***CBX1* variants cause a neurodevelopmental syndrome due to facultative heterochromatin dysfunction**

Aiko Iwata-Otsubo[1], Kerith-Rae Dias[2,3], Chun Su[4], Suzanna EL Temple[2,5], Ying Zhu[2], Sarah K Fiordaliso[1], Alyssa L Ritter[1], Samuel W Baker[6], Yukiko Kuroda[1], Beth A Keena[1], Struan F.A. Grant[1,4,7,8] , Gopinath Musuwadi Subramanian[9], Elaine H Zackai[1,7], Matt Edwards[10,11], Carey-Anne Evans[2,3], Matthew C Dulik[6], Michael F Buckley[2], Tony Roscioli[2,3,5*], Kosuke Izumi[1,7,12*]

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

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

[5] Centre for Clinical Genetics, Sydney Children's Hospital, Randwick, NSW, Australia[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] Department of Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA.

[9] Paediatric Neurology Unit, John Hunter Children's Hospital, New South Wales, Australia

[10] Hunter Genetics, Newcastle, New South Wales, Australia

[11] University of Western Sydney School of Medicine, New South Wales. Australia

[12] Laboratory of Rare Disease Research, Institute for Quantitative Biosciences, The University of Tokyo, Tokyo, Japan

[*] These authors contributed equally

Author Contributions

Conceptualization: Aiko Iwata-Otsubo, Tony Roscioli, and Kosuke Izumi

Experiments: Aiko Iwata-Otsubo, Sarah K Fiordaliso, Chun Su, Yukiko Kuroda, and Kosuke Izumi

Data analysis: Aiko Iwata-Otsubo, Chun Su, Struan F.A. Grant, Carey-Anne Evans, Ying Zhu, Suzanna Lindsay-Temple, Michael Buckley, Kerith-Rae Dias, Tony Roscioli, and Kosuke Izumi

Clinical evaluation: Alyssa Ritter, Samuel W Baker, Beth Keena, Elaine Zackai, Matt Edwards, Matthew Dulik, and Kosuke Izumi

Manuscript writing: Aiko Iwata-Otsubo, Chun Su, Tony Roscioli, Kerith-Rae Dias, and Kosuke Izumi

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.

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. TR, MB, K-R D and SELT are supported

through the Sydney Partnership for Health, Education, Research and Endeavour (SPHERE) and with C-A E, the NHMRC Centre for Research Excellence in Neurocognitive Disorders. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

The heterochromatin protein 1 (HP1) family of proteins represents an essential structural component of heterochromatin formation and organization. Although HP1β, which is encoded by the *CBX1* gene, is essential for brain development in mouse, no human disorders involving HP1 proteins have ever been reported. Through exome sequencing, we identified two heterozygous *de novo CBX1* variants in two unrelated individuals with developmental delay, hypotonia and autistic features. Identified variants are in the known functional domain of HP1β, chromodomain, which mediates interaction with chromatin. We examined the effects of the variants using protein structural prediction models and molecular assays. Initial *in silico* analyses of these missense variants predict that they are highly pathogenic and disrupt protein structure for chromatin binding. Subsequent molecular assays confirmed that the identified variants abolished HP1β-chromatin interactions. Transcriptome and epigenome analyses of human patient-derived lymphoblastoid cell lines with RNA-seq and ATAC-seq, respectively, detected global transcriptional and chromatin organizational alterations, particularly in the context of facultative heterochromatin (i.e. reversibly repressed genomic regions), while chromatin organization was unchanged in constitutional heterochromatin. Overall, the genes harboring H3K27me3 were upregulated, while genes harboring H3K9me3 did not reveal transcriptional alterations. Chromatin was globally more open in the *CBX1* mutant sample detected by ATAC-seq. These chromodomain HP1β variants highlight the importance of HP1β chromatin binding particularly in the functional regulation of facultative heterochromatin during neurocognitive development, and confirm the role of *CBX1* in intellectual disability and autism spectrum disorder.

Introduction

The human genome is compartmentalized into two major chromatin structures, euchromatin and heterochromatin. Euchromatin is an open chromatin region associated with active transcription, while heterochromatin is a closed chromatin conformation representing a repressed genomic function. Precise determination of euchromatin/heterochromatin structures is a key transcriptional regulatory mechanism, with an increasing number of neurodevelopmental disorders being linked to disruption of such chromatin organization [1, 2].

Heterochromatin is further divided into constitutive and facultative heterochromatin. Constitutive heterochromatin is a stably repressed chromatin region, while facultative heterochromatin is a temporarily inactivated genomic region. Constitutive heterochromatin is composed of repetitive sequences, and plays important roles in ensuring proper chromosomal segregation [3]. Conversely, facultative heterochromatin is instrumental in cell lineage-specific transcriptional regulation including genes associated with neuronal lineage [4, 5].

The key regulatory molecule of heterochromatin formation and organization is heterochromatin protein 1 (HP1) [6]. HP1 proteins form homo- and heterodimers, and they bind to methylated histone H3K9 and bring two methylated H3K9s into close proximity, resulting in chromatin compaction [7]. HP1 is a predominant mark residing in constitutive heterochromatin, and is mainly known to regulate chromatin organization of constitutive heterochromatin, a stably repressed genomic region [8]. However, recent reports suggest HP1 functioning outside of constitutive heterochromatin such as facultative heterochromatin, as reversibly repressed genomic regions [9].

In humans, there are three HP1 proteins, HP1α, HP1β, and HP1γ, which exhibit a high degree of homology. Despite the critical importance of HP1 proteins in chromatin organization and differentiation regulation, no human disorders have been associated to date with pathogenic variants in the genes encoding HP1 proteins. Here, we report two individuals with a neurodevelopmental syndrome due to *de novo* missense variants in the *CBX1*, which encodes

HP1β. Protein structural prediction models provide strong evidence that the variants are deleterious. Functional analysis confirmed that the *CBX1* variants abolish the interaction between HP1β and histone repressive marks, resulting in upregulation of its target genes within facultative heterochromatin.

Results

Clinical description of individuals with *CBX1* variants

Individual 1 was a 6 year old female who presented for evaluation at 19 months of age due to global developmental delay. She was delivered after an uncomplicated pregnancy to a 29 year old G1P1 mother. She was born at 41 weeks gestational age with a birth weight of 3.6 kg (50th percentile), and length of 55.9 cm (90th percentile). The neonatal period was complicated by poor weight gain, which was attributed to ankyloglossia, treated with frenotomy, and reduced maternal milk supply. Concerns regarding her developmental milestones were first raised at 9 months of age when she was not yet sitting. Her development had been globally delayed, with independent sitting achieved at 10 months of age, pulling to stand at 14 months of age, walking at 23 months of age, and significant expressive speech delay. Family history was non-contributory.

Her medical history included hypotonia, chronic constipation, recurrent otitis media treated with bilateral myringotomy tubes, easy bruising, and transient erythematous patches with dermographism. She experienced rapid weight gain after the age of 12-14 months. At initial evaluation at 19 months of age, her weight was 14.23 kg (99th percentile), height was 84.4 cm (79th percentile), and head circumference was 47.5 cm (50th percentile). A number of dysmorphic features were noted including round-shaped face, midface hypoplasia, epicanthal folds, slightly upslanting palpebral fissures, a flat nasal bridge with a short nose, a tented upper lip with thinning of the lateral upper lip, closely spaced nipples (chest circumference/internipple distance ratio= 0.175, <3rd percentile), short fingers and toes, and bilateral 2-3 toe syndactyly.

At the most recent evaluation at 5 years 8 months of age, her weight was 33 kg (>99th percentile), her height was 118.4 cm (85th percentile), and BMI was 23.54 kg/m2 (>99th percentile).

Her development has remained delayed, but she is making developmental progress. She was diagnosed with autism spectrum disorder at 3.5 years of age. The behavioral phenotype included self-stimulation, sensory seeking behaviors, repetitive hand movements, head banging, and obsessive-compulsive movements. At 6 years old, she uses some sign language and an augmentative communication device to communicate. Her receptive language is much better than her expressive language.

Her diagnostic work-up has included a normal brain MRI, normal female karyotype, normal Fragile X molecular testing, normal Prader-Willi/Angelman methylation analysis, and a normal SNP microarray result. Clinical trio exome sequencing was performed [10-12], and a *de novo* heterozygous single nucleotide variant in *CBX1* was identified, predicted to result in a missense change, NM_001127228.1: c.169A>G, p.(Asn57Asp).

Individual 2 was a 5 year old boy born to non-consanguineous parents of European origins. He was conceived naturally with an uncomplicated pregnancy. He was born via caesarean section at 37 weeks gestation for rising anti-E antibody levels for which he required phototherapy. His birth weight was 3350 g (90th percentile), and length was 51 cm (90th percentile). Occipitofrontal circumference at age 24 months was 51.5 cm (90th-97th percentile). The neonatal period was otherwise unremarkable. Two blood transfusions were required for anemia.

There were first concerns about global developmental delay when he was 24 months old. It was initially thought that this was secondary to a cardiorespiratory arrest after RSV bronchiolitis at 6 weeks of age. He sat independently at 6 months and was walking at 25 months. He was noted to be hypotonic, with hypermobile joints. His first words were at 30 months of age. He was suspected of having autism spectrum disorder at 3 years of age and at 6

years of age he has mild to moderate cognitive impairment and attends a supported class in a mainstream school.

Bilateral orchiopexy was performed at 15 months of age for incomplete testicular descent. He has had chronic relapsing multifocal osteomyelitis with a biopsy of the right hip showing an intramedullary femur inflammatory lesion. The right leg was 1 cm longer than the left, attributed to increased perfusion associated with chronic inflammation in the right hip.

On examination, at 5 years old, his weight was 17 kg (50-70th percentile), his height 103.7 cm (50-70th percentile) and his head circumference 52.8cm (85-97th percentile). He was noted to have arched eyebrows, a broad forehead, a tall skull with flat occiput, small over-folded ears, small scalp lentigines and bilateral narrow flat feet. He also had a 3-4 cm long lentigines on the lateral aspect of the right lower limb.

Brain MRIs at 1 year of age and subsequently at age 2 years showed increased extra axial CSF and mild ventricular dilatation. Audiology was normal. Chromosome microarray detected a maternally inherited chromosome 2q31.3 deletion reported as a variant of uncertain significance and fragile X DNA analysis was normal. At age 4 a Noonan spectrum disorders/RASopathies DNA panel was negative. Clinical trio exome sequencing was performed and identified a *de novo* heterozygous missense variant in *CBX1* (chr17(GRCh37):*g.46153526C>A*: NM_001127228: c.155G>T, p.(Trp52Leu).

Clinical features of these two patients are compared (**Table 1**) and their variant status is summarized (**Table 2**). The identified two variants, p.(Asn57Asp) and p.(Trp52Leu) reveal an *in silico* score interpretation consistent with an increased chance of pathogenicity with VarCards combined scores of 19/23 and 21/23, respectively. They are also not present in any sub-populations within gnomAD [13], therefore, representing rare and likely deleterious variants.

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

	Individual 1	Individual 2
Gender	Female	Male
Age at initial genetics evaluation	1 year 6 months	5 years
Age at last visit	6 years	5 years
OFC at last visit	51.5 cm (80th percentile)(at 4 years 9 months)	52.8 cm (85-97th percentile)
height at last visit	118.4 cm (85th percentile)	103.7 cm (50-70th percentile)
weight at last visit	33 kg (>99th percentile)	17 kg (50-70th percentile)
Feeding difficulties	Present as neonate; attributed to ankyloglossia s/p frenotomy and reduced 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	Autism 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, closely spaced nipples, short fingers and toes, 2-3 syndactyly of toes bilaterally	Arched eyebrows, broad forehead, tall skull with flat occiput, small overfolded ears. Increased extra axial CSF and mild ventricular dilatation
Genitourinary Issues	No issues	Incomplete testicular descent treated by 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	Individual 1	Individual 2
Variant VarCards D:A algorithm	p.(Asn57Asp) 19:23	p.(Trp52Leu) 21:23
ClinPred	0.9968995 / Pathogenic	0.99984 / Pathogenic
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
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

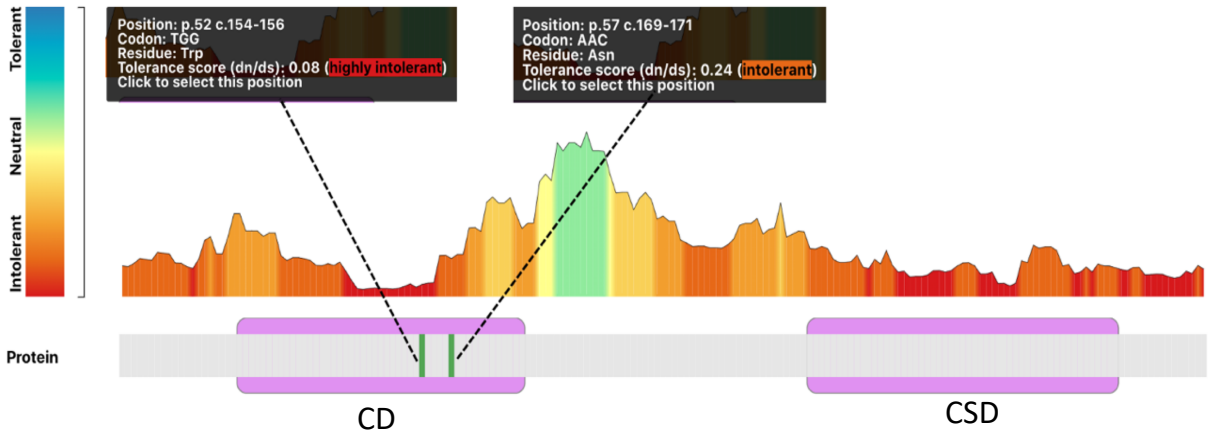
GERP (score/prediction)	5.85 / Conserved	5.85 / Conserved
phyloP (score/prediction)	7.943 / Conserved	7.832 / Conerved
phastCons (score/prediction)	1 / Conserved	1 / Conserved
SiPhy (score/prediction)	15.208 / Conserved	18.925 / Conserved

Molecular features of patient-identified *CBX1* variants

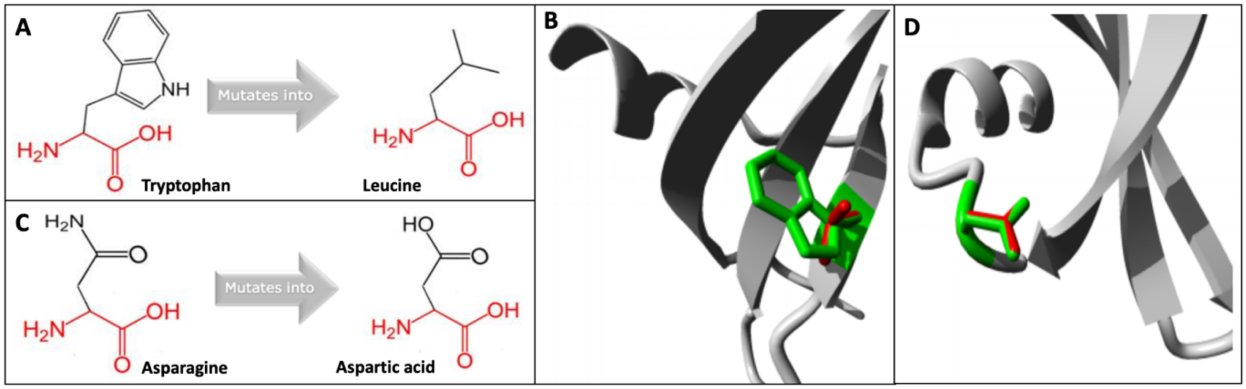
HP1β consists of a chromodomain (CD) at the amino-terminal, a chromoshadow domain (CSD) at the carboxy-terminal connected by a hinge region (HR) [14], and the identified variants in the two individuals reside within the CD (**Fig 1a**). CD directly interacts with methylated H3K9, whereas CSD serves as a dimerization interface with other HP1 proteins [15]. Various *in silico* computational programs predicted these variants to be deleterious (**Table 2**). The W52L variant alters tryptophan to a leucine at position 52. The variant residue is smaller than the wild-type residue due to the loss of tryptophan's large side chain with two fused rings (**Fig 1b**). The residue position is very conserved and is located near a highly conserved position. Neither the variant residue nor another residue type with similar properties have been observed at this position in other homologous sequences, and is therefore predicted to be damaging to the protein. The residue is located on the surface of the domain and may cause possible loss of external interactions. The N57D variant alters the asparagine into an aspartic acid at position 57 (**Fig 1b**), and is predicted to cause a change of residue charge from neutral to negative, which may cause repulsion between the variant residue and neighbouring residues. A variant close to the residues that make up the binding pocket residues could affect the local structure and as a consequence affect this binding site. The variant is located within the CD, which could disrupt this domain and abolish its function. The magnitude of the W residue reflecting the uniqueness of the sequence in homologous proteins indicates that the tryptophan at position 52 is highly conserved and the asparagine at position 57 although less so, is still conserved between species and related proteins (**S1 Fig**). No gnomAD population variation data was observed at

CBX1 position 52 position or any of the 27 aligned homologous positions, suggesting variation in not tolerated at this position in the general population [13].

(a)



(b)



<u>Fig 1: Mutational spectrum of *CBX1* in individuals with neurodevelopmental symptoms.</u>

(a) Tolerance and meta-domain landscape of *CBX1* gene. The schematic protein representation shows the 2 domains in *CBX1* as purple blocks. CD: chromodomain. CSD: chromoshadow domain. The selected positions highlighted in green represent the variants in this study. The tolerance landscape depicts a missense over synonymous ratio, showing that our variants are contained in the red/orange or intolerant to variation region. Position p.52 is highly intolerant and p.57 is intolerant to variation.

(b) Structural effects of the *CBX1* variants on the protein. The backbone colored in red is the same for both residues and the unique side chain is colored black (A&C). The protein is colored

grey, the side chain of the wildtype residue is green, and the variant residue is red (B&D). (A) Schematic structure of wildtype tryptophan on the left and the variant leucine residue on the right at HP1β position 52. (B) Close up of the W52L variant. (C) Schematic structure of the wildtype asparagine residue on the left and the variant aspartic acid residue on the right at HP1β position 57. (D) Close up of the N57D variant.

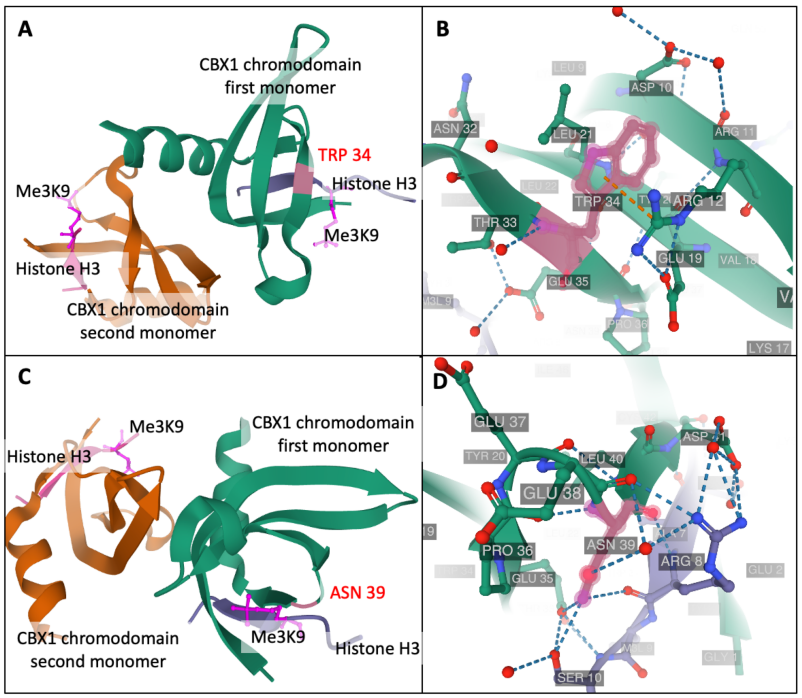
Patient-identified *CBX1* variants disrupt HP1β-chromatin interactions

The CD of HP1β mediates its binding to methylated lysine 9 in histone H3 (H3K9me3), which is a typical constitutive heterochromatin mark, and HP1 subsequently organizes heterochromatin by compacting chromatin [7]. W52 and N57 residues reside within a peptide sequence shown previously to be directly involved in the interaction between HP1β and H3K9me3 [16]. Using a protein structural prediction model, we further evaluated the effects of these variants in this interaction. High resolution structures of HP1β describe a conserved CD fold motif containing three b strands against an alpha helix. The structure referenced represents the complex of the HP1β CD and the histone H3K9me3 peptide. In the structure used for analysis, numbering of the residues begins at the start of the CD rather than the start of HP1β therefore W52 is equivalent to Trp34 and N57 is equivalent to Asn39. Trp34 is an especially conserved residue as it is located structurally on the periphery of the aromatic cage (**Fig 2a**). This residue also forms a cation-pi interaction with CD Arg12, which is an electrostatic interaction that might stabilise the two beta strands and consequently the monomer (**Fig 2a**). Interestingly, Asn39 forms hydrogen bonds with Arg8 and Serine10 of the histone H3 and a hydrogen bond to CD Glu 35 (**Fig 2a**). In the complex, the histone tail interaction might be stabilised by the formation of hydrogen bonds and a complementary surface (**Fig 2a**). Based on the structural prediction models, we hypothesized that the identified variants disrupt HP1β-chromatin interactions, particularly H3K9me3 marked chromatin.

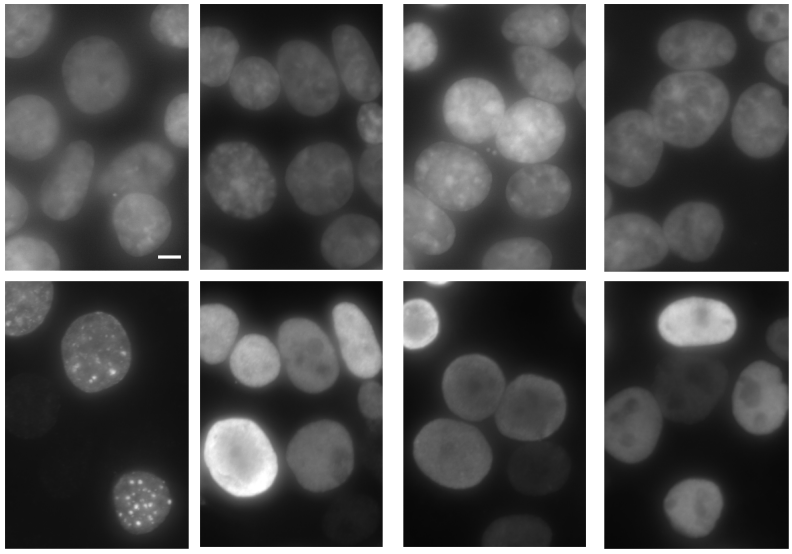
In the nucleus, wild type HP1β co-localizes with the chromocenter, which is DAPI-dense foci composed of constitutive heterochromatin [17]. Previous articles have described that HP1β either lacking the CD or with a variant in the CD (HP1β V23M) localize throughout the nucleus instead of accumulating in chromocenters [17, 18]. We tested if HP1β with patient-identified variants localize 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 (IF) using an antibody against FLAG. While WT HP1β co-localized with the chromocenter, signals of HP1β with CD variants (W52L and N57D) showed homogeneous nuclear staining, indicating mislocalization of the mutant HP1β, similar to that of V23M (**Fig 2b**). These observations are consistent with mutant HP1β lacking the ability to recognize heterochromatin.

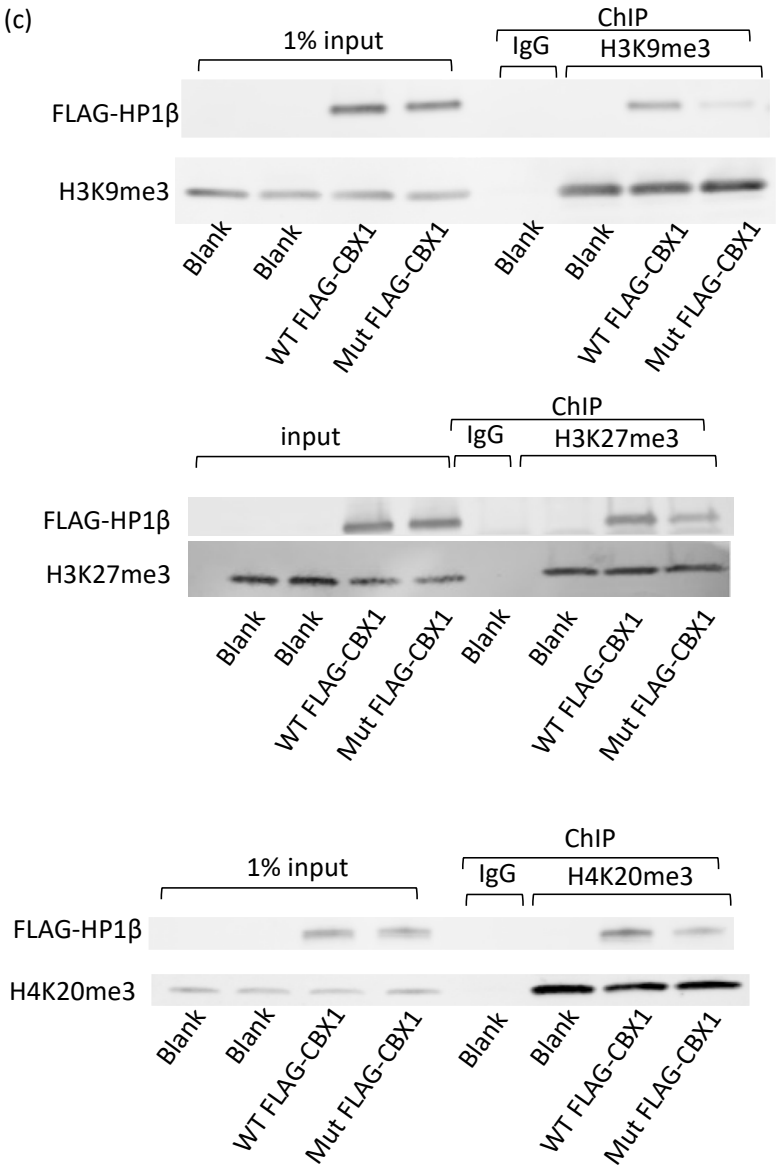
Based on the IF findings, we hypothesized that mutant HP1β binds less efficiently to compacted heterochromatin. Using wild type (WT) and N57D mutant (Mut) FLAG-*CBX1* cDNA vector, the amount of H3K9me3-bound HP1β was evaluated by using Chromatin immunoprecipitation (ChIP)-Western blot (WB) [19]. ChIP-WB showed a reduction of Mut FLAG-*CBX1* binding to H3K9me3, compared to WT (**Fig 2c**). Interactions between HP1β and other repressive histone marks were also evaluated, and similar results were observed for H4K20me3 as well as H3K27me3, histone modification seen in facultative heterochromatin (**Fig 2c**). Similar results were obtained with the W52L mutant for H3K9me3 chromatin binding (**S2 Fig**).

(a)


A
CBX1 chromodomain
first monomer
TRP 34
Histone H3
Me3K9
Me3K9
Histone H3
CBX1 chromodomain
second monomer
B
ASP 10
ASN 32
LEU 21
ARG 11
TRP 34
ARG 12
THR 33
GLU 19
GLU 35
LYS 17
C
Me3K9
CBX1 chromodomain
first monomer
Histone H3
ASN 39
CBX1 chromodomain
second monomer
Me3K9
Histone H3
D
GLU 37
TYR 20
LEU 40
GLU 38
PRO 36
ASN 39
GLU 35
SER 10

(b)


WT
V23M
W52L
N57D
DAPI
Flag



<u>Fig 2: Mutant HP1β abolishes HP1β-chromatin interactions.</u>

(a) Representation of the crystal structure of the complex between human *CBX1* homodimer and H3K9me3 histone peptide (PDB:6D07). The *CBX1* chromodomain (CD) first monomer is shown in green, the second monomer is in orange, the two H3 peptides are in magenta and purple, their methylated lysines are in fuschia and the residue positions affected by variants in salmon. Hydrogen bonds are drawn in blue broken lines and the cation-pi interaction in the orange broken line. Numbering of the structure begins at the start of the CD rather than the

protein so residue positions Trp52 is equivalent to Trp34 and Asn57 is equivalent to Asn39. Trp34 and Asn39 are located in the CD and in close proximity to the binding pocket for H3K9me3 (A, C). A cation-pi interaction between Trp34 and Arg12 in the CD (B). Asn39 forms hydrogen bonds with Arg8 and Serine10 of the histone H3 peptide and a hydrogen bond with Glu 35 of the CD (D).

(b) Mislocalization 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.

(c) 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 (**S2 Fig**).

Transcriptional effects of *CBX1* variants

The transcriptomic and epigenomic effects of the *CBX1* variant were evaluated by performing RNA-sequencing (RNA-seq) comparing the *CBX1* individual 1 and control lymphoblastoid cell lines (LCLs). RNA-seq stats is available as **S1 Table**.

Transcriptome analyses detected 205 upregulated genes and 69 downregulated genes in the patient sample (FDR < 0.05)(**Fig 3, S3 Table**). The majority (75%) of the differentially expressed genes (DEGs) were upregulated in the patient sample, suggesting that the patient mutation of *CBX1* leads to transcriptional activation. Gene list enrichment analysis of the identified upregulated DEGs using Enricher [20] indicated the enrichment of H3K27me3 marked genes, despite HP1 being best known previously as a regulator constitutive heterochromatin

marked by H3K9me3 (**S4 Table**). Particularly, Enricher revealed that H3K27me3-marked genes in a LCL sample (GM12878) used in the ENCODE epigenome project showed the highest overlap. Conversely, H3K9me3-marked genes in the GM12878 LCL sample were not enriched within the patient DEGs (**S4 Table**).

We further evaluated the possibility of H3K27me3 marked gene misexpression by using the ENCODE data set [21]. ChIP-seq results of H3K9me3, H3K27me3, and corresponding input data in a LCL sample (GM12878) were obtained from the ENCODE database. There were 4278 genes and 5845 genes identified to overlap with H3K9me3 and H3K27me3 marks, respectively. Fold gene expression differences of these H3K9me3-and H3K27me3-markerd genes were compared using our RNA-seq dataset. H3K27me3-marked genes tended to be upregulated in the patient sample , while H3K9me3-marked genes did not show major alterations of gene expression, (**Fig 3b**).

To identify the molecular pathways affected by *CBX1* variants, we performed an Ingenuity Pathway Analysis (IPA) using the *CBX1* mutant DEGs of LCL sample. IPA pathway analysis using the *CBX1* mutant DEGs revealed enrichment of genes involved in development of neurons and neuritogenesis, suggesting that the misexpression of these genes underlies the neurodevelopmental symptoms seen in individuals with the *CBX1* variants (**Table 3**).

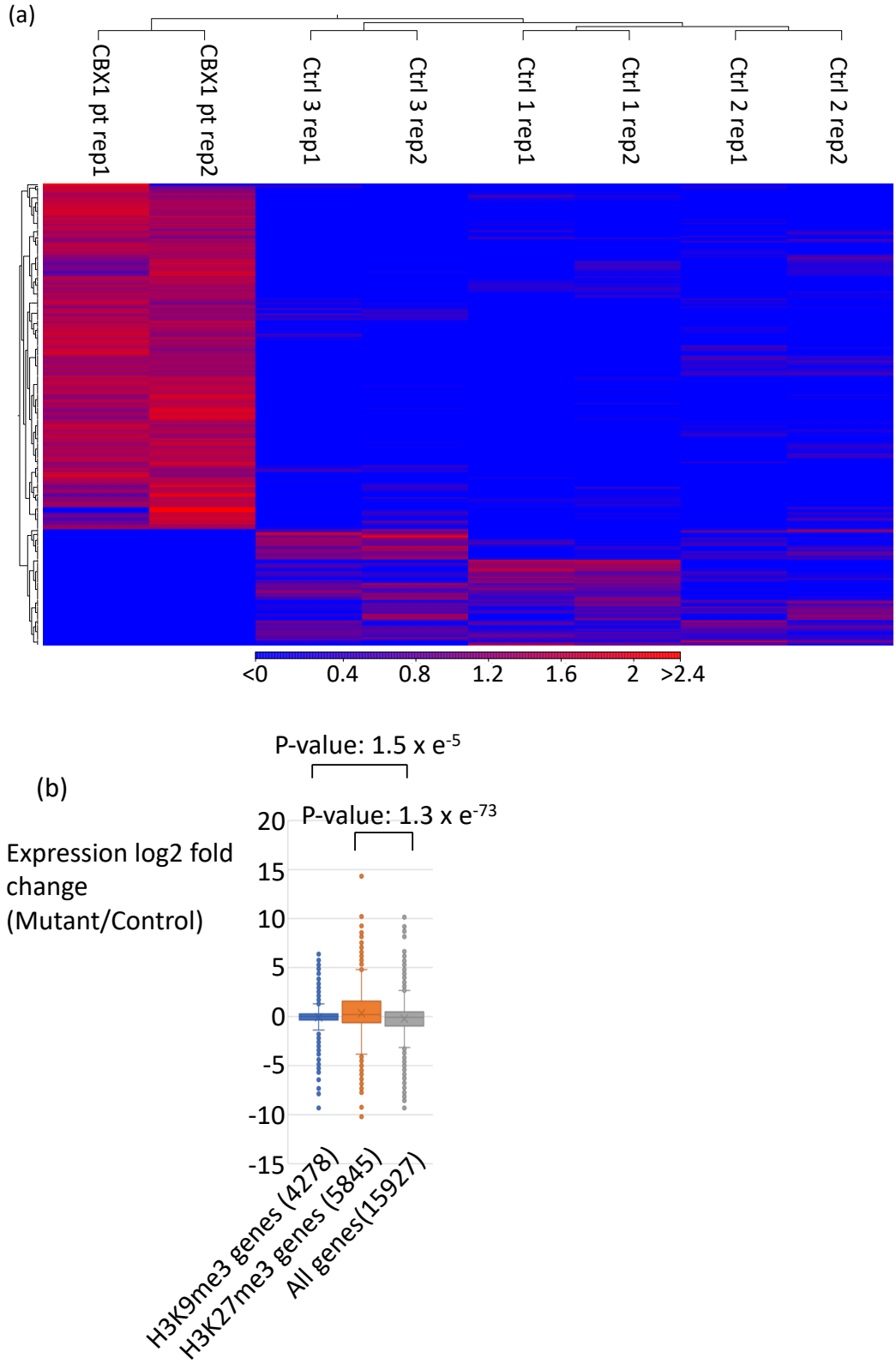


Fig 3: Transcriptomic alterations of *CBX1* mutant sample.

(a) Hierarchal clustering and heat map demonstrated the differentially expressed genes of the *CBX1* mutant LCL. Differentially expressed genes in the LCLs with *CBX1* variants 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.

(b) Box plot demonstrates the gene expression fold difference of H3K9me3- and H3K27me3-marked genes between individual 1 and control samples. Y-axis represents log 2 fold gene expression changes of the *CBX1* mutant sample, compared to the control. Student t-test was used to evaluate the distribution differences between two groups. Mean difference all genes vs H3K9me3= 0.13. Mean difference all genes vs H3K27me3 = 0.56.

Table 3: Ingenuity Pathway Analysis of differentially expressed genes of the *CBX1* mutant sample. Top 15 functional classes enriched in DEGs are listed.

Diseases or Functions Annotation	B-H p-value	Predicted Activation State	Activation z-score	# Molecules
Organization of cytoskeleton	1.64E-09		1.448	61
Organization of cytoplasm	8.52E-09		1.566	63
Development of neurons	2.52E-08	Increased	3.015	41
Cell movement	2.67E-08	Increased	2.179	80
Glucose metabolism disorder	4.35E-08		1.386	59
Migration of cells	8.95E-08	Increased	2.161	73
Microtubule dynamics	8.95E-08	Increased	2.166	51
Formation of cellular protrusions	1.5E-07	Increased	2.159	43
Proliferation of muscle cells	3.11E-07		-0.583	24
Diabetes mellitus	4.36E-07		-0.105	50
Cell movement of phagocytes	5.27E-07		0.745	31
Development of vasculature	8.31E-07		1.216	46
Neuritogenesis	0.0000021	Increased	2.24	31
Accumulation of cells	2.11E-06		1.119	21
Phosphorylation of protein	2.63E-06		0.159	32

Chromatin organizational effects of *CBX1* variants

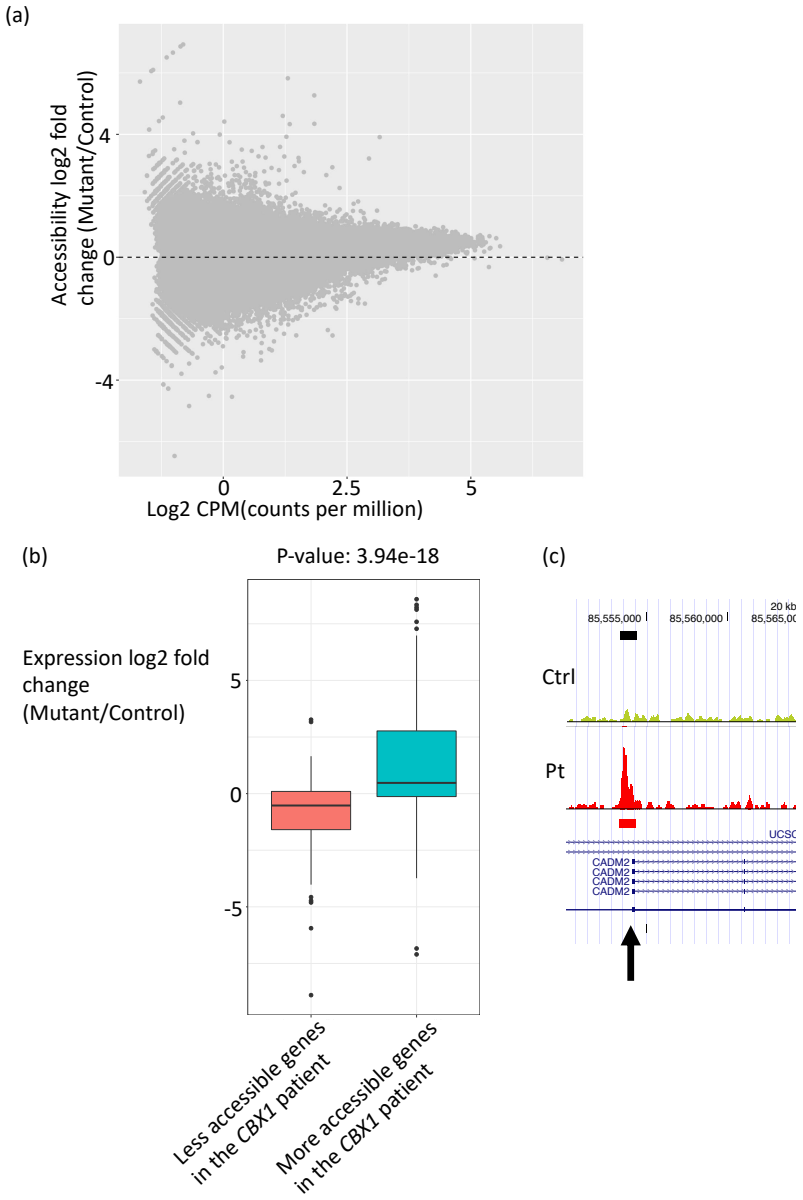
Given the known role of HP1 in organizing heterochromatin, we hypothesized that reduced HP1β bound to repressive histone marks may cause an increase in chromatin

accessibility, resulting in mis-expression of genes. To evaluate this, we performed ATAC-seq using LCLs of individual 1 and control LCL. ATAC-seq statistics are available in **S1 Table**. ATAC-seq identified 181,195 open chromatin regions (OCRs) in either individual 1 or control samples. Overall, chromatin accessibility was slightly increased in the patient sample, compared to the control sample (**Fig 4a**), consistent with an idea that reduced HP1β binding to chromatin results in an increase in chromatin accessibility. Differential analysis identified 3130 differentially accessible regions (DARs) between the two (FDR < 0.05 and log2 Fold Change > 1 or < -1), in which 57% (1813) of DARs showed increased accessibility in the sample from individual 1. Among these, 541 DARs were located at gene promoter regions (-1500/+500bp around transcription start side) and corresponded to 574 genes (**S4 Table**). Among these 574 differentially accessible genes, 17 showed differential expression between the patient and control samples based on our RNA-seq data. We found a positive correlation between the gene expression change and the promoter accessibility change in these genes (**Fig 4b and c,** wilcox rank sum test, $P = 3.94 \times 10^{-18}$)).

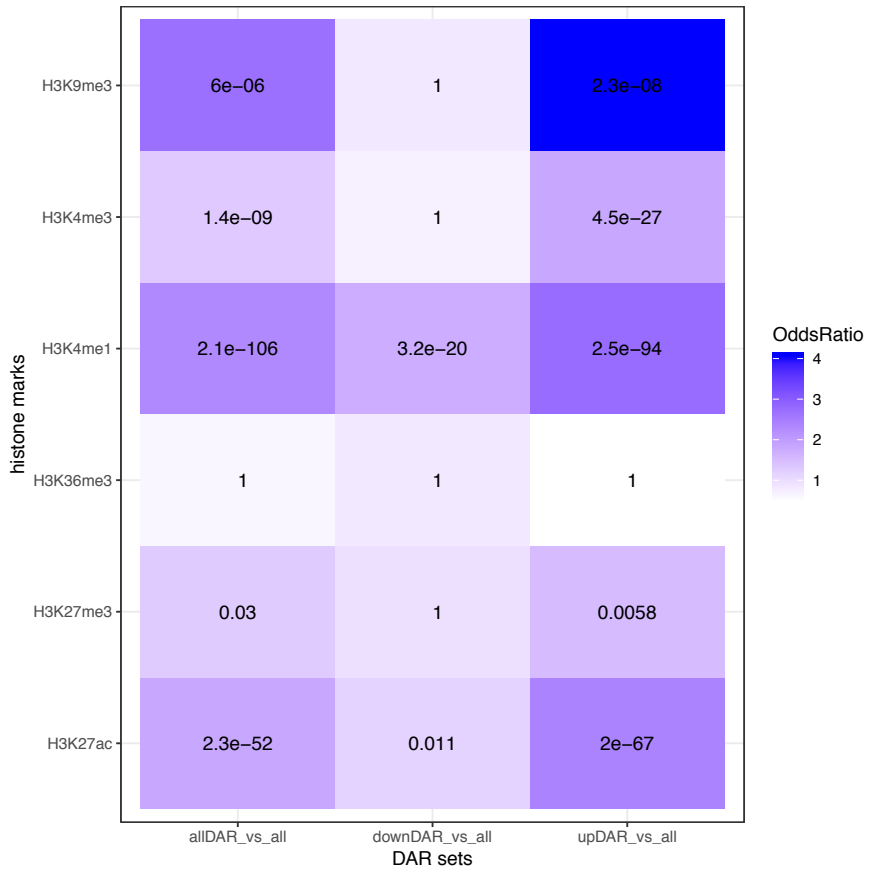
Next, the epigenetic characteristics of OCRs due to reduced HP1β chromatin binding were evaluated by using ATAC-seq data and 15 ChromHMM states derived LCL sample (GM12878)(**S3 Fig**) [22]. As expected, in the comparison to other OCRs, the regions with increased chromatin accessibility in the patient sample were significantly enriched at H3K9me3-marked genomic regions with highest enrichment odds ratio (**Fig 4d**. Fisher exact test, odds ratio = 4.05, $P = 2.3e{-}08$) among other histone marks. H3K27me3-marked genomic regions were also enriched within upregulated DARs found in the patient sample, although the degree of enrichment was less noticeable than those of H3K9me3 (**Fig 4d**).

Lastly, to evaluate the effects of HP1β chromatin binding reduction in the functional regulation of constitutive heterochromatin, chromatin accessibility change over the repetitive regions, the main components of constitutive heterochromatin, was evaluated [23]. The enrichment of differential chromatin accessibilities over repetitive DNA across the genome,

including SINE, LINE, LTR, DNA satellites and simple repeats were evaluated. The regions with differential chromatin accessibility only showed a moderate enrichment for SINEs (**S5 Table**, Fisher exact test odds ratio = 1.49, $P = 1.1 \times 10^{-17}$) and no major alterations of chromatin accessibility were enriched at the LINE, LTR, simple repeats and DNA satellites (**S5 Table**). These findings are consistent with reduced binding of HP1β to repressive chromatin not dramatically affecting organization of constitutive heterochromatin composed of repetitive DNA. Instead, it results in chromatin opening and expression of genes residing in facultative heterochromatin marked by H3K27me3.



(d)



<u>Fig 4: Chromatin organizational alterations of *CBX1* mutant sample.</u>

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

(b) 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 sample with the *CBX1* mutation demonstrated upregulation of gene expression. Three biological replicates were used. Wilcox rank sum test revealed the p-value of $3.94e^{-18}$.

(c) A representative locus of DARs in the patient sample with the *CBX1* variant. Arrow indicates the transcription start site of *CADM2* gene, and increased chromatin accessibility was identified in the patient sample.

(d) 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.

Discussion

Here we report the identification of *de novo CBX1* missense variants in two individuals with developmental delay, hypotonia and autism spectrum disorder. Previously, a mouse model with a homozygous *Cbx1* variant was shown to have abnormal neocortical columnar organization, suggesting a unique role of *Cbx1* in neuronal development [24]. Our data supports HP1β playing a unique non-redundant role in brain development, despite the presence of two other closely related homologues, namely HP1α and HP1γ.

CBX1 encodes a highly conserved, 185-residue homodimer which is involved in condensing the structure of chromatin to repress gene expression [25]. The CD recognizes and binds to histone H3K9me3, the CSD mediates dimerization of *CBX1* and forms a binding surface for interaction with chromatin-associated proteins and the HR interacts with DNA and RNA via a nuclear localisation signal [6, 26-28]. On recognition of the methylated H3K9, the H3 histone tail inserts as a beta strand peptide into the binding pocket caged by three aromatic side chains, Y21, W42 and F45. Complementary interactions between the CD and adjacent residues of the H3 tail stabilise the complex [29]. The W52 residue sits adjacent to a well-conserved phosphorylatable residue threonine, T51 known to be phosphorylated by casein kinase II (CKII) in HP1β, which has a critical effect on nucleosome binding properties [30], important for mobilization of HP1β, leads to phosphorylation of histone H2AX and initiation of the DNA damage response [31]. Interestingly, residue N57 was randomly picked to model protein-ligand binding affinity between *CBX1* and H3K9me3 using neural networking. In multiple rounds of directed evolution testing, the model displayed a strong preference for asparagine or alanine at position 57 [32].

Our analyses provide strong evidence that *CBX1* variants p.(W52L) and p.(N57D) are predicted to have a damaging effect of the structure and function of HP1β. Both variants reside at a highly conserved region, intolerant to variation within the CD. The pLI score in gnomAD (v.2.1.1) [13] is 0.98 consistent with fewer observed than expected missense variants in *CBX1* (expected 99.3 and observed 28). Both residues are located on the surface of the CD domain, and are proximal to the aromatic cage that acts as a binding pocket for methylated histone H3. Structural alterations in this region could therefore perturb the intact hydrophobic pocket that is essential for optimum recognition of H3K9me3. Bonds formed by these residues might be integral to the stabilisation of the CD and to forming closed compact chromatin. Consistent with this prediction, molecular assays confirmed the HP1β variants abolished HP1β-chromatin interactions, resulting in global transcriptional alterations detected by RNA-seq. These *CBX1* variants have therefore confirmed the significance of the interaction between the CD and H3K9me3 in chromatin organization and transcriptional regulation.

Although HP1β is known as essential for the formation of constitutive heterochromatin, direct roles of HP1β in transcriptional regulation outside of heterochromatin have been described [33, 34]. Our data are consistent with the transcriptional regulatory role of HP1β being particularly sensitive for the reduction of chromatin bound HP1β (**S4 Fig**). Transcriptome analysis suggested gene expression within the constitutive heterochromatin was not majorly affected, despite HP1β being a major structural component of constitutive heterochromatin. Instead, genes residing in the facultative heterochromatin marked with H3K27me3 were more likely to be affected. This is supported by 1) upregulated genes in the patient are enriched with H3K27me3 marked genes (**Fig 3**), and 2) reduced mutant HP1β binding to histone H3K27me3 mark (**Fig 2**). A recent study reported that HP1 depletion altered distributions H3K27me3, suggesting the contribution of HP1 in facultative heterochromatin function [9]. In addition, ATAC-seq evaluation showed that chromatin accessibility of repeat DNA sequences, major components of constitutive heterochromatin, was not altered in the sample from individual 1.

Collectively, these findings are consistent with facultative heterochromatin being more sensitive to the reduced HP1β chromatin binding, compared to the constitutive heterochromatin. The binding kinetics of HP1β to chromatin differ depending on the condensation level of chromatin; turn-over of HP1β is slower in highly condensed heterochromatin compared to euchromatin [18]. The difference in dynamics of HP1β binding to chromatin may partially explain why facultative heterochromatin is particularly sensitive to the reduction of chromatin bound HP1β.

Previously, *Cbx1*$^{-/-}$-null or hemizygous mice showed HP1β is essential in maintaining genome integrity as these mice exhibit significant or moderate chromosomal aberrations in their neurosphere cells [24]. However, there is no indication that the patient-identified *CBX1* variants caused genome instability given that the patients' karyotypes are normal. This also supports our hypothesis that the major effect of reduced HP1β chromatin binding is to transcriptional regulation. There are at least two possible explanations for the lack of constitutive heterochromatin dysfunction observed in the affected individuals. One is that *CBX1*-related heterochromatin dysfunction results from biallelic inactivation and that in the affected individuals reported here, the remaining HP1β from the wild type allele may be sufficient to maintain constitutive heterochromatin organization. Another is that other HP1 proteins, such as HP1α and HP1γ, may compensate for the reduced HP1β in heterochromatin. Further investigations are warranted to understand the role of HP1β 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 [35]. ChAHP, comprised CHD4 and ADNP, and mutations in the ChAHP complex components cause similar neurodevelopmental syndromes to those seen in individuals with *CBX1* variants [36, 37]. Therefore, disruption of components of this common molecular pathway may underlie the neurocognitive impairment seen in individuals with *CBX1*, *CHD4*, and *ADNP* mutations.

Here we report the identification of *CBX1* missense variants in individuals with neurocognitive disorders. This represents the first genetic disorder secondary to specific

germline mutations in genes encoding HP1 proteins. We propose the name "*CBX1*-related syndrome" to denote this condition. Our findings provide key evidence of 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 and transcriptional regulation in human development.

Material and Methods

Ethics statement:

Individuals with *CBX1* variants were recruited through GeneMatcher [38] and a network of collaborating clinical geneticists. All individuals were enrolled in the research study under an institutional review board protocol.

Exome sequencing:

Individual 1: Exome sequencing was performed as previously described [10-12]. Briefly, Genomic DNA was extracted from peripheral blood following standard DNA extraction protocols. After genomic DNA extraction, targeted exons were captured using the Agilent SureSelect XT Clinical Research Exome Version 1 kit and sequenced on the Illumina HiSeq 2500 platform with 100 bp paired-end reads. Sequence data was processed using an in-house custom-built bioinformatics pipeline.

Individual 2: Exome sequencing was performed with DNA extracted from EDTA blood from the proband and each parent, which was enriched using the Agilent SureSelect CREv2 and sequenced using Illumina NextSeq 6000.

Conservation and molecular modelling analysis:

The variants were analysed using the Pfam database [39], a large collection of protein domain families represented by curated multiple sequence alignments and a hidden Markov model and which provides a framework that annotates protein domains to proteins. The genetic tolerance of the variants was analysed using MetaDome [40], which visualises the position of

the variant in the context of general population-based genetic variation as well as a schematic representation of protein tolerance using the missense over synonymous variant ratio. The structural effects of a single nucleotide variant in a protein sequence using protein structure analysis tool HOPE, have generated insights into the impact of the variants on the function of *CBX1* [41]. The native crystal structure of human *CBX1* in a complex with H3K9me3 peptide was resolved [29], deposited in the Protein Data Bank under code 6907 and used here for molecular modelling analysis.

Cell culture:

Lymphoblastoid cell lines (LCLs) obtained from individual 1 with a *CBX1* variant N57D and three gender-ethnicity matched control subjects were used. 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 4.5 g/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 °C 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 suspected change and ruled out additional, nonspecific changes.

Chromatin immunoprecipitation (ChIP)-western blot (WB):

ChIP-WB was performed according to the protocol described previously [19]. Following 48 hours after the *CBX1* cDNA vector transfection, cells were fixed with 1% of formaldehyde for

5-10 min. Cross-linking reaction was quenched by 2.5 M Glycine. Cells were washed with cold PBS twice, and then PBS or lysis buffer 1 (LB1; 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, and were centrifuged at 1,500 ×g for 5 min at 4°C. Supernatant was discarded, then, pellets were washed with 1 mL of PBS, and transferred to 1.5 ml tubes followed by centrifugation 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 with BSA/PBS twice at 4°C. H3K9me3 (Abcam: ab8898), H4K20me3 (Abcam: ab9053) and H3K27me3 (Active Motif: MABI0323) antibodies were added to Protein A/G magnetic beads, and they were rotated at 4 °C for more than 3 hours. After the bead-antibody reaction, they were washed twice by BSA/PBS, and once with LB 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 in 100 µl of LB3. Collected cells were resuspended with 1 ml of LB1 and were lysed on ice for 10 min. After centrifugation at 2,000 xg for 5 min, pellets were resuspended in 1 ml of LB2 (20 mM Tris–HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1 mM PMSF). The tubes were placed on ice for 10 min, then spun again to remove supernatant. Obtained pellets were lysed with 1 ml of LB3 for 10 min. After the reaction, the tubes were centrifuged at 2,000 xg for 5 min, and the pellet was resuspended in 400 µl of LB3, and placed on ice for 10 min. After centrifugation at 2,000 g for 1 min, obtained pellets were sonicated, and centrifuged at 20,000 xg for 15 min and supernatant was transferred to new tubes, 30 µl aliquots were taken as whole cell lysate (WCE) samples and additional 10 ul aliquots for Western Blot. Harvested input samples were mixed with antibody beads complexes, and were rotated overnight at 4 ^{o}C. Magnet beads were collected using magnetic stands, and washed with 1 mL cold RIPA buffer five times and then washed once with 1 mL cold TE50. After this, the beads were centrifuged at 2,000 xg for 1 min, and placed in a magnetic holder. Harvested washed magnet beads were resuspended in 50 µl of elution buffer. Then, 2X Laemmli sample buffer was added to the beads, and they were placed in a 95 ^{o}C heat block for

30 min with vigorous vortexing every 5 min. Tubes were then centrifuged at 2,000 xg for 1 min, and the soluble ChIP lysates were transferred to a new tube. Collected ChIP samples were used for Western blot.

Immunofluorescence (IF):

IF was performed as previously described [42]. 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 the 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 a mouse primary antibody against FLAG (Sigma, F1804) for 1 hour at RT, followed by three washes in 0.2% Tween in PBS. 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).

RNA-seq and bioinformatic analysis:

LCLs obtained from the individual with *CBX1* variant N57D and three gender-ethnicity matched 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 150 bp 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-seq analysis. ENCODE ChIP-seq data analyses was performed by using MASER with a default analysis setting [43]. RNA-seq statistics are available in **S1 Table**. 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[44] was used for the peak-calling, and ChIPPeakAnno was used for the peak annotations[45]. Ingenuity Pathway Analysis (QIAGEN) was performed by as previously described [46].

ATAC-seq and bioinformatic analysis:

ATAC-seq was performed as previously described [47]. The triplicates derived from the LCL from individual 1 with N57D variant 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 #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). 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 (v 1.8.1) [48]. 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) [49] and lmFit from limmaVoom (v 3.30.13) [50] but with the normalization scaling factors calculated from csaw [48]. 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 [51] with differentially open regions as user set and all open regions as universal set. ATAC-seq summary statistics are available in **S1 Table**.

Acknowledgements

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

References:

1. Ciptasari U, van Bokhoven H. The phenomenal epigenome in neurodevelopmental disorders. Hum Mol Genet. 2020;29(R1):R42-R50. Epub 2020/08/09. doi: 10.1093/hmg/ddaa175. PubMed PMID: 32766754; PubMed Central PMCID: PMCPMC7530535.

2. Izumi K. Disorders of Transcriptional Regulation: An Emerging Category of Multiple Malformation Syndromes. Mol Syndromol. 2016;7(5):262-73. Epub 2016/11/22. doi: 10.1159/000448747. PubMed PMID: 27867341; PubMed Central PMCID: PMCPMC5109993.

3. Janssen A, Colmenares SU, Karpen GH. Heterochromatin: Guardian of the Genome. Annu Rev Cell Dev Biol. 2018;34:265-88. Epub 2018/07/26. doi: 10.1146/annurev-cellbio-100617-062653. PubMed PMID: 30044650.

4. Le HQ, Ghatak S, Yeung CY, Tellkamp F, Gunschmann C, Dieterich C, et al. Mechanical regulation of transcription controls Polycomb-mediated gene silencing during lineage commitment. Nat Cell Biol. 2016;18(8):864-75. Epub 2016/07/12. doi: 10.1038/ncb3387. PubMed PMID: 27398909.

5. Liu J, Wu X, Zhang H, Pfeifer GP, Lu Q. Dynamics of RNA Polymerase II Pausing and Bivalent Histone H3 Methylation during Neuronal Differentiation in Brain Development. Cell Rep. 2017;20(6):1307-18. Epub 2017/08/10. doi: 10.1016/j.celrep.2017.07.046. PubMed PMID: 28793256; PubMed Central PMCID: PMCPMC5564459.

6. Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, et al. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature. 2001;410(6824):120-4. Epub 2001/03/10. doi: 10.1038/35065138. PubMed PMID: 11242054.

7. Machida S, Takizawa Y, Ishimaru M, Sugita Y, Sekine S, Nakayama JI, et al. Structural Basis of Heterochromatin Formation by Human HP1. Mol Cell. 2018;69(3):385-97 e8. Epub 2018/01/18. doi: 10.1016/j.molcel.2017.12.011. PubMed PMID: 29336876.

8. Bosch-Presegue L, Raurell-Vila H, Thackray JK, Gonzalez J, Casal C, Kane-Goldsmith N, et al. Mammalian HP1 Isoforms Have Specific Roles in Heterochromatin Structure and

Organization. Cell Rep. 2017;21(8):2048-57. Epub 2017/11/23. doi: 10.1016/j.celrep.2017.10.092. PubMed PMID: 29166597.

9. Jamieson K, Wiles ET, McNaught KJ, Sidoli S, Leggett N, Shao Y, et al. Loss of HP1 causes depletion of H3K27me3 from facultative heterochromatin and gain of H3K27me2 at constitutive heterochromatin. Genome Res. 2016;26(1):97-107. Epub 2015/11/06. doi: 10.1101/gr.194555.115. PubMed PMID: 26537359; PubMed Central PMCID: PMCPMC4691754.

10. Baker SW, Murrell JR, Nesbitt AI, Pechter KB, Balciuniene J, Zhao X, et al. Automated Clinical Exome Reanalysis Reveals Novel Diagnoses. J Mol Diagn. 2019;21(1):38-48. Epub 2018/12/24. doi: 10.1016/j.jmoldx.2018.07.008. PubMed PMID: 30577886.

11. Gibson KM, Nesbitt A, Cao K, Yu Z, Denenberg E, DeChene E, et al. Novel findings with reassessment of exome data: implications for validation testing and interpretation of genomic data. Genet Med. 2018;20(3):329-36. Epub 2018/02/02. doi: 10.1038/gim.2017.153. PubMed PMID: 29389922.

12. Wu C, Devkota B, Evans P, Zhao X, Baker SW, Niazi R, et al. Rapid and accurate interpretation of clinical exomes using Phenoxome: a computational phenotype-driven approach. Eur J Hum Genet. 2019;27(4):612-20. Epub 2019/01/11. doi: 10.1038/s41431-018-0328-7. PubMed PMID: 30626929; PubMed Central PMCID: PMCPMC6460638.

13. Karczewski KF, 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. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv. 2019. doi: https://doi.org/10.1101/531210

14. Liu Y, Qin S, Lei M, Tempel W, Zhang Y, Loppnau P, et al. Peptide recognition by heterochromatin protein 1 (HP1) chromoshadow domains revisited: Plasticity in the pseudosymmetric histone binding site of human HP1. J Biol Chem. 2017;292(14):5655-64. Epub 2017/02/23. doi: 10.1074/jbc.M116.768374. PubMed PMID: 28223359; PubMed Central PMCID: PMCPMC5392561.

15. Ye Q, Callebaut I, Pezhman A, Courvalin JC, Worman HJ. Domain-specific interactions of human HP1-type chromodomain proteins and inner nuclear membrane protein LBR. J Biol Chem. 1997;272(23):14983-9. Epub 1997/06/06. doi: 10.1074/jbc.272.23.14983. PubMed PMID: 9169472.

16. Nielsen PR, Nietlispach D, Mott HR, Callaghan J, Bannister A, Kouzarides T, et al. Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. Nature. 2002;416(6876):103-7. Epub 2002/03/08. doi: 10.1038/nature722. PubMed PMID: 11882902.

17. Mateos-Langerak J, Brink MC, Luijsterburg MS, van der Kraan I, van Driel R, Verschure PJ. Pericentromeric heterochromatin domains are maintained without accumulation of HP1. Mol Biol Cell. 2007;18(4):1464-71. Epub 2007/02/23. doi: 10.1091/mbc.e06-01-0025. PubMed PMID: 17314413; PubMed Central PMCID: PMCPMC1838966.

18. Cheutin T, McNairn AJ, Jenuwein T, Gilbert DM, Singh PB, Misteli T. Maintenance of stable heterochromatin domains by dynamic HP1 binding. Science. 2003;299(5607):721-5. Epub 2003/02/01. doi: 10.1126/science.1078572. PubMed PMID: 12560555.

19. Ricke RM, Bielinsky AK. Easy detection of chromatin binding proteins by the Histone Association Assay. Biol Proced Online. 2005;7:60-9. Epub 2005/09/02. doi: 10.1251/bpo106. PubMed PMID: 16136225; PubMed Central PMCID: PMCPMC1190380.

20. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013;14:128. Epub 2013/04/17. doi: 10.1186/1471-2105-14-128. PubMed PMID: 23586463; PubMed Central PMCID: PMCPMC3637064.

21. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res. 2018;46(D1):D794-D801. Epub 2017/11/11. doi: 10.1093/nar/gkx1081. PubMed PMID: 29126249; PubMed Central PMCID: PMCPMC5753278.

22. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nat Methods. 2012;9(3):215-6. Epub 2012/03/01. doi: 10.1038/nmeth.1906. PubMed PMID: 22373907; PubMed Central PMCID: PMCPMC3577932.

23. Saksouk N, Simboeck E, Dejardin J. Constitutive heterochromatin formation and transcription in mammals. Epigenetics Chromatin. 2015;8:3. Epub 2015/03/20. doi: 10.1186/1756-8935-8-3. PubMed PMID: 25788984; PubMed Central PMCID: PMCPMC4363358.

24. Aucott R, Bullwinkel J, Yu Y, Shi W, Billur M, Brown JP, et al. HP1-beta is required for development of the cerebral neocortex and neuromuscular junctions. J Cell Biol. 2008;183(4):597-606. doi: 10.1083/jcb.200804041. PubMed PMID: 19015315; PubMed Central PMCID: PMCPMC2582898.

25. Maison C, Almouzni G. HP1 and the dynamics of heterochromatin maintenance. Nat Rev Mol Cell Biol. 2004;5(4):296-304. Epub 2004/04/09. doi: 10.1038/nrm1355. PubMed PMID: 15071554.

26. Eissenberg JC, Elgin SC. The HP1 protein family: getting a grip on chromatin. Curr Opin Genet Dev. 2000;10(2):204-10. Epub 2000/04/08. doi: 10.1016/s0959-437x(00)00058-7. PubMed PMID: 10753776.

27. Meehan RR, Kao CF, Pennings S. HP1 binding to native chromatin in vitro is determined by the hinge region and not by the chromodomain. EMBO J. 2003;22(12):3164-74. Epub 2003/06/14. doi: 10.1093/emboj/cdg306. PubMed PMID: 12805230; PubMed Central PMCID: PMCPMC162158.

28. Smothers JF, Henikoff S. The HP1 chromo shadow domain binds a consensus peptide pentamer. Curr Biol. 2000;10(1):27-30. Epub 2000/02/05. doi: 10.1016/s0960-9822(99)00260-2. PubMed PMID: 10660299.

29. Arora S, Horne WS, Islam K. Engineering Methyllysine Writers and Readers for Allele-Specific Regulation of Protein-Protein Interactions. J Am Chem Soc. 2019;141(39):15466-70. Epub 2019/09/14. doi: 10.1021/jacs.9b05725. PubMed PMID: 31518125; PubMed Central PMCID: PMCPMC6988089.

30. Yap KL, Zhou MM. Structure and mechanisms of lysine methylation recognition by the chromodomain in gene transcription. Biochemistry. 2011;50(12):1966-80. Epub 2011/02/04. doi: 10.1021/bi101885m. PubMed PMID: 21288002; PubMed Central PMCID: PMCPMC3062707.

31. Ayoub N, Jeyasekharan AD, Bernal JA, Venkitaraman AR. HP1-beta mobilization promotes chromatin changes that initiate the DNA damage response. Nature. 2008;453(7195):682-6. Epub 2008/04/29. doi: 10.1038/nature06875. PubMed PMID: 18438399.

32. Parkinson J, Hard R, Ainsworth RI, Li N, Wang W. Engineering a Histone Reader Protein by Combining Directed Evolution, Sequencing, and Neural Network Based Ordinal Regression. J Chem Inf Model. 2020;60(8):3992-4004. Epub 2020/08/14. doi: 10.1021/acs.jcim.0c00441. PubMed PMID: 32786513.

33. Ayyanathan K, Lechner MS, Bell P, Maul GG, Schultz DC, Yamada Y, et al. Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. Genes Dev. 2003;17(15):1855-69. Epub 2003/07/19. doi: 10.1101/gad.1102803. PubMed PMID: 12869583; PubMed Central PMCID: PMCPMC196232.

34. Cryderman DE, Grade SK, Li Y, Fanti L, Pimpinelli S, Wallrath LL. Role of Drosophila HP1 in euchromatic gene expression. Dev Dyn. 2005;232(3):767-74. Epub 2005/02/11. doi: 10.1002/dvdy.20310. PubMed PMID: 15704177.

35. Ostapcuk V, Mohn F, Carl SH, Basters A, Hess D, Iesmantavicius V, et al. Activity-dependent neuroprotective protein recruits HP1 and CHD4 to control lineage-specifying genes. Nature. 2018;557(7707):739-43. Epub 2018/05/26. doi: 10.1038/s41586-018-0153-8. PubMed PMID: 29795351.

36. Helsmoortel C, Vulto-van Silfhout AT, Coe BP, Vandeweyer G, Rooms L, van den Ende J, et al. A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. Nat Genet. 2014;46(4):380-4. Epub 2014/02/18. doi: 10.1038/ng.2899. PubMed PMID: 24531329; PubMed Central PMCID: PMCPMC3990853.

37. Weiss K, Terhal PA, Cohen L, Bruccoleri M, Irving M, Martinez AF, et al. De Novo Mutations in CHD4, an ATP-Dependent Chromatin Remodeler Gene, Cause an Intellectual Disability Syndrome with Distinctive Dysmorphisms. Am J Hum Genet. 2016;99(4):934-41. Epub 2016/09/13. doi: 10.1016/j.ajhg.2016.08.001. PubMed PMID: 27616479; PubMed Central PMCID: PMCPMC5065651.

38. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat. 2015;36(10):928-30. Epub 2015/07/30. doi: 10.1002/humu.22844. PubMed PMID: 26220891; PubMed Central PMCID: PMCPMC4833888.

39. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. Nucleic Acids Res. 2019;47(D1):D427-D32. Epub 2018/10/26. doi: 10.1093/nar/gky995. PubMed PMID: 30357350; PubMed Central PMCID: PMCPMC6324024.

40. Wiel L, Baakman C, Gilissen D, Veltman JA, Vriend G, Gilissen C. MetaDome: Pathogenicity analysis of genetic variants through aggregation of homologous human protein

domains. Hum Mutat. 2019;40(8):1030-8. Epub 2019/05/23. doi: 10.1002/humu.23798. PubMed PMID: 31116477; PubMed Central PMCID: PMCPMC6772141.

41. Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinformatics. 2010;11:548. Epub 2010/11/10. doi: 10.1186/1471-2105-11-548. PubMed PMID: 21059217; PubMed Central PMCID: PMCPMC2992548.

42. Fiordaliso SK, Iwata-Otsubo A, Ritter AL, Quesnel-Vallieres M, Fujiki K, Nishi E, et al. Missense Mutations in NKAP Cause a Disorder of Transcriptional Regulation Characterized by Marfanoid Habitus and Cognitive Impairment. Am J Hum Genet. 2019;105(5):987-95. Epub 2019/10/08. doi: 10.1016/j.ajhg.2019.09.009. PubMed PMID: 31587868; PubMed Central PMCID: PMCPMC6848994.

43. Kinjo S, Monma N, Misu S, Kitamura N, Imoto J, Yoshitake K, et al. Maser: one-stop platform for NGS big data from analysis to visualization. Database (Oxford). 2018;2018. Epub 2018/04/25. doi: 10.1093/database/bay027. PubMed PMID: 29688385; PubMed Central PMCID: PMCPMC5905357.

44. Xu S, Grullon S, Ge K, Peng W. Spatial clustering for identification of ChIP-enriched regions (SICER) to map regions of histone methylation patterns in embryonic stem cells. Methods Mol Biol. 2014;1150:97-111. Epub 2014/04/20. doi: 10.1007/978-1-4939-0512-6_5. PubMed PMID: 24743992; PubMed Central PMCID: PMCPMC4152844.

45. Zhu LJ, Gazin C, Lawson ND, Pages H, Lin SM, Lapointe DS, et al. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics. 2010;11:237. Epub 2010/05/13. doi: 10.1186/1471-2105-11-237. PubMed PMID: 20459804; PubMed Central PMCID: PMCPMC3098059.

46. Kaur M, Izumi K, Wilkens AB, Chatfield KC, Spinner NB, Conlin LK, et al. Genome-wide expression analysis in fibroblast cell lines from probands with Pallister Killian syndrome. PLoS

One. 2014;9(10):e108853. Epub 2014/10/21. doi: 10.1371/journal.pone.0108853. PubMed PMID: 25329894; PubMed Central PMCID: PMCPMC4199614.

47. Su C, Johnson ME, Torres A, Thomas RM, Manduchi E, Sharma P, et al. Mapping effector genes at lupus GWAS loci using promoter Capture-C in follicular helper T cells. Nat Commun. 2020;11(1):3294. Epub 2020/07/06. doi: 10.1038/s41467-020-17089-5. PubMed PMID: 32620744; PubMed Central PMCID: PMCPMC7335045.

48. Lun AT, Smyth GK. csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. Nucleic Acids Res. 2016;44(5):e45. Epub 2015/11/19. doi: 10.1093/nar/gkv1191. PubMed PMID: 26578583; PubMed Central PMCID: PMCPMC4797262.

49. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40. Epub 2009/11/17. doi: 10.1093/bioinformatics/btp616. PubMed PMID: 19910308; PubMed Central PMCID: PMCPMC2796818.

50. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. Epub 2015/01/22. doi: 10.1093/nar/gkv007. PubMed PMID: 25605792; PubMed Central PMCID: PMCPMC4402510.

51. Sheffield NC, Bock C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. Bioinformatics. 2016;32(4):587-9. Epub 2015/10/29. doi: 10.1093/bioinformatics/btv612. PubMed PMID: 26508757; PubMed Central PMCID: PMCPMC4743627.

Supporting information

S1 Fig: The *CBX1* chromodomain (CD) represented as a meta-domain. The x axis along the top shows the amino acid position in the CD domain. The panel denotes the Pfam hidden Markov Model, where the height of each stack of residues indicates the relative entropy for that position.

S2 Fig: 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-*CBX1* 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.

S3 Fig: ChromHMM states. Generally, the open regions are enriched at active transcriptional regions. Compared to general open regions, regions with elevated openness in chromatin were enriched both H3K27me3 and H3K9me3 dominant states like flanking bivalent TSS or enhancers and repeats, while the regions with closed chromatin in mutant were enriched at bivalent enhancers and polycomb repressive complex binding regions.

S4 Fig: Model of *CBX1*-related syndrome. Reduced HP1β chromatin binding cause increased chromatin accessibility, resulting in overexpression of facultative heterochromatin genes.

S1 Table: RNA-seq and ATAC-seq statistics.

S2 Table: Differentially expressed genes of the *CBX1* mutant LCL sample.

S3 Table: Histone modification enrichment in the *CBX1* mutant LCL differentially expressed genes.

S4 Table: Differentially accessible genes identified by ATAC-seq.

S5 Table: Chromatin accessibility evaluation of heterochromatin repeats.