Children's Hospital of Philadelphia, Philadelphia, USA

[6] Division of Genomic Diagnostics, Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, USA

[7] Department of Pediatrics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA.

[8] Hunter Genetics, Newcastle, NSW, Australia

Corresponding Author:

Kosuke Izumi, MD, PhD, 1010A, ARC, Division of Human Genetics, The Children's Hospital of Philadelphia, 3615 Civic Center Blvd. Philadelphia, PA 19104. USA. E-mail: izumik1@email.chop.edu

Declaration of Interests

The authors declare no competing interests.

ABSTRACT

The heterochromatin protein 1 (HP1) family of proteins represents an essential structural component of heterochromatin formation and organization. There are three HP1 proteins (HP1α, HP1β, and HP1γ), and HP1β is the only essential HP1 protein during mammalian development. Here we report a neurodevelopmental disorder due to missense mutations in the *CBX1* gene which encodes HP1β. Two unrelated individuals were found to have *de novo* missense mutations in *CBX1*. Their clinical features include developmental delay, hypotonia and autistic features, suggesting the importance of HP1β in neuronal function. Identified mutations are in a known functional domain, chromodomain, and the identified mutations abolished HP1β-chromatin interactions. Our transcriptome and epigenome analyses revealed that reduced HP1β chromatin affects gene expression of H3K27me3 marked genes, suggesting the importance of HP1β in facultative hetrochromatin organization during human development. This diagnosis represents the first genetic disorder due to germline mutations in genes encoding HP1 proteins.

INTRODUCTION

The heterochromatin protein 1 (HP1) family of proteins represents an essential structural component of heterochromatin formation and organization[1]. HP1 proteins form homo- and heterodimers of HP1, and they bind to methylated H3K9 and bring two methylated H3K9s in close proximity, resulting in chromatin compaction[2]. HP1 is mainly known to regulate chromatin organization of constitutive heterochromatin, stably repressed genomic region[3]. However, recent reports suggest the possibility of HP1 functioning outside of constitutive heterochromatin such as facultative heterochromatin, reversibly repressed genomic regions[4].

In humans, there are three HP1 proteins (HP1α, HP1β, and HP1γ), and their protein sequences are highly conserved. Despite the critical importance of HP1 proteins in chromatin organization and differentiation regulation, no human disorders have been associated previously with HP1 protein mutations. Here we report a neurodevelopmental syndrome due to *de novo* missense mutations in *CBX1*, encoding HP1β. Our functional analyses of *CBX1* mutations confirmed that the reduced chromatin binding of HP1 proteins causes overexpression of its target within H3K27me3-marked facultative heterochromatin.

MATERIALS AND METHODS

Human Subjects: Individuals with *CBX1* mutations were recruited through GeneMatcher[9] and a network of collaborating clinical geneticists. All individuals were enrolled in the research study under an institutional review board-protocol.

Cell culture: Lymphoblastoid cell lines (LCLs) were cultured in RPMI 1640 with 300mg/L L-glutamine (Life Technologies, 11875085) supplemented with 20% HyClone FBS (Fisher Scientific, SH3007103), 0.2% penicillin-streptomycin (Life Technologies, 15140122), 0.2% Plasmocin (Invivogen, ant-mpp), and 1% Glutamax (Life Technologies, 35050061). HEK293T cell line was cultured in DMEM with 4.5g/L D-Glucose and 110mg/L sodium pyruvate (Life Technologies, 11360070) supplemented with 10% HyClone FBS, and 0.2% penicillin-streptomycin. All cells were

cultured at 37 degrees in 5% CO2. *CBX1* cDNA vectors were transfected to HEK293T cells using Lipofectamine 2000 (Life Technologies, 11668-030) according to the manufacturer's protocol.

Reagents: Myc-DDK-tagged-*CBX1* cDNA vector was purchased from Origene (RC205672). *CBX1* mutations were introduced using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs Inc., E0554S) following the manufacturer's protocol. Sanger sequencing confirmed the intended change and to ruled out additional, nonspecific changes.

Histone association assay was performed according to the protocol described previously [10]. Following 48 hours after the *CBX1* cDNA vector transfection, cells were fixed with 1% of formaldehyde for 5-10 min. Cross-linkning reaction was quenched by 2.5 M Glycine. Cells were washed with cold PBS twice, and then PBS or lysis buffer 1 (20 mM Tris–HCl, pH 7.5, 10 mM NaCl, 1 mM EDTA, 0.2 % NP-40, 1 mM PMSF) was added. Cells were harvested into a 15ml conical tube, and were centrifuged at 1,500 × g for 5 min (4 degrees). Supernatant were discarded, then, pellets were washed with 1mL of PBS, and transferred to 1.5ml tube. Cells were centrifuged at 5,000 × g for 5 min. For the immunoprecipitation, Protein A (for rabbit antibody) or Protein G (for mouse antibody) magnetic beads were washed by ~500ul BSA/PBS (4°) twice. H3K9me3, H4K20me3 and H3K27me4 antibodies were added to Protein A/G magnetic beads, and they were rotated at 4°C for more than 3 h. After the beads-antibody reaction, they were washed twice by BSA/PBS, and once with lysis buffer 3 (20 mM Tris–HC l, pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1 % Triton X-100, 0.1 % Na-Deoxycholate, 0.1 % SDS, protease inhibitors), and resuspended with 100µl of LB3. Collected cells were resuspended with 1 ml of LB1 (995ul LB1 +5ul PMSF) and were lysed on ice for 10 min. After cellular lysis, the tubes were centrifuged at 2,000×g for 5 min, and pellets were resuspend in 1 ml of LB2 (995ul LB2 +5ul PMSF). The tubes were placed on ice for 10 min, then span again to remove supernatant. Obtained pellets are lysed with 1ml of LB3 for 10 min. After the reaction, the tubes were centrifuged at 2,000 × g for 5 min, and supernatant were discarded. 400 µl of LB3 was added to the tube, and placed on ice for 10 min. After the reaction, tubes were centrifuged at 2,000×g for 1 min. Obtained pellets were

sonicated by using Branson Sonifier 250D (40% amp, 12 sec, 8 cycles). After the sonication, tubes were centrifuged at 20,000 × g for 15 min and transfer supernatant to new 1.7 ml tube (Coster) take 30 µl aliquot as WCE sample and an additional 10ul aliquot for WB. Harvested input samples were mixed with antibody beads complex, and were rotated overnight at 4°C. After overnight antibody reaction, magnet beads were collected by using magnetic stand. Beads were washed with 1 mL cold RIPA buffer five times and then washed once with 1mL cold TE50. After the beads wash, the beads were centrifuged at 2,000 × g for 1 min, be placed in magnetic holder. Harvested washed magnet beads were re-suspend in 50 µl of EB to make bead suspension. Then, 2X Laemmli sample buffer (50 µl) were added to the beads, and placed at 95 °C heat block for 30 min with vigorous mixing by vortex every 5 min. After the boiling, tubes were centrifuge at 2,000 × g for 1 min, and the soluble ChIP lysates were collected to a new tube. Collected ChIP samples were used for western blot. Antibodies used for ChIP reactions are H3K9me3 (Abcam: ab8898), H3K27me3(Active Motif: MABI0323) and H4K20me3 (Abcam: ab9053).

Immunofluorescence (IF) staining: HEK293T cells were settled on poly-lysine coated coverslips and transfected with the CBX1 cDNA vector. Cells were fixed in 4% formaldehyde in PBS for 10 min at room temperature (RT) 48 hours after transfection, permeabilized in 0.5% Triton X-100 in PBS for 5 min at RT, then incubated with IF block (2% FBS, 2% BSA, 0.1% Tween, 0.02% Sodium Azide in PBS) for 20 min at RT. Cells were incubated with primary antibodies for 1 hour at RT, followed by three washes in 0.2% Tween in PBS. Primary antibody used was mouse anti FLAG (Sigma F1804-200uG). Subsequently, cells were incubated with Alexa Fluor 488 donkey anti-mouse IgG (1:500) for 1 hour at RT. After three washes in 0.2% Tween in PBS, and one wash with PBS, cells were mounted in Vectashield containing 4′,6-diamino-2-phenilinodole (DAPI, Vector Laboratories, H-1200). Images were captured on a Leica wide-field fluorescence microscope, using a 1.4 NA 63x oil-immersion objective (Leica) and an ORCA-Flash4.0 V2 Digital CMOS camera (Hamamatsu, C11440-22CU), controlled by LAS X software (Leica).

Next generation sequencing and bioinformatic analysis: LCLs obtained from one female patient with *CBX1* mutation and gender-ethnicity matched three control subjects were used. Total RNA was extracted using TRIzol (Life Technologies) and Nucleospin RNA (Macherey-Nagel) following the manufacturer's instructions. These RNA samples were sequenced at Genewiz (South Plainfield, NJ). About 78-120M raw reads were generated by HiSeq with paired-end 150bp sequencing for each sample. Sequenced reads were mapped to GRCh38 reference genome. RNA-sequencing data was analyzed by CLC genomics workbench with a default analysis setting (https://www.qiagenbioinformatics.com/products/clc-genomics-workbench). Coding genes and ncRNA genes were considered for RNA-sequencing analysis. ATAC-seq was performed as previously described[11]. The triplicates derived from one female *CBX1* mutant LCL and gender-ethnicity matched LCL are harvested and counted. 50,000 cells of each sample were spun down at 550 ×g for 5 min at 4°C. The cell pellet was then resuspended in 50 µl cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% IGEPAL CA-630) and spun down immediately at 550 ×g for 10 min, 4°C. The nuclei were resuspended on ice in the transposition reaction mix (2x TD Buffer, 2.5ul Tn5 Transposes and Nuclease Free H_2O) (Illumina Cat #FC-121-1030, Nextera) on ice and the transposition reaction was incubated at 37°C for 45 min. The transposed DNA was then purified using a the MinElute Kit (Qiagen) adjusted to 10.5 µl elution buffer. The transposed DNA was converted into libraries using NEBNext High Fidelity 2x PCR Master Mix (NEB) and the Nextera Index Kit (illumina) by PCR amplification for 12 cycles. The PCR reaction was subsequently cleaned up using AMPureXP beads (Agencourt), checked on a Bioanalyzer 2100 (Agilent) high sensitivity DNA Chip (Aglient), and paired-end sequenced on the Illumina NovaSeq 6000 platform (51bp read length) at the Center for Spatial and Functional Genomics at CHOP. ATAC-seq peaks were called using the ENCODE ATAC-seq pipeline (https://www.encodeproject.org/atac-seq/). Briefly, pair-end reads from all replicates for each condition (mutant vs control) were aligned to hg19 genome using bowtie2, and duplicate reads were removed from the alignment. Aligned tags were generated by modifying the reads alignment

by offsetting +4bp for all the reads aligned to the forward strand, and -5bp for all the reads aligned to the reverse strand. Narrow peaks were called independently for pooled replicates for each cell type using macs2 (-p 0.01 --nomodel --shift -75 --extsize 150 -B --SPMR --keep-dup all --call-summits) and ENCODE blacklist regions were removed from called peaks. Finally, a consensus of open chromatin regions (OCRs) were obtained by consolidating the peak sets from both mutants and controls using bedtools merge (v2.25.0). To determine whether an OCR is differentially accessible between mutant and control, the de-duplicated read counts for consensus OCRs were calculated for each replicate and normalized against background (10K bins of genome) using the R package Csaw[76] (v 1.8.1). OCRs with median value of less than 0.45 CPM (5~8 reads per OCR) across all replicates were removed from further differential analysis. Accessibility differential analysis was performed using glmQLFit approach fitting model ~condition in edgeR (v 3.16.5) and lmFit from limmaVoom (v 3.30.13) but with the normalization scaling factors calculated from csaw. Differential OCRs between conditions were identified if FDR<0.05 and absolute log2 fold change >1. Region enrichment analysis was performed using R package LOLA {2016 Feb 15;32(4):587-9. doi: 10.1093/bioinformatics/btv612} with differentially open regions as user set and all open regions as universal set. ENCODE ChIP-seq data analyses was performed by using MASER with a default analysis setting[12]. RNA-seq and ATAC-seq stats are available as Supplementary Table 1. H3K9me3 ChIP-seq result (ENCFF000AUK), H3K27me3 ChIP-seq data (ENCFF000ASZ), and corresponding input data (ENCFF000ARK) were obtained from the ENCODE database (encodeproject.org). Downloaded data were used to identify H3K9me3- and H3K27me3-marked genes in the LCL. For the identification of H3K9me3- and H3K27me3-marked genes SICER V1.1[17] was used for the peak-calling, and ChIPPeakAnno was used for the peak annotations[18].

RESULTS

Through independent exome sequencing analyses at the Children's Hospital of Philadelphia and Randwick Genomics Laboratory, two individuals with *de novo* heterozygous *CBX1* mutations were identified. Clinical features are compared in the Table 1, and their mutation information are summarized in the Table 2. Identified mutations reside within a known functional domain of HP1β, chromodomain (CD) (Fig 1). HP1 has two known functional domains, CD and a chromoshadow domain (CSD). CD directly interacts with methylated H3K9, whereas CSD serves as a dimerization interface with other HP1 proteins[8]. Various *in silico* computer programs predicted these variants to be deleterious.

The CD of HP1β mediates its binding to methylated lysine 9 in histone H3 (H3K9), which is constitutive heterochromatin marks, and subsequently organizes chromatin structure by forming closed compacted chromatin. It was previously shown that both HP1β that lacks CD and with a mutation in CD (V23M) localize throughout the nucleus instead of accumulating in heterochromatin[13]. We tested if patient-identified HP1β mutations localizes throughout the nucleus differently from the wild type (WT) protein. By using FLAG-CBX1-cDNA overexpression in HEK293T cells, we evaluated the intranuclear distribution of WT and mutant HP1β by immunofluorescence using an antibody against FLAG. WT HP1β co-localizes with chromocenter, which is composed of constitutive heterochromatin. Signals of HP1β with patient-identified CD mutants (W52L and N57D) showed homogeneous nuclear staining, indicating mislocalization of the mutant HP1β, similar to that of V23M (Fig 2).

These missense mutations reside within a peptide sequence shown previously to be directly involved in the interaction between HP1β and H3K9me3[14]. Hence, we suspected that mutant HP1β binds less efficiently to chromatin. Using WT and N57D mutant (Mut) FLAG-CBX1 cDNA vector, we evaluated the amount of H3K9me3-bound HP1β. Chromatin immunoprecipitation (ChIP)-western blot (WB) showed the reduction of Mut FLAG-CBX1 binding to H3K9me3, compared to WT(Fig 3). We also evaluated interactions between HP1β and other repressive histone marks, and similar results were observed for H4K20me3, as well as

H3K27me3, histone modification seen in facultative heterochromatin (Fig 3). Similarly, HP1β with W52L mutation showed reduced binding to H3K9me3 chromatin (Supplementary Figure 1).

Next, to evaluate the transcriptomic and epigenomic effects of the *CBX1* mutation, we performed RNA-sequencing (RNA-seq) using the patient and control LCLs. Transcriptome analysis detected 205 upregulated genes and 69 downregulated genes in the patient sample (Supplementary Table 2) (Fig 4a,b). A majority of the differentially expressed genes (DEGs) was upregulated genes in the patient sample. Gene list enrichment analysis of the identified DEGs using Enricher[15] indicates the enrichment of H3K27me3-marked genes, despite HP1 is most known as a regulator of constitutive heterochromatin marked by H3K9me3 (Supplementary Table 3). Particularly, H3K27me3-marked genes in a LCL sample (GM12878) used in the ENCODE epigenome project showed the highest overlap (Supplementary Table 4). Conversely, H3K9me3-marked genes in the GM12878 LCL sample were not enriched within CBX1 mutant DEGs (Supplementary Table 3).

We further evaluated the possibility of H3K27me3 marked gene misexpression by using the ENCODE data set[16]. H3K9me3 ChIP-seq result (ENCFF000AUK), H3K27me3 ChIP-seq data (ENCFF000ASZ), and corresponding input data (ENCFF000ARK) were obtained, and 4278 and 5845 genes were identified to have H3K9me3 and H3K27me3 peaks nearby, respectively. Fold gene expression differences of these H3K9me3, H3K27me3 genes were compared by using RNA-seq dataset. Although H3K9me3-marked genes did not show major alterations of gene expression, H3K27me3-marked genes showed a tendency for upregulation in the patient sample, compared to the controls (Student t-test $p= 1.3 \times e^{-73}$)(Fig 4c).

Given the known role of HP1 in forming heterochromatin, we hypothesized that reduced HP1β would cause increase chromatin accessibility, resulting in overexpression of genes residing within open chromatin regions. To evaluate the alteration of chromatin accessibility due to the reduction of HP1β chromatin binding, we performed ATAC-seq and identified 181,195 open chromatin regions (OCRs) in either mutant or control (Supplemental Table 4). Overall,

chromatin accessibility was slightly increased in the CBX1 mutant sample, compared to the control (Fig 4d), confirming that HP1 involves in forming heterochromatin and reduced HP1β would cause increase chromatin accessibility. Differential analysis identified 3130 differentially accessible regions (DARs) between the mutant and control samples, in which 57% (1813) of DARs showed increased accessibility in the *CBX1* mutant samples. Among these, 541 DARs were located on gene promoter regions (-1500/+500bp around transcription start side) and corresponded to 574 genes (Supplemental Table 4). Among these 574 differentially accessible genes, 17 genes were also differentially expressed mutant and control. We also found positive correlation between gene expression change and promoter accessibility change (Fig 4e, f).

Since HP1 is known to bind to constitutive heterochromatin marks H3K9 and regulates chromatin compaction across genomic repeats, we evaluated chromatin accessibility change over these regions. As expected, in the comparison to other OCRs, the regions with increased chromatin accessibility due to *CBX1* mutation significantly enriched at H3K9me3 with highest enrichment odds ratio (Fig 4g. Fisher exact test odds ratio = 4.05, P-value = 2.3e−08). H3K27me3- marked genomic regions were also enriched within upregulated DARs seen in the mutant sample, although the degree of enrichment was less noticeable than those of H3K9me3 (Fig 4g). Furthermore, we performed the enrichment over predefined heterochromatin repeats across genome, including SINE, LINE, LTR, DNA satellites and simple repeats. However, the regions with differential chromatin accessibility only shows a moderate enrichment for SINE (Supplemental Table 5, Fisher exact test odds ratio = 1.49, P-value = 1.1e-17) and no major alterations of chromatin accessibility were enriched at the LIN, LTR, simple repeats and DNA satellites (Supplemental Table 5). This suggests that the missense mutation of *CBX1* from patient may mainly disrupt the repression on facultative heterochromatin which retains the potential to convert to euchromatin.

DISCUSSION

Here we report a neurodevelopmental syndrome in two families secondary to *de novo* *CBX1* missense mutations. The two variants, p.(Trp52Leu) and p.(Asn57Asp) show high consistency of *in silico* score interpretation consistent with an increased chance of pathogenicity with VarCards combined scores of 19/21 and 19/23 respectively (Table 2). They are also not present in any sub-populations in gnomAD therefore representing rare and more likely damaging variants. The two variants are clustered in the CD, which is known to recognize methylated H3K9 histones[14]. As predicted by CD mediation of HP1β-chromatin interactions, the HP1β mutations identified in these two families abolished HP1β-chromatin interactions, resulting in global transcriptional alterations. These *CBX1* variants have therefore confirmed the significance of chromodomain-methylated H3K9 interactions in chromatin organization and transcriptional regulation. The pLI score in gnomAD (v.2.1.1) [19] is 0.98 consistent with fewer observed than expected missense mutations in *CBX1* (expected 99.3 and observed 28). These data are consistent with HP1β playing essential roles in human development, particularly in the brain. Previously, a mouse model with a heterozygous *Cbx1* mutation was shown to have abnormal neocortical columnar organization, providing confirmatory evidence of the unique role of *Cbx1* in neuronal development [20]. Our data further support the hypothesis that HP1β plays a unique non-redundant role in brain development. These findings represents the first genetic disorder due to germline mutations in genes encoding HP1 proteins. We propose the name "*CBX1*-related syndrome" to denote this condition.

As a mechanism of this developmental syndrome, our data supports the model that *CBX1* mutations cause reduced chromatin binding of HP1β, then, resulting in upregulation of genes residing in the facultative heterochromatin regions (Fig 5). This notion is supported by 1) overexpression of H3K27me3 marked genes (Fig 4), and 2) reduced mutant HP1β binding to H3K27me3 marked histones (Fig 3). A recent study of a *Neurospora crassa* model, HP1 depletion altered distributions H3K27me3, suggesting the contribution of HP1 in facultative heterochromatin function[4]. Therefore, reduced HP1β chromatin binding may influence establishment of

H3K27me3-marked facultative heterochromatin. Meanwhile, our ATAC-seq epigenome evaluation indicates that constitutive heterochromatin organization is not majorly affected by *CBX1* mutations. Collectively, these findings indicate that facultative heterochromatin is more sensitive to the reduced HP1β chromatin binding, compared to the constitutive heterochromatin.

Although HP1β is known as an essential for the formation of compacted heterochromatin, direct roles of HP1β in transcriptional regulation outside of heterochromatin have been described [21,22]. Our data are consistent with the transcriptional regulatory role of HP1β being particularly sensitive for the reduction of chromatin bound HP1β. In euchromatin/facultative heterochromatin, the binding of HP1 proteins is transient and characterized by rapid turnover within 5 seconds[13]. On the other hand, in constitutive heterochromatin, its turn over time is about 60 seconds[13]. Hence, mutant HP1β associated with this *CBX1*-related syndrome may affect the rapid turnover commonly seen in the euchromatin/facultative heterochromatin.

While our data showed the major effects of reduced HP1β chromatin binding in transcriptional regulation, the ATAC-seq data suggests that constitutive heterochromatin organization is not disrupted by the reduction of HP1β chromatin binding. There are at least two possible explanations to the lack of constitutive heterochromatin dysfunction in this diagnosis. First is that reduced HP1β constitutive heterochromatin binding is sufficient enough to maintain constitutive heterochromatin organization. Second is that lack of constitutive heterochromatin bindings of HP1β is compensated by the presence of HP1α and HP1γ. Further investigations are warranted to understand the role of HP1 in heterochromatin regulation.

Recently, HP1 was found to comprise a subunit of the ChAHP complex, which binds to promoter regions throughout the genome repressing gene expression[23]. ChAHP, comprised of CHD4 and ADNP, and mutations in the ChAHP complex components cause similar neurodevelopmental syndromes to those seen in individuals with *CBX1* mutations[24; 25]. Therefore, disruption of similar molecular pathway may underlies neurocognitive impairment seen in individuals with *CBX1*, *CHD4*, and *ADNP* mutations.

Here we described a novel neurodevelopmental syndrome due to missense mutations in CBX1, encoding HP1β. Our identification of *CBX1*-related syndrome represents key evidence which supports the importance of HP1β-chromatin binding in human development. Further studies to elucidate the molecular pathogenesis of *CBX1*-related syndrome will uncover the significance of HP1-mediated heterochromatin organization in human development.

ACKNOWLEDGEMENTS

We thank subjects with *CBX1* mutations and their families who participated in this research study. The authors appreciate the GeneMatcher website for facilitating the collaboration.

FUNDING

KI is supported by the Children's Hospital of Philadelphia Research Institute Institutional Developmental Fund, the Children's Hospital of Philadelphia Foerderer award, the Children's Hospital of Philadelphia Pathology Support, and the Children's Hospital of Philadelphia Roberts Collaborative. S.F.A.G. is supported by the Daniel B. Burke Endowed Chair for Diabetes Research and R01 HG010067.

REFERENCES

1. Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature 410, 120-124.
2. Machida, S., Takizawa, Y., Ishimaru, M., Sugita, Y., Sekine, S., Nakayama, J.I., Wolf, M., and Kurumizaka, H. (2018). Structural Basis of Heterochromatin Formation by Human HP1. Mol Cell 69, 385-397 e388.
3. Bosch-Presegue, L., Raurell-Vila, H., Thackray, J.K., Gonzalez, J., Casal, C., Kane-Goldsmith, N., Vizoso, M., Brown, J.P., Gomez, A., Ausio, J., et al. (2017). Mammalian HP1 Isoforms Have Specific Roles in Heterochromatin Structure and Organization. Cell Rep 21, 2048-2057.
4. Jamieson, K., Wiles, E.T., McNaught, K.J., Sidoli, S., Leggett, N., Shao, Y., Garcia, B.A., and Selker, E.U. (2016). Loss of HP1 causes depletion of H3K27me3 from facultative heterochromatin and gain of H3K27me2 at constitutive heterochromatin. Genome Res 26, 97-107.
5. Baker, S.W., Murrell, J.R., Nesbitt, A.I., Pechter, K.B., Balciuniene, J., Zhao, X., Yu, Z., Denenberg, E.H., DeChene, E.T., Wilkens, A.B., et al. (2019). Automated Clinical Exome Reanalysis Reveals Novel Diagnoses. J Mol Diagn 21, 38-48.
6. Gibson, K.M., Nesbitt, A., Cao, K., Yu, Z., Denenberg, E., DeChene, E., Guan, Q., Bhoj, E., Zhou, X., Zhang, B., et al. (2018). Novel findings with reassessment of exome data: implications for validation testing and interpretation of genomic data. Genet Med 20, 329-336.
7. Wu, C., Devkota, B., Evans, P., Zhao, X., Baker, S.W., Niazi, R., Cao, K., Gonzalez, M.A., Jayaraman, P., Conlin, L.K., et al. (2019). Rapid and accurate interpretation of clinical exomes using Phenoxome: a computational phenotype-driven approach. Eur J Hum Genet 27, 612-620.
8. Ye, Q., Callebaut, I., Pezhman, A., Courvalin, J.C., and Worman, H.J. (1997). Domain-specific interactions of human HP1-type chromodomain proteins and inner nuclear membrane protein LBR. J Biol Chem 272, 14983-14989.
9. Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat 36, 928-930.
10. Ricke, R.M., and Bielinsky, A.K. (2005). Easy detection of chromatin binding proteins by the Histone Association Assay. Biol Proced Online 7, 60-69.
11. Su, C., Johnson, M.E., Torres, A., Thomas, R.M., Manduchi, E., Sharma, P., Mehra, P., Le Coz, C., Leonard, M.E., Lu, S., et al. (2020). Mapping effector genes at lupus GWAS loci using promoter Capture-C in follicular helper T cells. Nat Commun 11, 3294.
12. Kinjo, S., Monma, N., Misu, S., Kitamura, N., Imoto, J., Yoshitake, K., Gojobori, T., and Ikeo, K. (2018). Maser: one-stop platform for NGS big data from analysis to visualization. Database (Oxford) 2018.
13. Cheutin, T., McNairn, A.J., Jenuwein, T., Gilbert, D.M., Singh, P.B., and Misteli, T. (2003). Maintenance of stable heterochromatin domains by dynamic HP1 binding. Science 299, 721-725.
14. Nielsen, P.R., Nietlispach, D., Mott, H.R., Callaghan, J., Bannister, A., Kouzarides, T., Murzin, A.G., Murzina, N.V., and Laue, E.D. (2002). Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. Nature 416, 103-107.
15. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14, 128.

16. Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K., et al. (2018). The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res 46, D794-D801.
17. Xu, S., Grullon, S., Ge, K., and Peng, W. (2014). Spatial clustering for identification of ChIP-enriched regions (SICER) to map regions of histone methylation patterns in embryonic stem cells. Methods Mol Biol 1150, 97-111.
18. Zhu, L.J., Gazin, C., Lawson, N.D., Pages, H., Lin, S.M., Lapointe, D.S., and Green, M.R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics 11, 237.
19. Karczewski, K.F., LC; Tiao, G; Cummings, BB; Alföldi,J; Wang, Q; Collins, RL; Laricchia, KM; Ganna, A; Birnbaum, DP; Gauthier, LD; Brand, H; Solomonson, M; Watts, NA; Rhodes, D; Singer-Berk, M; England, EM; Seaby, EG; Kosmicki, JA; Walters, RK; Tashman, K; Farjoun, Y; Banks, E; Poterba, T; Wang, A; Seed, C; Whiffin, N; Chong, JX; Samocha, KE; Pierce-Hoffman, E; Zappala, Z; O'Donnell-Luria, AH; Minikel, EV; Weisburd, B; Lek, M; Ware, JS; Vittal, C; Armean, IM; Bergelson, L; Cibulskis, K; Connolly, KM; Covarrubias, M; Donnelly, S; Ferriera, S; Gabriel, S; Gentry, J; Gupta, N; Jeandet, T; Kaplan, D; Llanwarne, C; Munshi, R; Novod, S; Petrillo, N; Roazen, D; Ruano-Rubio, V; Saltzman, A; Schleicher, M; Soto, S; Tibbetts, K; Tolonen, C; Wade, G; Talkowski, ME; The Genome Aggregation Database Consortium; Neale, BM; Daly, MJ; MacArthur, DG. (2019). Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv.
20. Aucott, R., Bullwinkel, J., Yu, Y., Shi, W., Billur, M., Brown, J.P., Menzel, U., Kioussis, D., Wang, G., Reisert, I., et al. (2008). HP1-beta is required for development of the cerebral neocortex and neuromuscular junctions. J Cell Biol 183, 597-606.
21. Ayyanathan, K., Lechner, M.S., Bell, P., Maul, G.G., Schultz, D.C., Yamada, Y., Tanaka, K., Torigoe, K., and Rauscher, F.J., 3rd. (2003). Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. Genes Dev 17, 1855-1869.
22. Cryderman, D.E., Grade, S.K., Li, Y., Fanti, L., Pimpinelli, S., and Wallrath, L.L. (2005). Role of Drosophila HP1 in euchromatic gene expression. Dev Dyn 232, 767-774.
23. Ostapcuk, V., Mohn, F., Carl, S.H., Basters, A., Hess, D., Iesmantavicius, V., Lampersberger, L., Flemr, M., Pandey, A., Thoma, N.H., et al. (2018). Activity-dependent neuroprotective protein recruits HP1 and CHD4 to control lineage-specifying genes. Nature 557, 739-743.
24. Helsmoortel, C., Vulto-van Silfhout, A.T., Coe, B.P., Vandeweyer, G., Rooms, L., van den Ende, J., Schuurs-Hoeijmakers, J.H., Marcelis, C.L., Willemsen, M.H., Vissers, L.E., et al. (2014). A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. Nat Genet 46, 380-384.
25. Weiss, K., Terhal, P.A., Cohen, L., Bruccoleri, M., Irving, M., Martinez, A.F., Rosenfeld, J.A., Machol, K., Yang, Y., Liu, P., et al. (2016). De Novo Mutations in CHD4, an ATP-Dependent Chromatin Remodeler Gene, Cause an Intellectual Disability Syndrome with Distinctive Dysmorphisms. Am J Hum Genet 99, 934-941.

Figure Legends:

Figure 1: Mutational spectrum of *CBX1*. Protein structure of HP1β. Heterozygous missense mutations are identified in the chromodomain (CD) of HP1β, where its amino acid sequence is highly conserved throughout evolution. CSD: chromoshadow domain.

Figure 2: Mislocazation of mutant HP1β. Mutant HP1β demonstrates diffuse HP1β signal within nucleus, although WT HP1β signals co-localize with chromocenter detected by DAPI staining. In all the IF experiments, at least two biological duplicates were performed. Bar is 5 um.

Figure 3: Reduced HP1β binding to repressive chromatin marked histones. Reduced H3K9me3, H4K20me3 and H3K27me3 binding of mutant HP1β. ChIP was performed after 48 hours of FLAG-CBX1 cDNA overexpression (wild type and N57D mutant). Although equal amount of FLAG-HP1β is expressed between wild type and N57D mutant, less N57D mutant was identified in H3K9me3/H4K20me3/H3K27me3 marked chromatin fraction compared to control. Biological duplicates revealed the consistent result. Similar result was obtained with p.W52L mutant for H3K9me3.

Figure 4: Transcriptomic and epigenomic alterations of *CBX1* mutant sample.

(a) Volcano plot demonstrating the differentially expressed genes of the *CBX1* mutant LCL.

(b) Hierarchal clustering and heat map demonstrated the differentially expressed genes of the *CBX1* mutant LCL. Differentially expressed genes in the LCLs with *CBX1* mutations defined by FDR<0.05 are depicted. Red indicates higher gene expression, and blue indicates lower gene expression. The distances between two clusters are depicted on the top and left of the heat map.

(c) Box plot demonstrates the gene expression fold difference between patient and control samples. Y-axis represents log 2 fold gene expression differences of the *CBX1* mutant sample, compared to the control. Overexpression of H3K27me3 marked genes are identified. Student t-test was used to evaluate the distribution differences between two groups.

(d) MA plot comparing the chromatin accessibility between the *CBX1* mutant and control LCL. *CBX1* mutant sample demonstrates slightly more accessibility than control.

(e) Box plot demonstrates correlation between RNA-seq and ATAC- seq. Y-axis represents log 2 fold gene expression differences of the *CBX1* mutant sample, compared to the control. Genes in the open differentially accessible regions (DARs) in the CBX1-related syndrome demonstrated upregulation of gene expression. Three biological replicates were used. Wilcox rank sum test revealed the p-value of $3.94e^{-18}$.

(f) Representative loci of DARs in the *CBX1*-related syndrome. Arrow indicates the transcription start site of *CADM2* gene, and increased chromatin accessibility was identified in the patient sample.

(g) Enrichment of DARs at histone mark binding region. The enrichment test was performed using fisher exact test. The color scale represents the odd ratio and value within cell indicates the p value.

Figure 5: Model of CBX1-related syndrome. Reduced HP1β chromatin binding cause increased chromatin accessibility, resulting in overexpression of facultative heterochromatin genes.

Table 1: Clinical features of two individuals with *CBX1* mutations.

	Patient 1	Patient 2
Gender	female	male
Age at initial presentation (years)	1.5	5
Age at last visit (years)	5 years 8 months	5 years
OFC at last visit	75th percentile	75th percentile
height at last visit	85th percentile	50-75th percentile
weight at last visit	>99th percentile	50-75th percentile
Feeding difficulties	Present as neonate; attributed to anklyoglossia s/p frenotomy and poor maternal milk supply	Colicky baby
Growth	Obesity (weight >99th percentile, BMI >99th percentile)	50th-75th percentiles for height and weight
Overall Development	Global developmental delay	Global developmental delay
Motor abilities	Hypotonia, gross motor delay (sitting at 10m, pulling to stand at 14m, and walking at 23m); still clumsy	Cardiorespiratory arrest with RSV bronchiolitis aged 6 weeks thought to have caused developmental delay initially
Cognitive abilities	Mild-to-moderate intellectual disability	Mild to borderline disability probable
Behavioral Issues	Autism spectrum disorder, obsessive-compulsive behaviors	Autistic spectrum disorder
Verbal abilities	Expressive language delay, non-verbal, uses AAC to communicate	Expressive language delay
MRI	normal brain MRI at 3y	Enlarged extra axial fluid spaces
Muscle tone	Hypotonia	Hypermobility and hypotonia
Ear Problems	Recurrent/chronic otitis media s/p bilateral myringotomy tubes	No
Craniofacial Abnormalities	Frontal bossing, midface hypoplasia, round-shaped face, epicanthal folds, upslanting palpebral fissures, flat nasal bridge with short nose, tented upper lip with thinning of lateral upper lip	Arched eybrows, broad forehead, tall skull with flat occiput, small overfolded ears.
Genitourinary Issues	No issues.	Incomplete testicular descent orchiopexy

Skeletal Issues	No issues.	Narrow flat feet. Right hip pain for months led to MRI showing right intramedullary femur lesion. Biopsy showed inflammation. No evidence of bacterial infection and subsequent slow progression suggests chronic relapsing multifocal osteomyelitis

Table 2: Molecular Summary of *CBX1* variants.

In silico scores	Patient 1	Patient 2
VarCards D:A algorithm	19:23	21:23
ClinPred	0.9968995 / Pathogenic	0.99984 / Pathogenic
Splice site predictions (ALAMUT)		
CADD (score/prediction)	27.8 / Damaging	35 / Damaging
Polyphen2 HVAR (score/prediction)	0.989 / probably_damaging	0.992 / Probably damaging
PROVEAN (score/prediction)	-4.81 / damaging	-12.52 / Damaging
SIFT (score/prediction)	0.002 / damaging	0.00 / Damaging
REVEL (score/prediction)	0.702 / Damaging	0.918 / Damaging
LRT (score/prediction)	0 / Unknown	0 / Unknown
Mutation Taster (score/prediction)	1 / Disease causing	1 / Disease causing
Mutation Assessor (score/prediction)	2.765 / Medium	3.9 / High
FATHMM (score/prediction)	-0.66 / Tolerable	-3.49 / Damaging
VEST3 (score/prediction)	0.491 / Tolerable	0.936 / Damaging
MetaSVM (score/prediction)	0.129 / Damaging	1.082 / Damaging
MetaLR (score/prediction)	0.537 / Damaging	0.948 / Damaging
M-CAP (score/prediction)	0.112 / Damaging	0.503 / Damaging
DANN (score/prediction)	0.998 / Damaging	0.985 / Tolerable
FATHMM-MKL (score/prediction)	0.989 / Damaging	0.992 / Damaging
Eigen (score/prediction)	0.826 / Damaging	0.997 / Damaging

GenoCanyon (score/prediction)	1 / Damaging	1 / Damaging
ReVe (score/prediction)	0.772 / Damaging	0.980 / Damaging
RVIS v2		
ExAC Missense constraint	z = 2.70	z = 2.70
ExAC LoF intolerance	pLI = 0.95	pLI = 0.95
fitCons (score/prediction)	0.707 / Damaging	0.707 / Damaging
GERP (score/prediction)	5.85 / Conserved	5.85 / Conserved
phyloP (score/prediction)	7.943 / Conserved	7.832 / Conerved
phastCons (score/prediction)	1 / Conserved	1 / Conserved
SiPhy (score/prediction)	15.208 / Conserved	18.925 / Conserved
gnomAD exome AFR	-	-
gnomAD exome AMR	-	-
gnomAD exome ASJ	-	-
gnomAD exome EAS	-	-
gnomAD exome FIN	-	-
gnomAD exome NFE	-	-
gnomAD exome OTH	-	-
gnomAD exome SAS	-	-
gnomAD exome TOTAL (number/frequency)	-	-
gnomAD genome AFR	-	-
gnomAD genome AMR	-	-
gnomAD genome ASJ	-	-
gnomAD genome EAS	-	-
gnomAD genome FIN	-	-
gnomAD genome NFE	-	-
gnomAD genome OTH	-	-

gnomAD genome TOTAL (number/frequency)	-	-

Supplementary Figure 1: Reduced p.W52L HP1β binding to H3K9me3 marked histones. Reduced H3K9me3 binding of p.W52L mutant HP1β. ChIP was performed after 48 hours of FLAG-CBX1 cDNA overexpression (wild type and W52L mutant). Although equal amount of FLAG-HP1β is expressed between wild type and W52L mutant, less W52L mutant was identified in H3K9me3 marked chromatin fraction compared to control. Biological duplicates revealed the consistent result.

Supplementary Table 1: RNA-seq and ATAC-seq statistics.

Supplementary Table 2: Differentially expressed genes of the *CBX1* mutant LCL sample.

Supplementary Table 3: Histone modification enrichment in the *CBX1* mutant LCL differentially expressed genes.

Supplementary Table 4: Differentially accessible genes identified by ATAC-seq.

Supplementary Table 5: Chromatin accessibility evaluation of heterochromatin repeats.

Fig 1

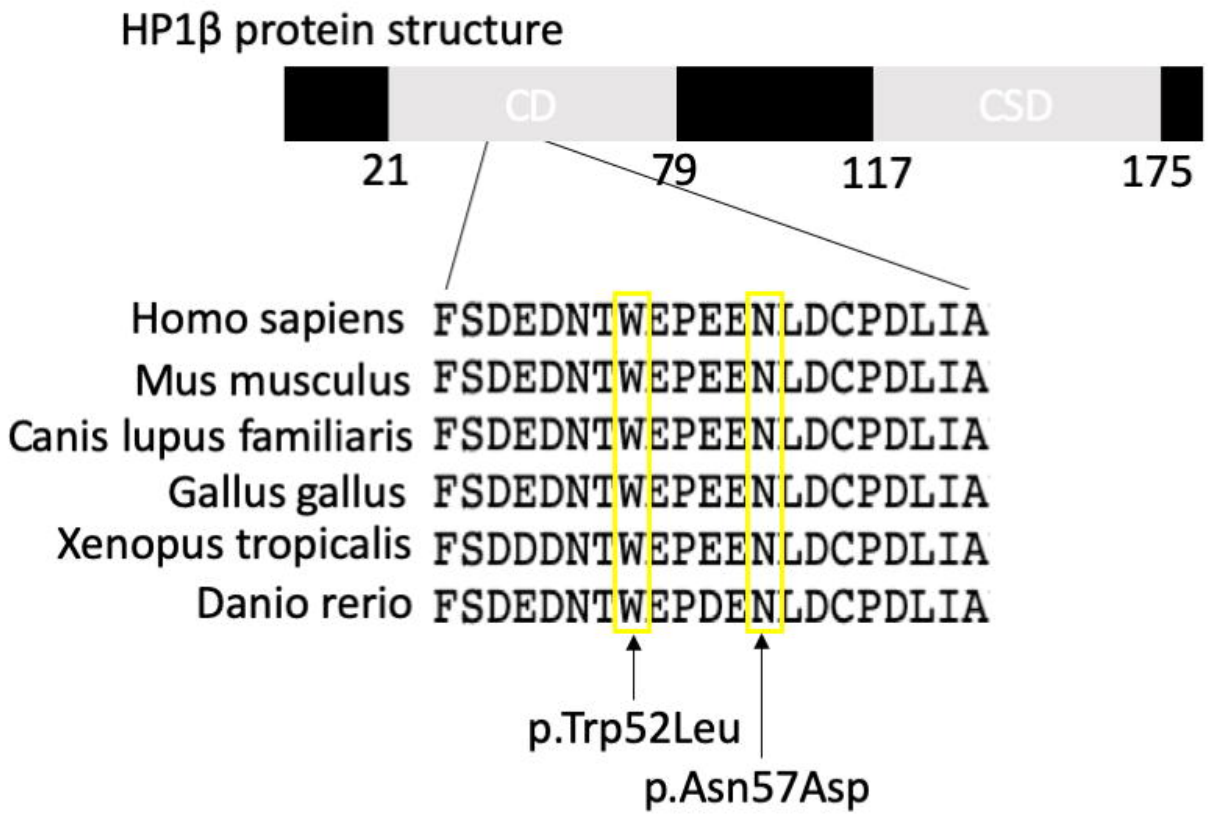

HP1β protein structure
CD
CSD
21
79
117
175
Homo sapiens FSDEDNTWEPEENLDCPDLIA
Mus musculus FSDEDNTWEPEENLDCPDLIA
Canis lupus familiaris FSDEDNTWEPEENLDCPDLIA
Gallus gallus FSDEDNTWEPEENLDCPDLIA
Xenopus tropicalis FSDDDNTWEPEENLDCPDLIA
Danio rerio FSDEDNTWEPDENLDCPDLIA
p.Trp52Leu
p.Asn57Asp

Fig 2

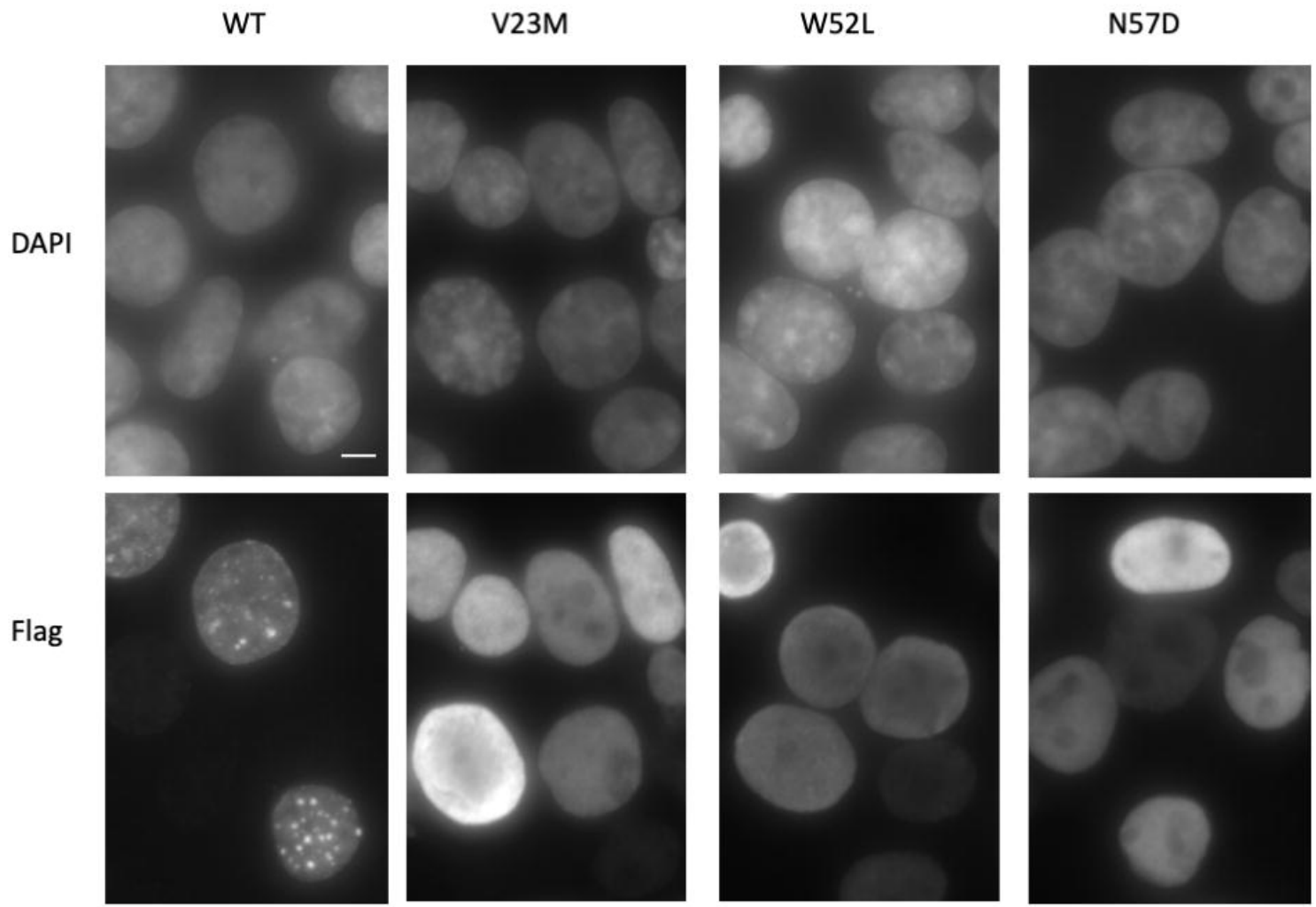
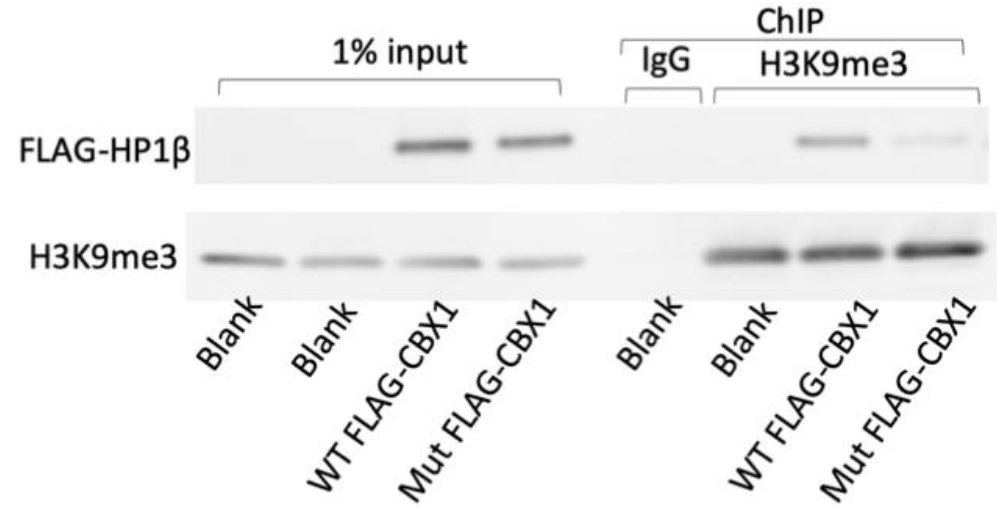
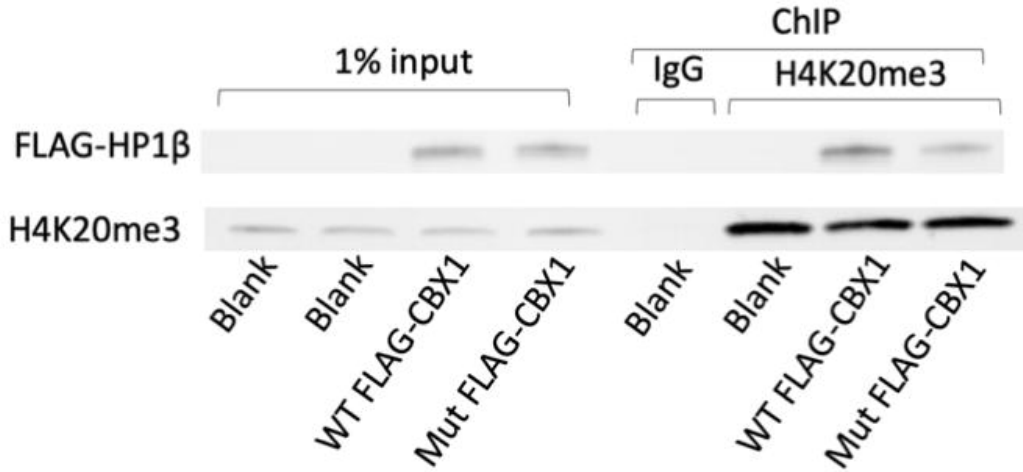


Fig 3


1% input
ChIP
IgG
H3K9me3
FLAG-HP1β
H3K9me3
Blank
Blank
WT FLAG-CBX1
Mut FLAG-CBX1
Blank
Blank
WT FLAG-CBX1
Mut FLAG-CBX1


1% input
ChIP
IgG
H4K20me3
FLAG-HP1β
H4K20me3
Blank
Blank
WT FLAG-CBX1
Mut FLAG-CBX1
Blank
Blank
WT FLAG-CBX1
Mut FLAG-CBX1

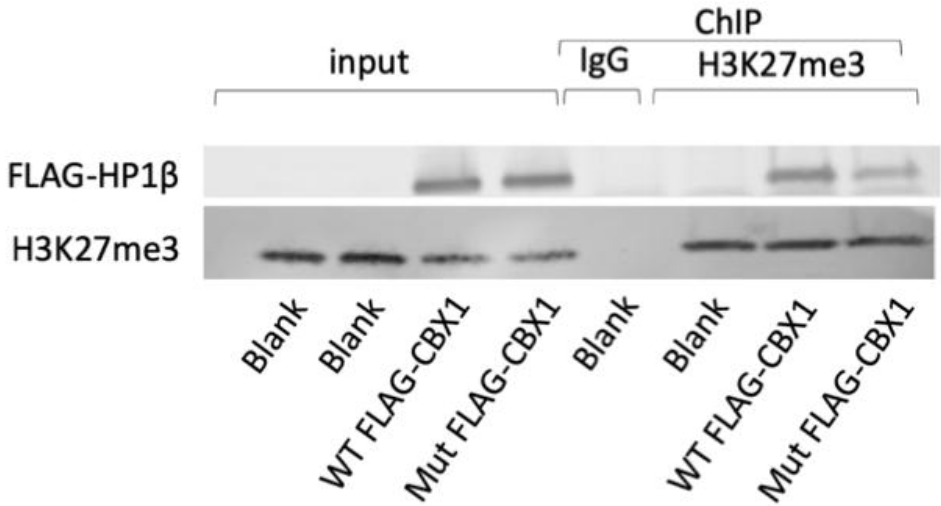

input
ChIP
IgG
H3K27me3
FLAG-HP1β
H3K27me3
Blank
Blank
WT FLAG-CBX1
Mut FLAG-CBX1
Blank
Blank
WT FLAG-CBX1
Mut FLAG-CBX1

Fig 4

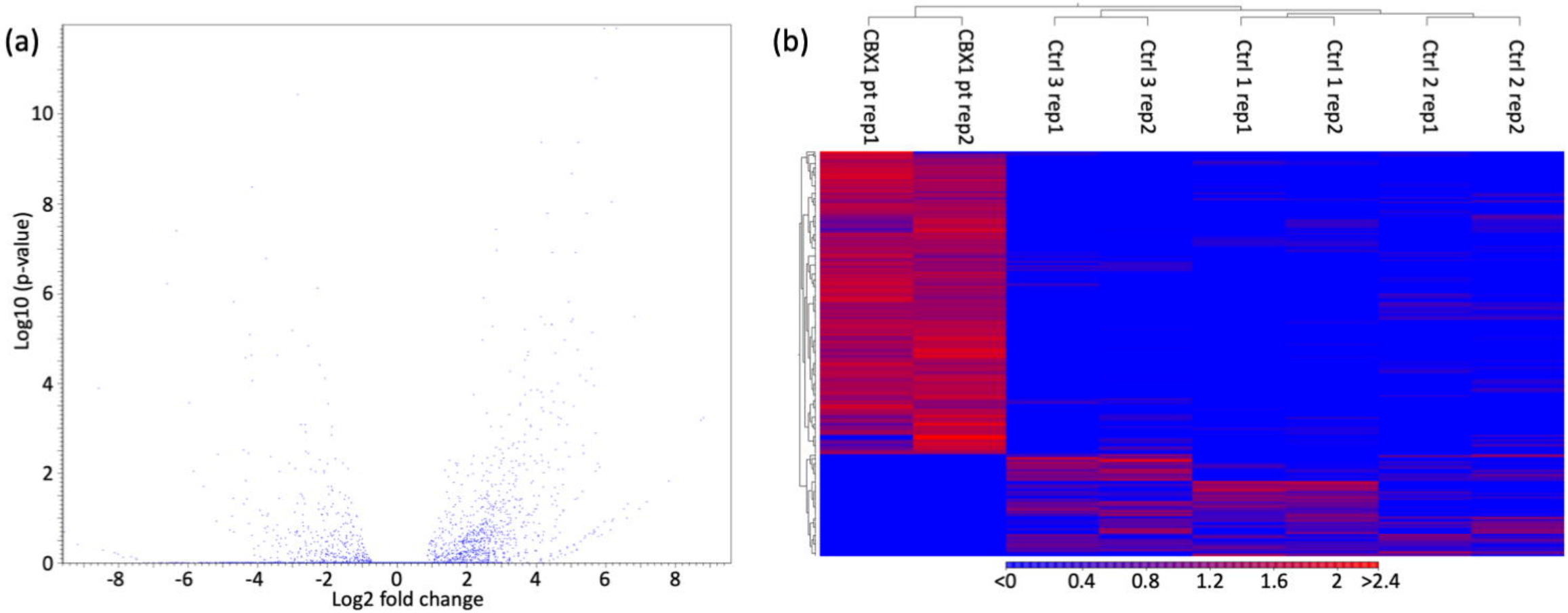
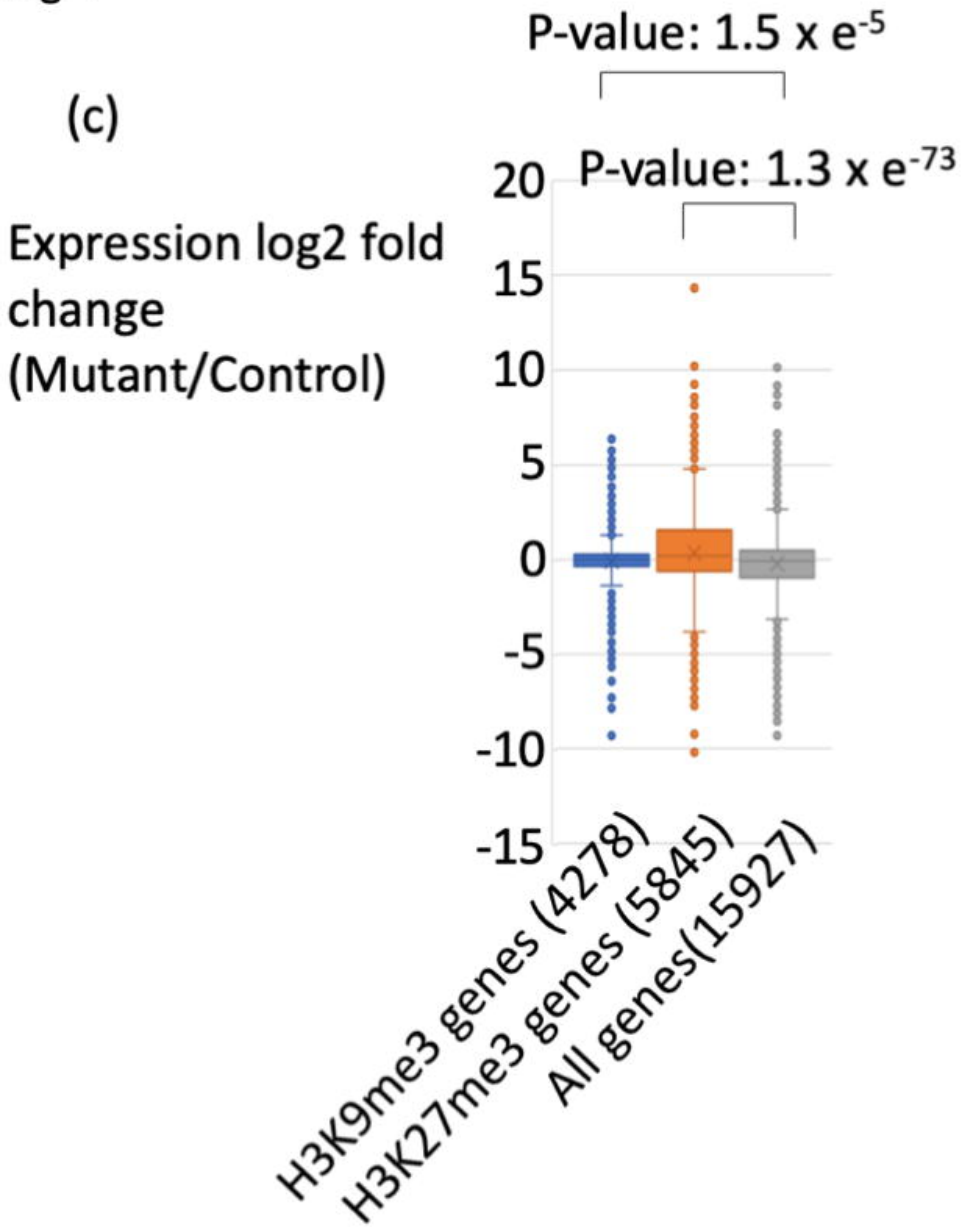

(a)
Log10 (p-value)
Log2 fold change
(b)
CBX1 pt rep1
CBX1 pt rep2
Ctrl 3 rep1
Ctrl 3 rep2
Ctrl 1 rep1
Ctrl 1 rep2
Ctrl 2 rep1
Ctrl 2 rep2
<0
0.4
0.8
1.2
1.6
2
>2.4

Fig 4

(c)


P-value: 1.5 x e-5
P-value: 1.3 x e-73
Expression log2 fold change (Mutant/Control)
20
15
10
5
0
-5
-10
-15
H3K9me3 genes (4278)
H3K27me3 genes (5845)
All genes(15927)

(d)

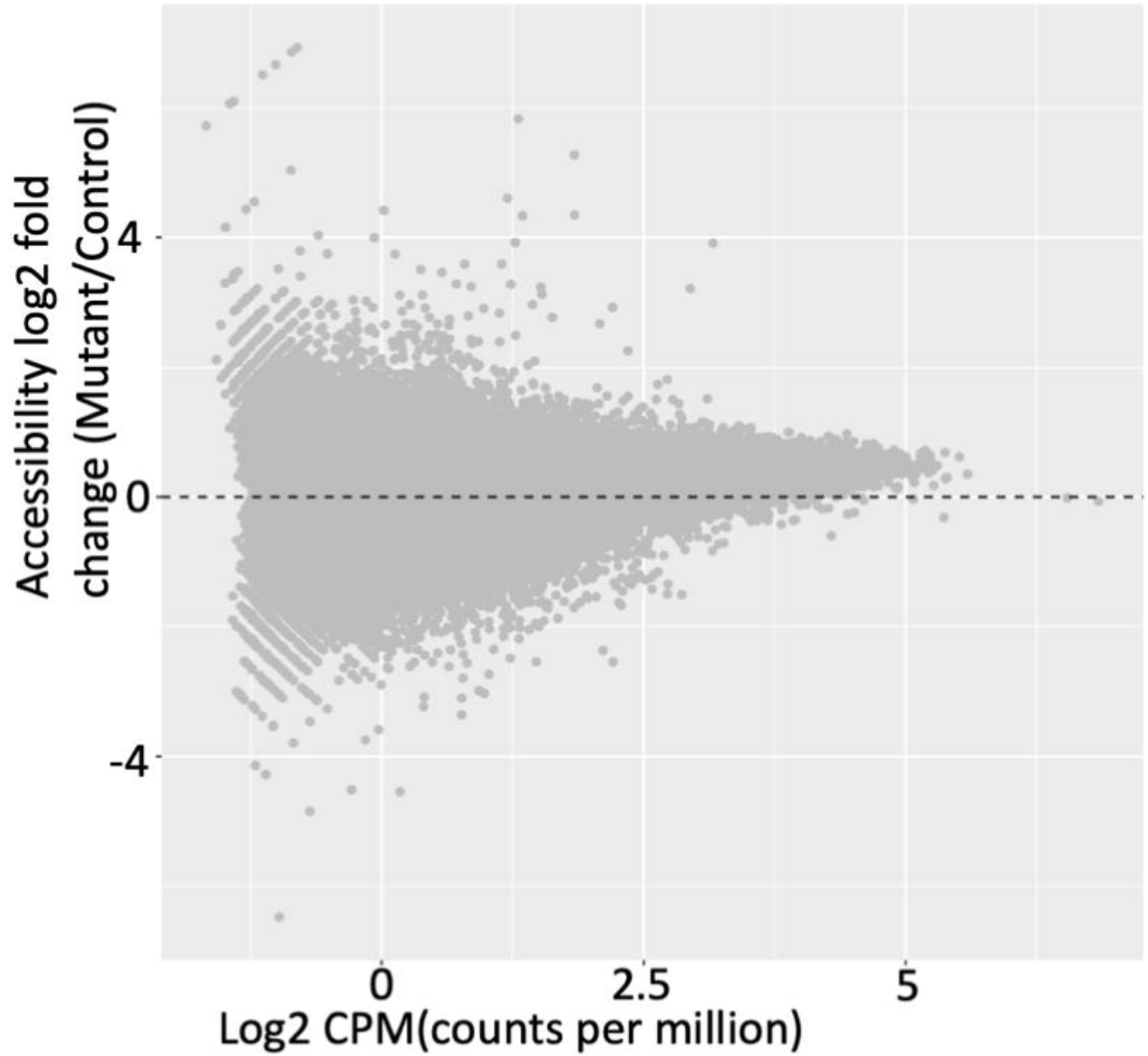
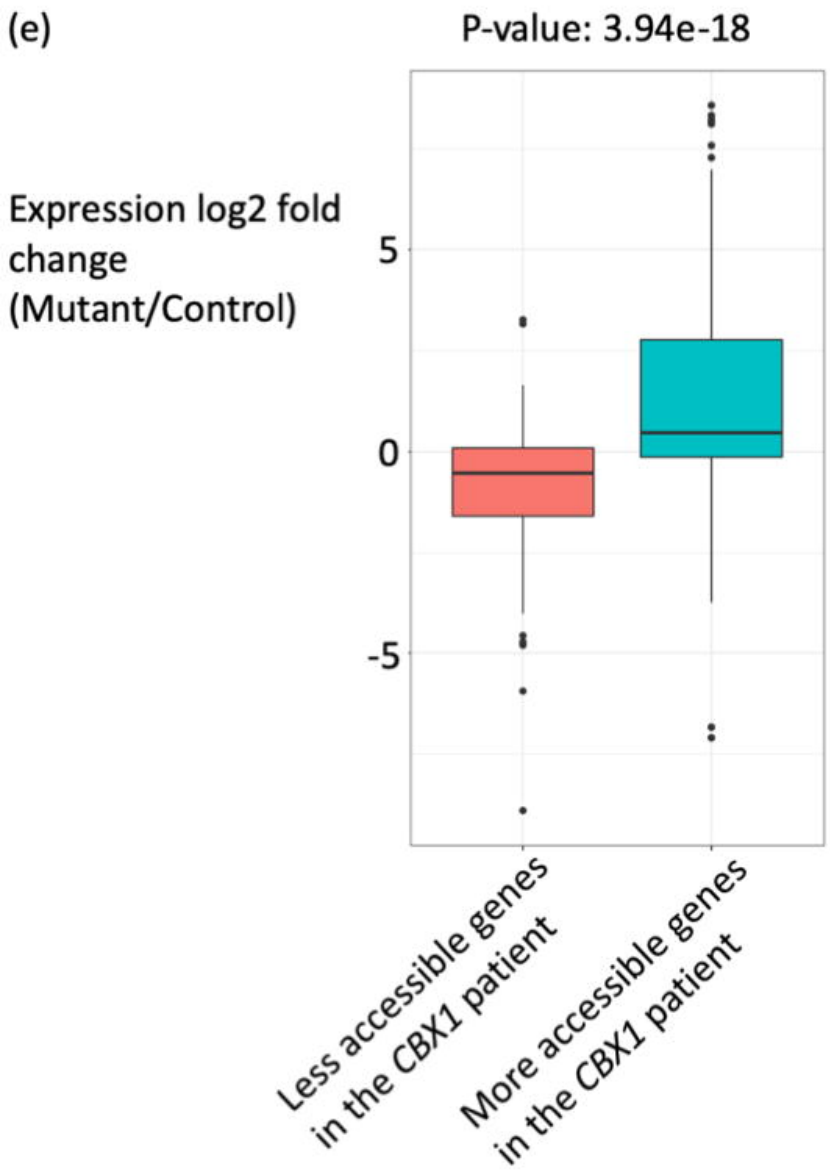
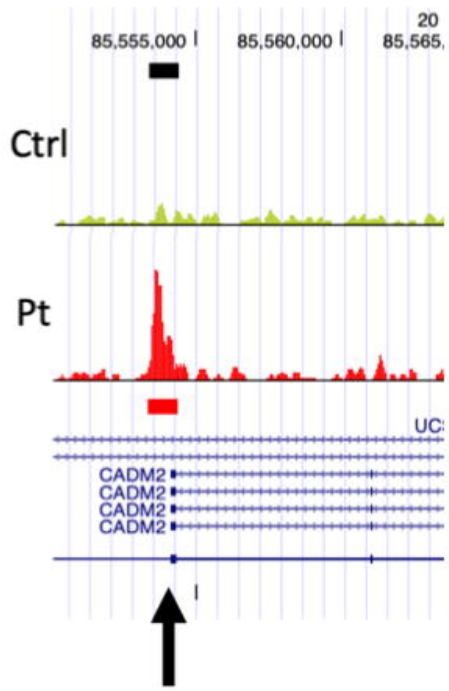

Accessibility log2 fold change (Mutant/Control)
4
0
-4
0
2.5
5
Log2 CPM(counts per million)

Fig 4


(e)
P-value: 3.94e-18
Expression log2 fold change (Mutant/Control)
5
0
-5
Less accessible genes in the CBX1 patient
More accessible genes in the CBX1 patient


(f)
85,555,000
85,560,000
85,565,
Ctrl
Pt
CADM2
CADM2
CADM2
CADM2

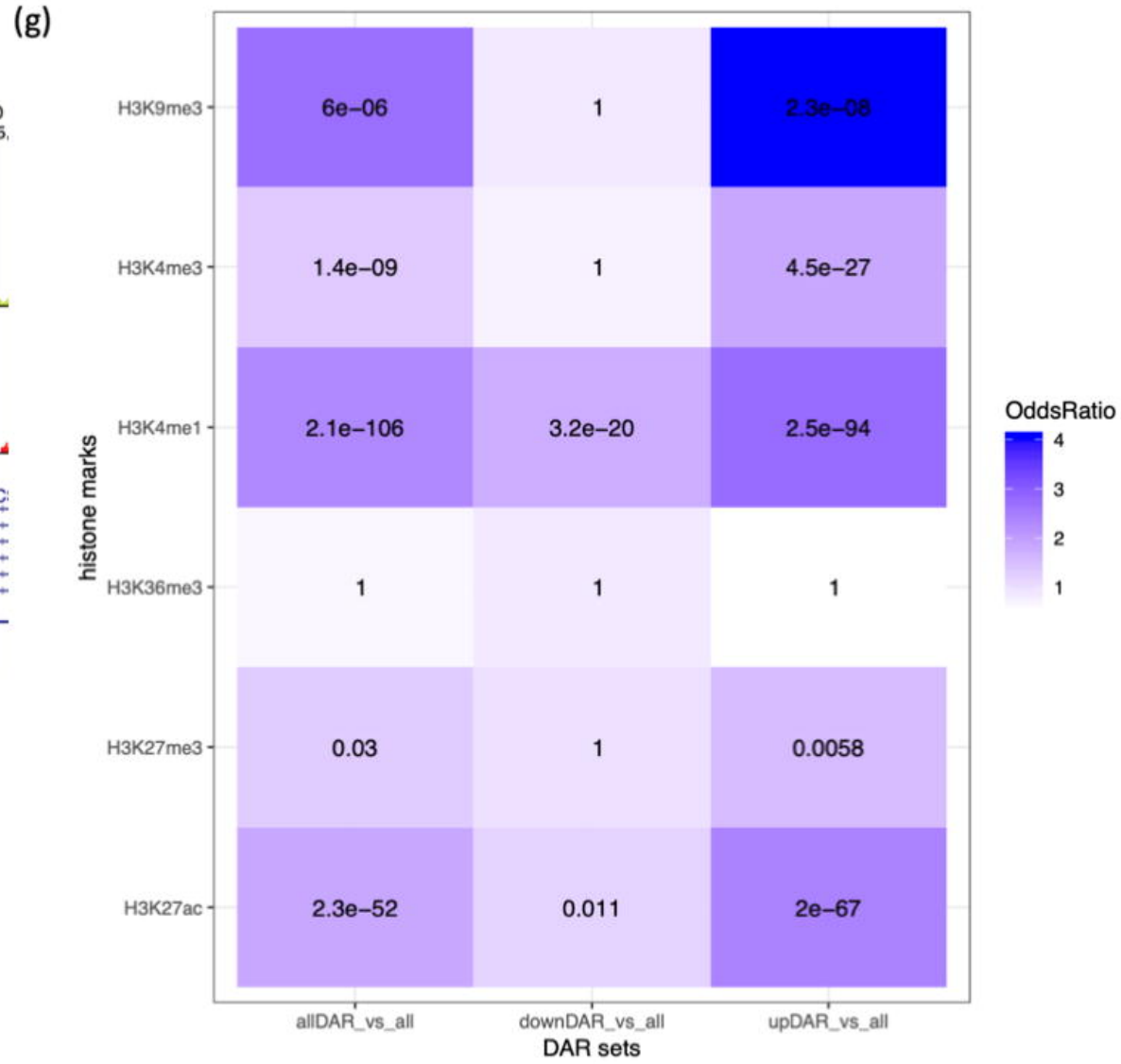
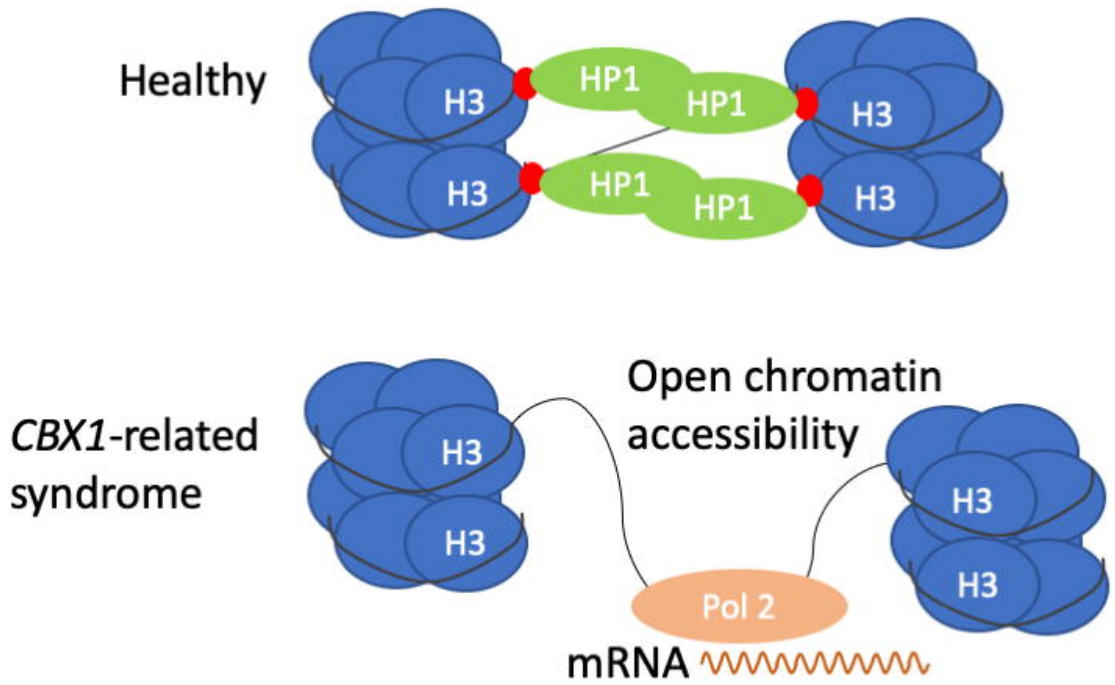

(g)
H3K9me3
H3K4me3
H3K4me1
H3K36me3
H3K27me3
H3K27ac
histone marks
6e−06
1
2.3e−08
1.4e−09
1
4.5e−27
2.1e−106
3.2e−20
2.5e−94
1
1
1
0.03
1
0.0058
2.3e−52
0.011
2e−67
allDAR_vs_all
downDAR_vs_all
upDAR_vs_all
DAR sets
OddsRatio
4
3
2
1

Fig 5


Healthy
H3
HP1
HP1
H3
H3
HP1
HP1
H3
CBX1-related syndrome
H3
H3
Open chromatin accessibility
H3
H3
Pol 2
mRNA