

The histone acetyltransferase KAT6A is recruited to unmethylated CpG islands via a DNA binding winged helix domain

Lisa Marie Weber^{1,†}, Yulin Jia^{2,†}, Bastian Stielow¹, Stephen S. Gisselbrecht³, Yinghua Cao², Yanpeng Ren², Iris Rohner¹, Jessica King³, Elisabeth Rothman³, Sabrina Fischer¹, Clara Simon¹, Ignasi Forné⁴, Andrea Nist⁵, Thorsten Stiewe⁵, Martha L. Bulyk^{3,6}, Zhanxin Wang^{2,*} and Robert Liefke^{1,7,*}

¹Institute of Molecular Biology and Tumor Research (IMT), Philipps University of Marburg, Marburg 35043, Germany, ²Key Laboratory of Cell Proliferation and Regulation Biology of Ministry of Education, College of Life Sciences, Beijing Normal University, Beijing 100875, China, ³Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA, ⁴Protein Analysis Unit, Biomedical Center (BMC), Faculty of Medicine, Ludwig-Maximilians-University (LMU) Munich, Martinsried 82152, Germany, ⁵Genomics Core Facility, Institute of Molecular Oncology, Member of the German Center for Lung Research (DZL), Philipps University of Marburg, Marburg 35043, Germany, ⁶Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA and ⁷Department of Hematology, Oncology, and Immunology, University Hospital Giessen and Marburg, Marburg 35043, Germany

Received June 23, 2022; Revised November 04, 2022; Editorial Decision November 21, 2022; Accepted November 29, 2022

ABSTRACT

The lysine acetyltransferase KAT6A (MOZ, MYST3) belongs to the MYST family of chromatin regulators, facilitating histone acetylation. Dysregulation of KAT6A has been implicated in developmental syndromes and the onset of acute myeloid leukemia (AML). Previous work suggests that KAT6A is recruited to its genomic targets by a combinatorial function of histone binding PHD fingers, transcription factors and chromatin binding interaction partners. Here, we demonstrate that a winged helix (WH) domain at the very N-terminus of KAT6A specifically interacts with unmethylated CpG motifs. This DNA binding function leads to the association of KAT6A with unmethylated CpG islands (CGIs) genome-wide. Mutation of the essential amino acids for DNA binding completely abrogates the enrichment of KAT6A at CGIs. In contrast, deletion of a second WH domain or the histone tail binding PHD fingers only subtly influences the binding of KAT6A to CGIs. Overexpression of a KAT6A WH1 mutant has a dominant negative effect on H3K9 histone acetylation, which is comparable to the effects upon overexpression of a KAT6A HAT domain mutant. Taken together, our

work revealed a previously unrecognized chromatin recruitment mechanism of KAT6A, offering a new perspective on the role of KAT6A in gene regulation and human diseases.

INTRODUCTION

Human KAT6A [Lysine Acetyltransferase 6A, also named MYST3 and MOZ (Monocytic leukemic zinc-finger)] and KAT6B [Lysine Acetyltransferase 6B, also named MYST4 and MORF (Monocytic leukemia zinc finger protein-related factor)] are related proteins that belong to the MYST family of histone acetyltransferases (1). KAT6A was first identified as fusion partner of CBP in the context of acute myeloid leukemia (2), while KAT6B was later identified as a homolog of KAT6A (3).

Both are very large proteins (~250 kDa) with highly homologous domain structures and functions. They consist of a so-called NEMM domain (N-terminal part of Enok, MOZ or MORF), a double PHD finger (DPF), the HAT domain and a long unstructured C-terminal region, which can be further subdivided into a glutamate/aspartate- and a serine/methionine-rich region (1). The DPF of KAT6A interacts with H3K9 and H3K14 acetylated histones (4). KAT6A not only acetylates histone substrates but has also

*To whom correspondence should be addressed. Tel: +49 6421 28 66697; Email: robert.liefke@imt.uni-marburg.de
Correspondence may also be addressed to Zhanxin Wang. Tel: +86 1058804966; Email: wangz@bnu.edu.cn

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

been implicated in the acetylation of non-histone substrates, such as p53 (5) and SMAD3 (6).

Despite their structural and biochemical similarities, KAT6A and KAT6B appear to have dissimilar biological roles. KAT6A knockout mice are embryonic lethal with several developmental defects and show strongly reduced numbers of hematopoietic cells (7,8). In contrast, KAT6B-deficient mice display defects during neurogenesis, suggesting that KAT6B plays a more dominant role in the brain. (9). Subsequent studies further emphasized the role of KAT6A in the hematopoietic system (10) and associated diseases (11,12). Experiments in zebrafish also supported the role of KAT6A during developmental processes (13,14). Consistently, KAT6A mutations have been found in patients with developmental disorders that are collectively referred to as ‘KAT6A syndrome’, also known as Arboleda-Tham Syndrome (ARTHS) (15–17), which is considered a rare genetic disease with currently fewer than 500 identified patients worldwide (15). Major characteristics of KAT6A syndrome patients are intellectual disability, speech delay, microcephaly, cardiac anomalies, and gastrointestinal complications (15). Mutations involved in KAT6A syndrome mostly lead to the removal of the C-terminal unstructured region, which has been implicated in activating gene transcription (18,19) and protein interactions (5,20). However, the precise mechanisms by which these truncations lead to developmental defects remain unclear.

KAT6A does not operate in isolation but together with other chromatin regulators, such as BRPF1/2/3, ING5 and EAF6 (21) and forms larger protein complexes. These interaction partners have been proposed to be important for the recruitment of KAT6A to its target genes (21) and enzymatic activity (12). In addition, KAT6A has also been shown to interact with sequence-specific transcription factors, such as RUNX1/2 (19,20), PU.1 (7) and PPAR γ (22), suggesting that KAT6A recruitment to chromatin may be highly sophisticated and could involve several distinct mechanisms (23).

Multiple studies have demonstrated that KAT6A plays an important role in HOX gene expression (13,14,24–27). The HOX gene clusters are predominantly regulated by Polycomb group proteins (PcG) and are critical for developmental processes (28). Consistent with a role of KAT6A in HOX gene regulation, studies in *Drosophila* suggest an interaction of the KAT6A-homolog Enok with the Polycomb repressive complex 1 (PRC1) (29,30), pointing to a possible link of KAT6A to the Polycomb system. Since no DNA binding domain has been identified in KAT6A, it has been speculated that unknown transcription factors may recruit KAT6A to Polycomb target genes (23). Recent work suggests that KAT6A is mostly localized to CpG island-containing promoters, including Polycomb target genes, and it has been proposed that this recruitment is facilitated by the interaction of the N-terminal part of the NEMM domain with RNA polymerase II (11).

CpG islands (CGIs) are locations in the genome that are characterized by the accumulation of CpG dinucleotides (31). CGIs are found at most gene promoters and are either in a methylated or unmethylated DNA state. Methylated CGIs are typically transcriptionally silenced (31). In contrast, unmethylated CGIs are in a more active state and are modulated in their activity by specific proteins

that recognize unmethylated CpG motifs. Many of those proteins, such as CXXC1, are part of larger protein complexes that facilitate gene activation (32). On the other hand, KDM2B and mammalian Polycomb-like proteins are associated with Polycomb repressive complexes and are involved in gene repression (33–35). Recently, we identified the SAM domain-containing protein 1 (SAMD1) as another protein that specifically binds to unmethylated CpG motifs (36,37). SAMD1 is associated with L3MBTL3 and the KDM1A histone demethylase complex and functions as a transcriptional repressor (36,38).

Here, we show that the very N-terminal region of KAT6A contains a highly conserved winged helix domain within the NEMM, which specifically interacts with unmethylated CpG-containing DNA motifs. Phylogenetic and structural studies demonstrated that this domain is related to the CpG binding winged helix domain of SAMD1 (36). Mutation of essential amino acids for DNA binding completely abrogates the CpG island binding of KAT6A genome-wide, supporting a pivotal function of this domain for the proper chromatin association of KAT6A. In contrast, this domain appears to be dispensable for the association of KAT6A with gene bodies. Taken together, this work establishes KAT6A as a histone acetyltransferase that possesses a specific DNA binding function, allowing its direct recruitment to unmethylated CpG islands.

MATERIALS AND METHODS

Protein expression and purification

The open reading frame of human KAT6A WH1 was chemically synthesized with codon optimization for efficient bacterial expression. The KAT6A WH1 construct that contains residues 1–85 was cloned into a hexahistidine-SUMO-tagged pRSFDuet-1 vector and expressed in the *Escherichia coli* Rosetta (DE3) strain. The cells were shaken at 37°C until the OD₆₀₀ reached ~1.0, and then cooled at 20°C for around an hour before 0.2 mM IPTG was added to induce expression overnight. Cells were collected by centrifugation at 5000 \times g for 10 min. Cell pellets were resuspended with the initial buffer (20 mM Tris at pH 7.0, 500 mM NaCl and 20 mM imidazole) and sonicated at 4°C for around 5 min. The supernatant was pooled by centrifuging the cell lysate at 18 000 \times g for an hour. Histidine-SUMO-tagged target protein was isolated through a nickel-charged HiTrap Chelating FF column from GE Healthcare. The histidine-SUMO tag was cleaved by incubating with a histidine-tagged ubiquitin-like-specific protease 1 (ULP1) at 4°C for around 1 h. After dialysis with the initial buffer at 4°C for around 3 h, the solution was then reloaded onto a nickel-charged chelating column to remove both the histidine-tagged SUMO and ULP1 protease. The flow through was collected and diluted with water until the NaCl concentration reached at ~200 mM before it was loaded onto a heparin column (GE Healthcare) to remove bound DNA. Target protein was eluted through linearly increasing the NaCl concentration from the low salt buffer (20 mM Tris pH 7.0, 2 mM DTT) to the same buffer containing 1 M NaCl. The target protein was further purified by a HiLoad 200 16/600 gel filtration column equilibrated with the low salt buffer. After gel filtration, the target protein fractions were

diluted with water until the NaCl concentration reached to around 100 mM, and then loaded directly onto a Source 15S 4.6/100 column for further purification. Target protein was separated by increasing the NaCl concentration of the low salt buffer (20 mM Tris-pH 7.0, 100 mM NaCl, 2 mM DTT) from 100 mM to 1 M through a linear gradient. After these purification steps, the target protein was concentrated to ~20 mg/ml and was stored in a -80°C freezer.

All mutations of KAT6A WH1 were generated by PCR-based methods and subcloned into a hexahistidine-SUMO-tagged pRSFDuet-1 vector. The purification steps of these mutants were similarly to wild-type protein, except that for some mutants, the purification step through the heparin column was omitted, since those mutants that had disrupted DNA binding ability and could not attach to the heparin column.

ZMYND11 WH (1–95) and KAT6B WH1 (1–100) were also cloned into hexahistidine-SUMO-tagged pRSFDuet-1 vectors and expressed in the *E. coli* Rosetta (DE3) strain. The purification procedures of both proteins were similar to KAT6A WH1.

Crystallization and structure resolution

Crystallization was carried out using the hanging-drop, vapor-diffusion method by mixing equal volumes of protein and well solutions. The complex of KAT6A WH1 (1–85) and DNA was prepared by mixing the target protein with a 14-bp palindromic CpG-containing dsDNA (5'-GG AGTGCGCACTCC-3' is the sequence of one strand) at the molar ratio of 2:1.1. Crystals of KAT6A WH1/DNA complex were grown in the solution containing 0.2 M magnesium acetate tetrahydrate, 20% polyethylene glycol 3350 at 20°C. Crystals were flash-frozen in the cryoprotectant with the crystallization buffer containing 10% 2,3-butanediol. Datasets for the human KAT6A WH1/DNA complex crystals were collected at the Shanghai Synchrotron Radiation Facility (SSRF) beamline BL18U1 in China at the wavelength of 0.97930 Å. The datasets were processed using the program HKL2000. The structure of the KAT6A WH1/DNA complex was solved by the molecular replacement method by PHENIX using the SAMD1 WH/DNA complex structure as the model. The initial partial model was manually rebuilt in Coot and further refined by PHENIX.

The complex of the dimer-formed KAT6A WH1 (1–85) and DNA was prepared by mixing the target protein with a 13-bp CpG-containing dsDNA (5'-GGTCCGTCGGAC C-3' is the sequence of one strand) at the molar ratio of 2:1.1. The crystals of KAT6A WH1/DNA complex were grown in the solution containing 0.01 M magnesium chloride hexahydrate, 0.005 M nickel(II) chloride hexahydrate, 0.1 M HEPES sodium pH 7.0, 12% w/v polyethylene glycol 3350 at 20°C. Crystals were flash-frozen in the cryoprotectant with the crystallization buffer containing 8% 2,3-butanediol. Datasets for the human KAT6A WH1/DNA complex crystals were collected at the Shanghai Synchrotron Radiation Facility (SSRF) beamline BL19U1 in China at the wavelength of 0.97979 Å. The datasets were processed using the program HKL2000. The structure of KAT6A WH1/DNA complex was solved by molecular replacement method by PHENIX using the monomer-formed

KAT6A WH1/DNA complex structure as the model. The initial partial model was manually rebuilt in Coot and further refined by PHENIX.

Gel filtration analysis

The complex assembly of KAT6A and DNA was assessed by analytical gel filtration. DsDNA (50 nmol) was mixed with excessive KAT6A WH1 to a final volume of 200 µl in the dilution buffer of 20 mM Tris-pH 7.0, 100 mM NaCl and 2 mM DTT. The mixtures were incubated at 4°C for 10–20 min and loaded to a Superdex 200 10/300 GL column (GE) equilibrated in the dilution buffer. Elution profiles were monitored at the OD of 280 nm.

Isothermal titration calorimetric measurement

Calorimetric experiments were carried out at 20°C with a MicroCal iTC200 instrument. Purified wild-type or mutant proteins and DNA duplexes were dialyzed overnight at 4°C in titration buffer containing 20 mM Tris pH 7.5, 100 mM NaCl and 2 mM β-mercaptoethanol. Titration was performed by injecting DNA duplexes into the proteins. Calorimetric titration data were fitted with the Origin software under the algorithm of one binding-site model.

Electrophoretic mobility shift assay

Fifty picomoles of dsDNA was mixed with increasing amount of KAT6A WH1 proteins in the buffer containing 20 mM Tris pH 7.0, 100 mM NaCl and 2 mM DTT, and incubated at 4°C for 10 min. The mixture was then loaded onto a 1.2% agarose gel in the TAE buffer for electrophoresis and detected by ethidium bromide staining. KAT6A WH1 (1 to 85) and its mutants were used for the assay. All EMSA experiments were repeated at least three times. One strand of the DNA sequences used in the EMSA assays are listed below. GC-rich: 5'-GGCCTGCGCAGGCC-3'; AT-rich: 5'-ATATATATATATAT-3'; 14-bp-CpG: 5'-GGAG TGCGCACTCC-3'; 14-bp-TpG: 5'-GGAGTGTGCACT CC-3'; 14-bp-GpG: 5'-GGAGTGGGCACTCC-3'; 14-bp-ApG: 5'-GGAGTGAGCACTCC-3'; 14-bp-CpA: 5'-GG AGTGCACTCC-3'; 14-bp-CpC: 5'-GGAGTGCCCA CTCC-3'; 14-bp-CpT: 5'-GGAGTGCTCACTCC-3'; 14-bp-mCpG: 5'-GGAGTG(m)CGCACTCC-3'.

Protein binding microarray

KAT6A WH1 (1–85), WH2 (85–181) or their combination (1–181) were cloned into the pT7CFE1-NHis-GST-CHA plasmid (Thermo Scientific: #88871). GST-fusion proteins were expressed using the 1-Step Human Coupled IVT Kit (ThermoFisher Scientific). Expressed protein concentrations were estimated from anti-GST Western blots. Subsequently, custom-designed 'all-10mer' universal oligonucleotide arrays in 8 × 60K GSE array format (Agilent Technologies; AMADID 030236) were double-stranded and PBM experiments were performed essentially as described previously (39) with Alexa488-conjugated anti-GST antibody (Invitrogen A-11131). The KAT6A WH1 domain was assayed in duplicate at a final concentration of 600 nM in PBS-based binding and wash buffers, on fresh slides. Scans

were acquired using a GenePix 4400A (Molecular Devices) microarray scanner. Microarray data quantification, normalization, and motif derivation were performed essentially as described previously using the Universal PBM Analysis Suite and the Seed-and-Wobble motif-derivation algorithm (39). Analysis of 4-mer scores was performed by selecting subsets of 60-bp probes matched for CpG content and counting the instances of each 4-mer in all relevant probe sequences, then calculating the Pearson correlation coefficient and significance of 4-mer count and signal intensity for all 4-mers. The results were highly consistent across a range of values of CpG count, each of which represents a non-overlapping set of probes considered; a value of 4 (the mode) was chosen for display.

Cell culture

E14 mouse ES cells (E14TG2a) were cultured in Dulbecco's modified Eagle's medium (DMEM) and GlutaMAX (Gibco; 61965-026), 15% Fetal Bovine Serum (FBS) (Merck; F7524) 1× nonessential amino acids (Gibco; 11140-035), 1× sodium pyruvate (Gibco; 11160-039), 1× penicillin/streptomycin (Gibco; 15140-122), 10 μM β-mercaptoethanol (Gibco; 31350-010), and with home-made human LIF produced in COS-7 cells, cultured in DMEM GlutaMAX, 10% FBS, 1× nonessential amino acids, 1× penicillin/streptomycin, transfected with pCAG-hLIF plasmid using polyethylenimine (PEI) reagent (Polysciences; 23966). Medium containing secreted hLIF was collected and filter-sterilized, subsequently assayed for working concentration on mES cells with Alkaline Phosphatase detection kit (Sigma-Aldrich; SCR004). Mouse ES cells were cultured on 0.2% gelatin-coated plates.

KAT6A mES KO cells were created by transfection using the jetPRIME transfection reagent (Polyplus; 101000046) with LentiCRISPRv2 (Addgene no. 52961) (40) constructs and two different single guide RNAs targeting mKAT6A (sg2: TGGCTGCCTTAGTGTGAGG; sg3: TGTGGAAGCAGTGACACGGG). After selection with 2 μg/ml puromycin (Merck; 58-58-2), single-cell clones were obtained and further validated. Sanger sequencing confirmed the knockout.

Human embryonic kidney-293 (HEK293) cells were cultured with DMEM/F-12 (Gibco, 31331-028), supplemented with 10% FBS and 1× penicillin/streptomycin. For ectopic, FLAG-HA tagged gene overexpression, the respective KAT6A cDNAs were cloned into pDEST expression vectors using the Gateway cloning system (Invitrogen). Cells were transiently transfected with the plasmids using PEI. Stably expressing single clones were selected with 2 μg/ml puromycin. Following modified KAT6A constructs were used:

KAT6A WH1 mut	K24A/Q24A
KAT6A HAT mut	Q654E/G657E (41)
KAT6A WH1 only	1-85
KAT6A WH1-WH2	1-181
KAT6A WH1-DPF	1-324
KAT6A WH1-HAT	1-785
KAT6A ΔWH2	Δ94-171
KAT6A ΔDPF	Δ205-313

Antibodies

The following antibodies were used:

Anti-KAT6A (MYST3)	Affinity Biosciences	DF9024
Anti-KAT6A	Sigma-Aldrich	HPA063266
Anti-FLAG	Sigma-Aldrich	F3165
Anti-H3K9ac	Diagenode	C15410004
Anti-H3K14ac	Active Motif	39697
Anti-Suz12	Santa Cruz	sc-271325
Anti-SP1	self-made	(42)
Anti-HA	Merck	11867423001
Anti-L3MBTL2	Active Motif	39569
Anti-PCGF6	Proteintech	24103
Anti-Tubulin	Merck	MAB3408
Anti-H3	Abcam	ab1791

Nuclear extract preparation

To obtain nuclear extract, the cytoplasmic fraction was removed by incubating harvested cells for 10 min at 4°C in low salt buffer (10 mM HEPES/KOH pH 7.9, 10 mM KCl, 1.5 mM MgCl₂, 1× PIC (Protease Inhibitor Cocktail) (Roche; 04693116001), 0.5 mM PMSF). After centrifugation, the remaining pellet was dissolved in high salt buffer (20 mM HEPES/KOH pH 7.9, 420 mM NaCl, 1.5 mM MgCl₂, 0.2 mM EDTA, 20% glycerol, 1× PIC, 0.5 mM PMSF) and incubated for 30 min, 4°C while shaking. Subsequently, the lysates were centrifuged and the supernatant containing the nuclear fraction was further analyzed by western blotting.

Subcellular fractionation

Cellular fractionations were performed using 'Subcellular Protein Fractionation Kit for Cultured Cells' (Thermo Fisher Scientific; 78840) according to the manufacturer's instructions, followed by Western blotting. As loading controls for the respective fractions, a self-made SP1 antibody (42), anti-Tubulin (Merck; MAB3408), and anti-H3 (Abcam; ab1791) were applied.

FLAG immunoprecipitation and mass spectrometry

FLAG-HA-tagged human KAT6A wild-type, the WH1 KQ-AA mutant, and FLAG-HA-GFP as a control were stably expressed in HEK293 cells. Whole cell extracts were prepared from around 1 × 10⁸ cells with NP-40 lysis buffer (20 mM Tris-HCl, pH 7.6, 300 mM KCl, 12.5 mM MgCl₂, 0.5% NP-40, 10% glycerol and protease inhibitors). Immunoprecipitation was performed using Flag M2 agarose beads (Sigma) with around 5 mg of total protein per IP in NP-40 lysis buffer containing 150 mM KCl and 0.25% NP-40 for 3 h at 4°C. After three washes with TBS, 0.25% NP-40, bound proteins were eluted by the addition of Laemmli buffer or 0.2 mg/ml FLAG peptide and subsequently analyzed by western blotting or silver staining.

For mass spectrometric analysis, beads were washed additional 3 times with 50 mM NH₄HCO₃ and digested with trypsin after reduction and alkylation. For LC-MS purposes, desalted peptides were injected in an Ultimate 3000 RSLCnano system (Thermo) and separated in a 25-cm analytical column (75 μm ID, 1.6 μm C18, IonOpticks)

with a 50-min gradient from 2 to 37% acetonitrile in 0.1% formic acid. The effluent from the HPLC was directly electrosprayed into a Qexactive HF (Thermo) operated in data dependent mode to automatically switch between full scan MS and MS/MS acquisition. Survey full scan MS spectra (from m/z 375–1600) were acquired with resolution $R = 60\,000$ at m/z 400 (AGC target of 3×10^6). The 10 most intense peptide ions with charge states between 2 and 5 were sequentially isolated to a target value of 1×10^5 , and fragmented at 27% normalized collision energy. Typical mass spectrometric conditions were: spray voltage, 1.5 kV; no sheath and auxiliary gas flow; heated capillary temperature, 250°C; ion selection threshold, 33 000 counts. MaxQuant 2.0.1.0 was used to identify proteins and quantify by iBAQ with the following parameters: Database, uniprot_P000005640_Hsapiens_20191126.fasta; MS tol, 10 ppm; MS/MS tol, 20 ppm Da; peptide FDR, 0.1; protein FDR, 0.01 min. peptide length, 7; variable modifications, oxidation (M); fixed modifications, carbamidomethyl (C); peptides for protein quantitation, razor and unique; min. peptides, 1; min. ratio count, 2. results were analyzed in Perseus 1.6.15.0. The processed MS data are available as Supplementary Table S4.

Immunofluorescence staining

HEK293 and mouse ES cells were seeded on 0.2% gelatin-coated coverslips. Cells were fixed with 4% formaldehyde (w/v), methanol-free (Thermo Fisher Scientific; PI28906), and subsequently permeabilized with wash buffer (0.5% Triton X-100 in PBS). Blocking was performed with 10% FCS in wash buffer. Primary antibodies were diluted 1:500 in blocking solution and incubated in a wet chamber, overnight at 4°C. Three washing steps of the cells were performed before incubation with secondary antibody, using Alexa Fluor 488 goat anti-rabbit IgG (H + L) (Thermo Fisher Scientific; A-11008), at 1:2000 dilution. Following three washing steps, the coverslips were mounted onto microscopy slides using VECTASHIELD® Antifade Mounting Medium with DAPI (Vector Laboratories; H-1200), and sealed. Microscopy was performed using a Leica DM5500 microscope, and data was analyzed using ImageJ (Fiji).

Proliferation assay

To determine proliferation rates, cells were seeded on 6-well plates at a density of 1×10^5 cells per well. The cell viability was determined 1, 3 and 6 days after seeding using MTT assay by adding 90 μ l of 5 mg/ml Thiazolyl blue $\geq 98\%$ (Carl Roth; 4022) to each well. After 1h, the medium was aspirated, and stained cells were dissolved in 400 μ l lysis buffer (80% isopropanol, 10% 1 M HCl, 10% Triton X-100) and diluted further if necessary. Absorption was measured at 595 nm using a plate reader. All values were normalized to day 1 to compensate for variations in seeding density. The mean value of three biological replicates was determined.

Colony formation assay

To examine the ability of cells to form colonies, cells were seeded at low density (1×10^3 cells per well on 6-well

plates) and cultured for 7 days. Afterwards, the cells were washed once with PBS and fixed with 100% methanol for 20 min before staining for 5 min with 0.5% crystal violet in 25% methanol. To remove excess coloring, the plates were washed with dH₂O until single colonies were visible.

DNMT1 inhibition

HEK293 cells were transiently transfected with constructs for expression of FLAG-HA-tagged human KAT6A using PEI. Twenty-four hours after transfection, the cells were treated with the DNMT1 inhibitor GSK-3484862 (MedChemExpress, Nr.: HY-135146) at 2 μ M and 10 μ M for 3 days as well as 0.1% DMSO as a solvent control as described (43), followed by a chromatin immunoprecipitation using a FLAG antibody and qPCR.

RT-qPCR and RNA-seq

For RNA isolation, cells were cultivated on 6-well plates up to 80–100% confluency. RNA was prepared using the RNeasy Mini Kit (Qiagen; 74004) according to the manufacturer's manual; including an on-column DNA digest (Qiagen; 79254).

The PrimeScript RT Reagent Kit (TaKaRa; RR037A) was used to transcribe mRNA into cDNA according to the manufacturer's manual. Samples were incubated for 30 min at 37°C followed by 5 min at 85°C to inactivate PrimeScript RT enzymes. Subsequently, cDNA was diluted 1:20 to be used in RT-qPCR.

For analysis by real-time quantitative PCR, MyTaq Mix (Bioline; BIO-25041) was used. For gene expression analysis, values were normalized to either mActb and mGapdh or hHPRT and hGAPDH expression. The qPCR primers used are presented in Supplementary Table S2.

For RNA-Seq, RNA integrity was assessed on Experion StdSens RNA chips (Bio-Rad). RNA-Seq libraries were prepared using the TruSeq Stranded mRNA Library Prep Kit (Illumina). RNA-Seq libraries were quantified on a Bioanalyzer (Agilent Technologies). Next-generation sequencing was performed on Illumina HiSeq1500 or NextSeq550.

Chromatin immunoprecipitation

Chromatin-immunoprecipitation (ChIP) was performed in accordance with the Fast ChIP protocol (44) using antibodies described above. ChIP-qPCRs with gene-specific primers (Supplementary Table S2) were performed using the MyTaq PCR reagent (Bioline) in the presence of 0.1x SYBR Green (Molecular Probes). For ChIP-sequencing, two to three individual ChIPs were pooled. The precipitated chromatin was eluted from the beads with 100mM NaHCO₃, 1% SDS. Crosslinking was reversed by an overnight incubation at 65°C followed by a proteinase K digestion. The precipitated DNA was purified using QIAquick columns (Qiagen). Five to ten nanograms DNA were used for indexed sequencing library preparation using the Microplex Library preparation kit v2 (Diagenode). Libraries were purified and size-selected by AMPure magnetic beads (Beckman) and quantified on a Bioanalyzer (Agilent). Next-generation sequencing was performed on Illumina NextSeq 550.

Bioinformatics analysis

ChIP-Seq data were aligned to the human genome hg38 using Bowtie (45). Bigwig files, heatmaps and binding profiles were created using samtools and DeepTools (46). Peak calling was performed with MACS2 with standard settings (47). The genomic distribution of SAMD1 was determined using ChIPseeker (Galaxy Version 1.28.3) (48). Gene ontology analysis of KAT6A target loci was performed using GREAT (49). Enriched motifs at KAT6A bound locations were identified using HOMER (50). Genomic occurrences of palindromic CCGNCGG motifs (and variants) were determined via HOMER ('scanMotifGenomeWide')(50). ChIP-Seq tracks were visualized using the UCSC browser (51). Promoter definition and CpG islands were downloaded from the UCSC table browser. The correlation analysis of features of the KAT6A bound gene bodies was performed via the Cistrome data analysis platform (52), using the region of +1000 to +10 000 for each gene.

RNA-Seq samples were aligned to the mouse transcriptome GENCODE vM25 using RNA-Star (2.7.2b) (53). Reads per gene were calculated using feature counts (2.0.1). Differentially regulated genes and normalized read counts were determined using DESeq2 (2.11.40.6) (54). Genes with an at least 0.75-fold (\log_2) difference and a *P*-value below 0.01 were considered differentially expressed genes. Gene set enrichment analysis (GSEA) (55) was performed with standard settings.

The following public datasets for HEK293 cells were used: CIRA-Seq: DRR186438 (11); WGBS: GSM3791391 (56); H3K4me3: GSM1249885 (57); H3K4me2: GSM1249886 (57); H3K4me1: GSM2711410 (58); RNA Polymerase II: GSM3073973 (59); PCGF6: ERR2103747 (60), E2F6: ERR2103744 (60), L3MBTL2: ERR2103745 (60); RNA-Seq: DRR065497 (11). The GC content was downloaded from the UCSC genome browser (file name: hg38.gc5Base.bw) (51)

RESULTS

KAT6A possesses a conserved N-terminal winged helix domain

It has been proposed that KAT6A is recruited to its target genes via a combination of the histone reader function of its double PHD finger (DPF) domain (1,4,61,62), transcription factors (7,20,22) and chromatin binding of interaction partners, such as ING5 and BRPF1 (21). The N-terminal so-called NEMM domain has also been implicated in chromatin binding of KAT6A (1,63), but the specific mechanisms of this region are unclear. Investigation of the domain structure of KAT6A using the AlphaFold database (64,65) demonstrated that KAT6A possesses two well-defined globular domains at the N-terminus (Figure 1A–C), reflecting the NEMM. The globular domain at the very N-terminus has not yet been described, while the more C-terminal domain has previously been shown to have similarity to the linker histone H1/5 (19,63). Whether this domain indeed functions as a linker histone is unknown. Since both domains are characterized by three alpha-helices and two beta-sheets, each, typical for winged-helix (WH) domains (Figure 1B) (66), we named these two domains

winged helix domain 1 (WH1) and -2 (WH2) (Figure 1A–C). A very similar structural composition was also found for the paralog KAT6B (Supplementary Figure S1A–C), and the *Drosophila* ortholog Enok (Supplementary Figure S1D–F). In KAT6A and KAT6B, the two WH domains are separated by approximately 25 amino acids (Figure 1B, Supplementary Figure S1C), while in Enok these two domains are only separated by 5 amino acids (Supplementary Figure S1F). The different distances between the two domains raise the possibility that the functions of the double WH domains diverged during evolution. Although the WH1 domains of KAT6A and KAT6B are currently not annotated in the UniProt database (Q92794, Q8WYB5), phylogenetic analysis demonstrates that this domain is evolutionarily ancient. A homologous domain can be identified in the KAT6A homolog of the freshwater polyp (*Hydra vulgaris*) (Figure 1D), suggesting that the WH1 has already evolved during early metazoan development. Comparing the N-terminal domains of KAT6A, KAT6B and Enok shows that the WH1 is evolutionarily more conserved than the WH2 and the DPF domain (Supplementary Figure S1G).

Furthermore, the WH1 domains of KAT6A and KAT6B together with the WH domains of SAMD1 and ZMYND11 form a group of WH domains with shared sequence homology. We previously named this group 'SAMD1-like WH domains' (36). Within this group, the WH1 domains of KAT6A, -B and the WH of SAMD1 are similar to each other, while the WH domain of ZMYND11 is more distant (Figure 1D, E). Previously, we demonstrated that the WH domain of SAMD1 directly interacts with unmethylated CpG motifs, which is important for the recruitment of SAMD1 to CGIs (36). Thus, given the observed localization of KAT6A to CpG islands (11) and the sequence similarity of the WH1 domain to the CpG binding WH domain of SAMD1 (Figure 1E), we speculated that the WH1 domains of KAT6A and -B may also bind to CpG-containing motifs. To address this possibility, we performed EMSA (electrophoretic mobility shift assays) experiments using a CpG-rich DNA sequence. We found that the WH1 domains of both KAT6A and KAT6B efficiently bind to CpG-rich DNA (Figure 1F). In contrast, the WH domain of ZMYND11 did not show any DNA binding capacity in this experiment (Figure 1F). To further validate the DNA binding preference of KAT6A, we performed unbiased protein binding microarray (PBM) experiments (39), which confirmed that the WH1 of KAT6A preferentially binds to CpG-rich motifs *in vitro* (Figure 1G, Supplementary Table S3). We found that both a simple GCGCG motif, and a palindromic CCGNCGG motif were enriched. Importantly, in both replicates, we observed approximately one hundred 8-mers with an E-score of more than 0.45, which indicates a highly robust DNA binding (Supplementary Table S3). Since the WH2 has been implicated in chromatin binding of KAT6A (19), we also investigated the DNA binding ability of KAT6A WH2. In EMSA, we could not observe any binding of this domain to CpG-rich DNA (Supplementary Figure S2A). Additionally, in PBM experiments we could not detect any 8-mer with an E-score higher than 0.45, suggesting little or no sequence-specific DNA binding (Supplementary Figure S2B, Table S3). The com-

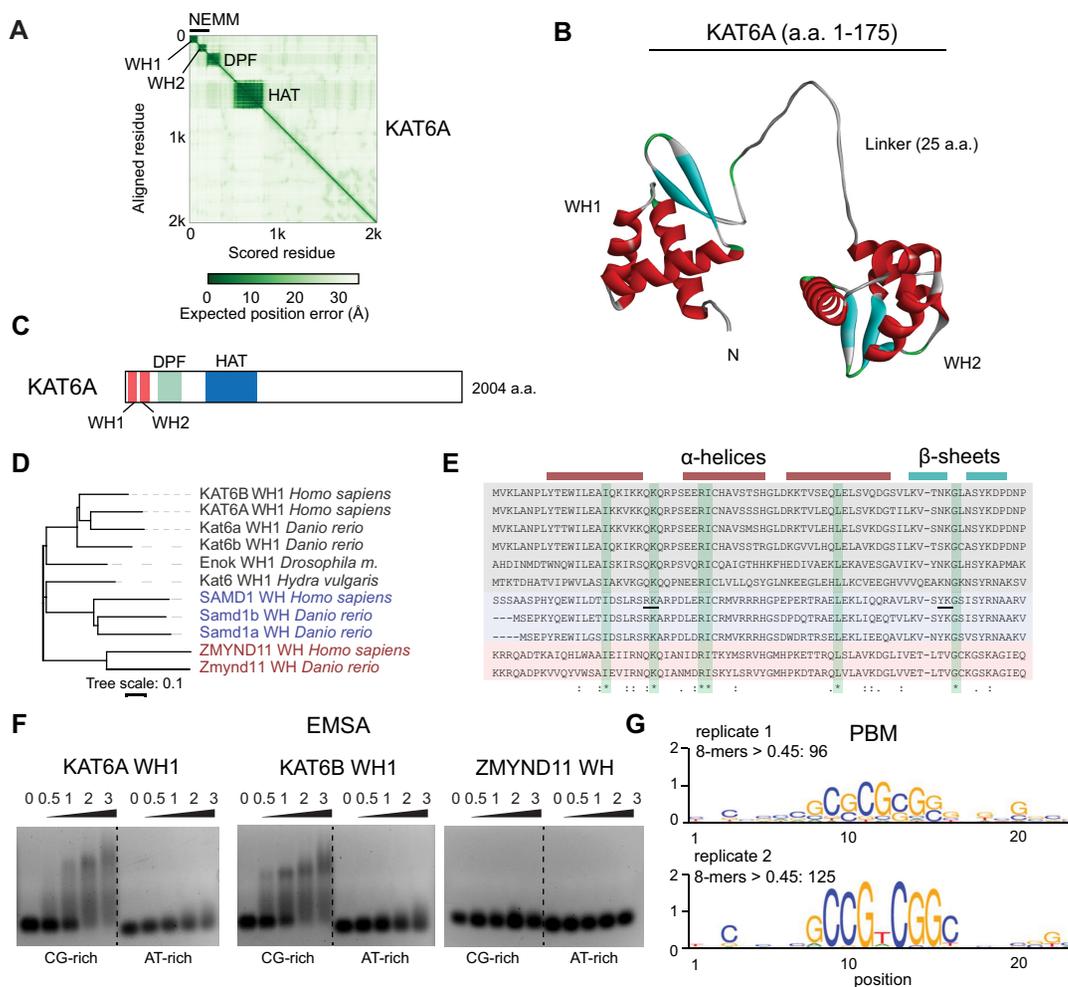


Figure 1. (A) AlphaFold predicted alignment error for KAT6A (65), showing two distinguishable WH domains at the N-terminus, reflecting the NEMM. DPF = double PHD finger. See also Supplementary Figure S1. (B) AlphaFold predicted structure of WH1 and WH2 of KAT6A. (C) Domain structure of KAT6A based on AlphaFold prediction. (D) Phylogenetic tree of WH1 domains with related domains in SAMD1 and ZMYND11. (E) Sequence alignment of the domains from (D). Amino acids that are important for the DNA binding of the WH domain of SAMD1 (36) are underlined. (F) EMSA of the WH1 domains of KAT6A and KAT6B as well as the related WH domain of ZMYND11 using CG-rich and AT-rich DNA. Numbers indicate molar ratio. (G) DNA binding specificity motifs derived from protein binding microarray (PBM) analysis of the KAT6A WH1 domain. The results of two technical replicates are shown (Supplementary Table S3). See also Supplementary Figure S2.

bination of WH1 and WH2 shows a similar binding preference as the WH1 alone (Figure 1G, Supplementary Figure S2C, Table S3), suggesting that WH1 is the major domain responsible for the recognition of CpG-rich DNA motifs.

KAT6A WH1 specifically associates with unmethylated CpG motifs

To understand the molecular basis of CpG recognition by the WH1 domain of KAT6A, we crystallized KAT6A WH1 with a 5'-GCGC-3'-containing double-stranded DNA and solved the complex structure at a resolution of 1.5 Å (Supplementary Table S1). In the complex structure, the WH1 domain has the fold of a winged helix domain that contains three α helices at the N-terminal half and two β strands at the C-terminal half (Figure 2A). Wing-like loops between strands β 1 and β 2 (W1) and after β 2 (W2) could also be observed. Both the major groove and the minor groove of the CpG-containing DNA are recognized (Fig-

ure 2A). KAT6A WH1 recognizes the major groove of the DNA molecule mainly through the α 1 helix. The C-terminal end of the α 1 helix inserts into the CpG-containing major groove and makes sequence-specific contacts with the base pairs containing both the unmethylated CpG motif, and the pair following the CpG motif (Figure 2B, C). C7 and its palindromic C8' in the CpG-motif form a hydrogen bond each with the main chain carbonyl oxygen of Gln23 and Lys24 of KAT6A (Figure 2B), respectively. In addition, the C7 base is also recognized by Gln25 through a side chain-mediated hydrogen bond. These main chain atom-mediated hydrogen bonds would bring the CpG-motif and KAT6A much closer, so that methylation of either C7, or C8' or both would result in a steric clash between two molecules that leads to disrupted interaction. This kind of CpG-motif binding mode is also observed in most of the other CpG island binding domains, verifying that the WH1 domain of KAT6A is a bona fide unmethylated CpG motif binding domain. In addition to the cytosines, both

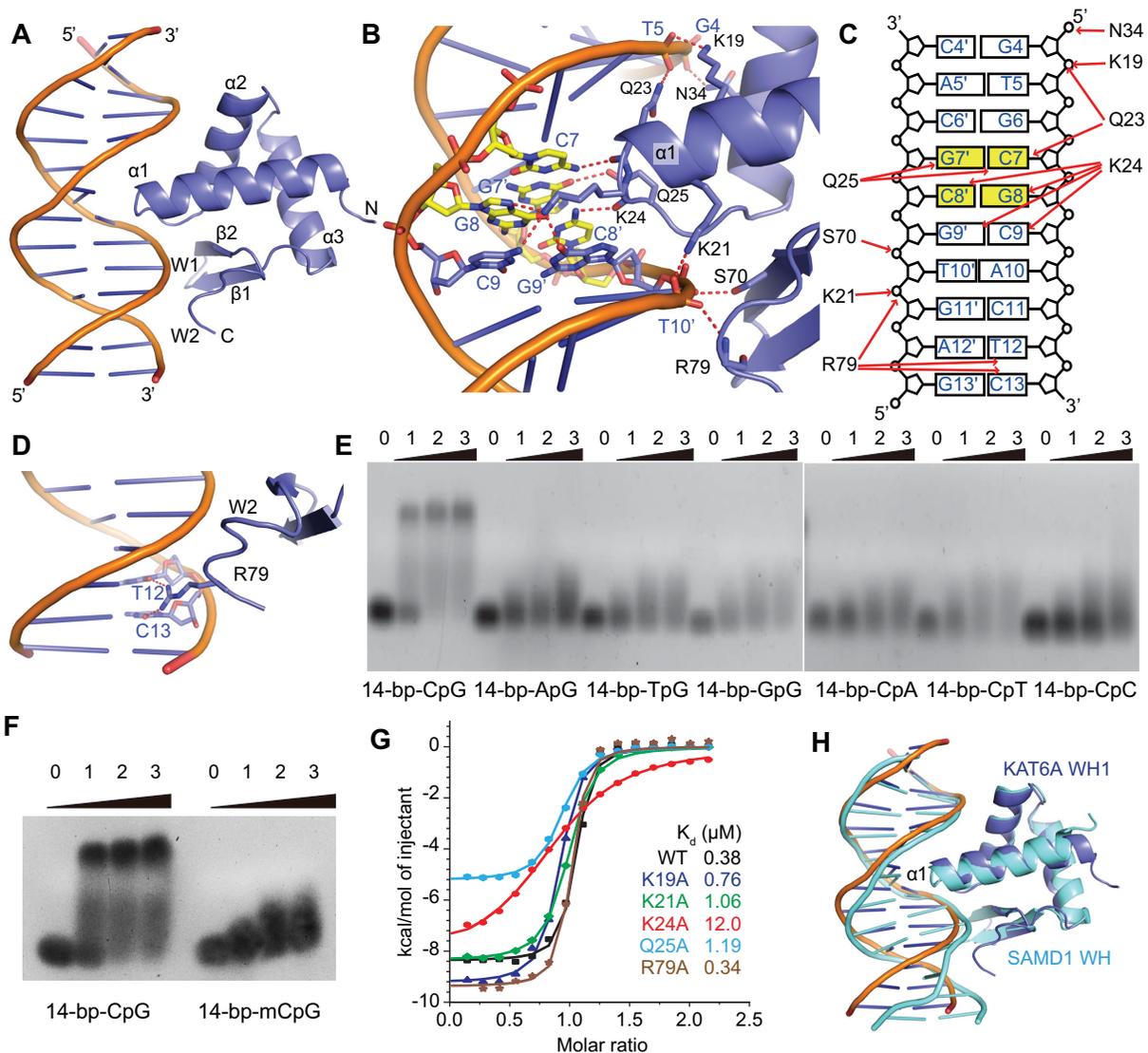


Figure 2. (A) The overall structure of the KAT6A WH1 and DNA complex (PDB: 7Y43). KAT6A is colored blue, and DNA molecules are colored orange. (B) Details of the major groove recognition by KAT6A WH1. (C) A schematic representation of the interactions between KAT6A WH1 and CpG-containing DNA. (D) Minor groove recognition by KAT6A WH1. (E) EMSA results of KAT6A WH1 with wild-type or CpG-mutated DNA molecules. Numbers indicate molar ratio. (F) EMSA results of KAT6A WH1 with unmethylated or methylated CpG-containing DNAs. Numbers indicate molar ratio. (G) ITC-based measurements of wild-type or KAT6A WH1 mutants with the 14-bp-CpG DNA. (H) Superimposed structures of the DNA bound KAT6A WH1 and SAMD1 WH complexes. The KAT6A/DNA complex is colored as in (A), and the SAMD1/DNA complex is colored in cyan.

guanines of the CpG motif are also recognized, with the G7' base forming a hydrogen bond with the side chain of Gln25, and G8 forming a hydrogen bond with the side chain of Lys24 (Figure 2B). Lys24 also recognizes the base pair following the CpG motif by forming a hydrogen bond each with both C9 and G9' through its long side chain (Figure 2B).

In addition to the above base-specific recognition within the major groove, the phosphate backbones of both sides of the major groove are also recognized. The phosphate backbone of G4 forms a hydrogen bond with the side chain of Asn34, and its following base T5 forms a hydrogen bond each with the side chain of Lys19 and Gln23 (Figure 2B), respectively. On the other side of the major groove, the phosphate backbone of G9' forms a hydrogen bond with Ser70,

and its neighboring base T10' forms a hydrogen bond each with both Lys21 and Arg79 (Figure 2B). Overall, KAT6A WH1 contacts the CpG-containing major groove at a 7-bp footprint, among which base-specific recognition is centered at the 5'-CGC-3' motif (Figure 2C). KAT6A contacts the minor groove of the DNA through the W2 loop, with the side chain of Arg79 reaching inside the minor groove and forming a hydrogen bond each with bases T12 and C13 (Figure 2D). The CpG motif recognition at the major groove is essential for the binding, as shown from the EMSA results where KAT6A WH1 displayed dramatically reduced binding affinities towards the CpG-mutated DNA molecules, whose bases in the CpG motif were replaced with other bases (Figure 2E). Similarly, CpG-methylated DNA also lost binding to KAT6A WH1 (Figure 2F). On

the protein side, when the CpG motif-recognizing residues Lys24 or Gln25 were mutated to an alanine, the resultant mutant of KAT6A WH1 displayed a 32-fold or 3-fold weaker binding affinity (Figure 2G), respectively, as calculated by ITC-based measurements. K19A or K21A mutations that disrupted the recognition of the phosphate backbone of the major groove also weakened the binding affinity of KAT6A for the CpG-containing DNA substrate by ~2.0–2.8-fold (Figure 2G). In contrast, recognition at the minor groove does not contribute much to the recognition, as when the minor groove-recognizing residue Arg79 was mutated to an alanine, the resultant mutant did not show noticeable change of binding affinity towards the 14-bp-CpG DNA (Figure 2G). Overall, the DNA-recognition mode by KAT6A WH1 is similar to that of SAMD1 WH (36), as both proteins recognize the unmethylated CpG motif mainly through the C-terminal end of the $\alpha 1$ helix (Figure 2H).

KAT6A WH1 can bind as a dimer to palindromic CCGNCGG motifs

In our PBM experiments KAT6A WH1 showed a high preference for the CCGNCGG motif (Figure 1G, Supplementary Table S3). To understand the recognition mechanism for this palindromic motif, we crystallized the complex of KAT6A WH1 with bound CCGTCGG-containing DNA and solved the structure at 1.93 Å resolution (Supplementary Table S1). In the complex structure, two KAT6A WH1 molecules bound one double stranded DNA in a head-to-head manner (Figure 3A). The dimer form of KAT6A WH1 is not only mediated by the DNA they bind, but also by direct protein-protein interactions between the dimer. Due to the closely positioned two oppositely directed CGG-motifs recognized by both KAT6A WH1, Glu30 from one molecule forms a hydrogen bond each with Ser28, Glu29 and Glu30 from another molecule, respectively (Figure 3B). The dimer form of KAT6A on DNA is very stable in solution, as verified by gel-filtration analysis (Figure 3C). The DNA recognition mode of the dimer-formed KAT6A is very similar to that of the monomer-formed KAT6A, as shown from the overlapped structures of both complexes (Figure 3D). In the overlapped structures, most parts of the KAT6A WH1 molecules were very well superimposed. Noticeable differences occurred mainly at the W1 and W2 loops (Figure 3D), both of which do not play a major role in the DNA recognition. In the dimer-formed complex structure, the CGG-motif was specifically recognized (Figure 3E, F), comparable with the monomer-formed structure in which only the CGC-motif was specifically recognized. The CGG-motif in the dimer-formed complex was recognized by three key residues, Gln23, Lys24 and Gln25, each of which formed 1–3 hydrogen bonds with the bases from the CGG-motif, similarly as the CGC-motif recognition mode in the monomer-formed KAT6A complex. Noticeable differences occurred mainly at the nonspecific DNA backbone recognition. In the dimer-formed complex, only Arg26 could be observed to recognize the phosphate backbone of one side of the CGG-motif (Figure 3E, F), while interactions with the other parts of the phosphate backbone and with the minor groove could not be observed, further

verifying that the CGG/CGC-motifs are the most critical sequences required for the recognition of KAT6A WH1.

KAT6A influences developmental pathways related to heart and neuronal development in mouse ES cells

Previous work suggested an important role of KAT6A in a variety of developmental processes (1,7,8,27). To gain further insight into the potential gene regulatory function of KAT6A, we used first mouse ES cells (mESCs) as a model. Using CRISPR/Cas9, we created KAT6A KO mESCs, which were validated by Western blotting, immunofluorescence and Sanger sequencing (Figure 4A, B; Supplementary Figure S3A). Consistent with the role of KAT6A as histone acetyltransferase, we observed reduced global levels of the histone acetylation marks H3K9ac and H3K14ac in the knockout cells (Figure 4C). These cells showed slightly increased proliferation (Supplementary Figure S3B) and they had an altered morphology in comparison to control cells (Supplementary Figure S3C). Using RT-qPCR, we confirmed the dysregulation of several previously described KAT6A target genes, such as *Rhox6*, *Skidal* and *Sox9* (27) (Figure 4D). In colony formation assays, KAT6A deletion led to fewer colonies (Figure 4E), suggesting a reduced ability to grow out from single-cell clones.

To address the impact of KAT6A deletion on the gene expression pattern at a global level, we performed RNA-Seq using two independent KAT6A KO clones. Principal component analysis demonstrated that the KO cells differed greatly from the control cells (Figure 4F). Consistent with an activating role of KAT6A as a histone acetyltransferase, we observed more down- ($n = 494$; \log_2 -fold change > 0.75 ; $P < 0.01$) than upregulated genes ($n = 238$) upon KAT6A depletion (Figure 4G). Gene set enrichment analysis (GSEA) revealed that gene sets related to development and signaling pathways were mostly downregulated upon KAT6A deletion (Supplementary Figure S3D). Conversely, gene sets related to brain development were upregulated (Supplementary Figure S3D). The dysregulation of developmental and signaling pathways due to impaired KAT6A function is consistent with the observed defects in patients with KAT6A syndrome (15). Closer inspection of the data further revealed that KAT6A deletion influences the gene expression pattern, similar to a pattern after deletion of the PRC2 core component SUZ12 (35,67). Namely, genes that were downregulated upon SUZ12 deletion were mostly downregulated in KAT6A KO cells, as well (Supplementary Figure S3D, E). The opposite was the case for genes that were upregulated upon SUZ12 deletion (Supplementary Figure S3D, E). This phenomenon can likely be explained by the downregulation of SUZ12 at both the mRNA and protein level in the KAT6A KO cells (Supplementary Figure S3F, G). We also confirmed reduced EZH2 recruitment to some PRC2 target genes via ChIP-qPCR (Supplementary Figure S3H). Thus, the observed gene expression changes after KAT6A deletion in mouse ES cells are likely due to direct effects caused by impaired histone acetyltransferase activity, and indirect effects, such as dysregulation of key chromatin regulators, such as SUZ12.

Analysis of our RNA-Seq data further demonstrated that the downregulated but not the upregulated genes were sig-

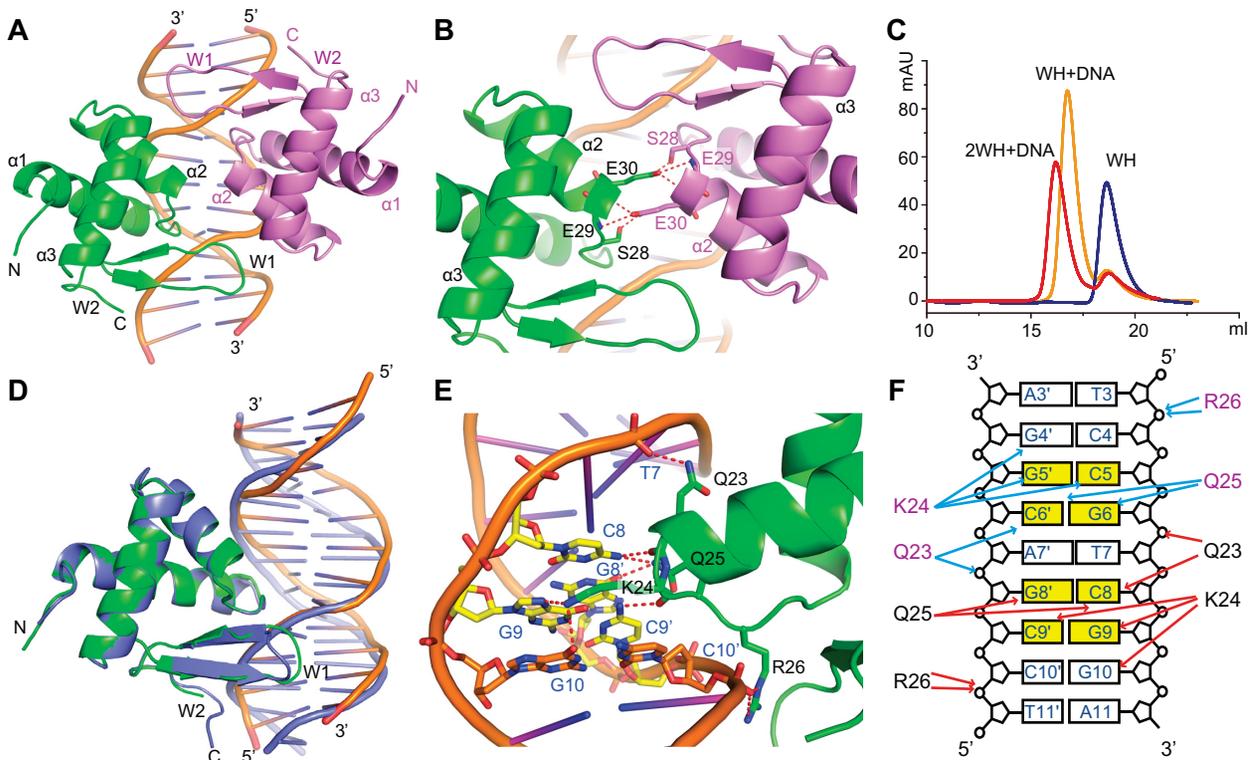


Figure 3. (A) The overall structure of the dimer-formed KAT6A WH1 and DNA complex (PDB: 8H7A). The KAT6A dimer is colored green and purple, respectively. DNA molecules are colored orange. (B) Interactions between the dimer molecules of KAT6A WH1. (C) Gel-filtration profiles of the KAT6A WH1 alone (blue), or with one CpG-motif containing DNA (orange), or with DNA containing two CpG-motifs (red). (D) Superimposed structures of the DNA bound monomer-formed KAT6A WH1 and dimer-formed KAT6A complexes. The monomer-formed KAT6A/DNA complex is colored in blue, and one copy of the dimer-formed KAT6A and its bound DNA are colored in green and orange, respectively. (E) Details of the major groove recognition by one molecule of the dimer-formed KAT6A WH1. (F) A schematic representation of the interactions between the dimer-formed KAT6A WH1 and palindromic CpG-containing DNA.

nificantly enriched for CGIs at their promoters (Figure 4H). This result supports that KAT6A may bind to CGI-containing promoters in mouse ES cells for gene activation, and that the WH1 domain could play a role in the chromatin binding of KAT6A in these cells. However, in mouse ES cells endogenous KAT6A was difficult to detect by Western blotting (Figure 4A), and we failed to immunoprecipitate KAT6A from chromatin in these cells. Additionally, our attempts to ectopically express KAT6A in mouse ES cells were not successful. Consequently, an investigation of the genome-wide binding pattern of KAT6A as well as an investigation of the role of the KAT6A WH1 domain was not possible in these cells.

KAT6A WH1 is required but not sufficient for chromatin association

As an alternative model, we chose human HEK293 cells, which can easily be transfected with KAT6A constructs (Figure 5A, Supplementary Figures S5A and S6A) and are therefore more suitable for the biochemical characterization of KAT6A chromatin recruitment mechanisms. Using these cells, we addressed the consequences of abrogating the DNA binding function of the WH1 domain by mutating the two most important amino acids lysine 24 and glutamine 25 to alanines (K24A/Q25A, ‘WH1 mut’) (Figure

2G). Via cellular fractionation experiments, we observed that ectopically expressed wild-type KAT6A in HEK293 cells was found in the chromatin fraction, but also in the nucleoplasm and in the cytoplasm (Figure 5B). Mutating the WH1 domain of KAT6A led to reduced levels of the protein in the chromatin fraction but increased levels in the nucleoplasm and cytoplasm fractions (Figure 5B, C), suggesting that chromatin binding is impaired. Via ChIP-qPCR experiments using a FLAG-antibody, we confirmed strong chromatin binding of KAT6A wild-type to CGI-containing promoters, which were selected based on previously published KAT6A ChIP-Seq data (11). Importantly, the KAT6A WH1 mutant showed almost no chromatin association with the investigated CGIs (Figure 5D), suggesting that mutating the DNA binding amino acids of KAT6A is sufficient to prevent its binding to CGIs. Thus, these results support the importance of a functional WH1 domain for the chromatin binding of KAT6A.

A previous report suggested that the KAT6A WH1 domain alone is sufficient for proper chromatin targeting (11). To investigate this aspect, we performed ChIP-qPCR experiments with only the WH1 domain of KAT6A. We found that neither the wild-type WH1 nor the mutant WH1 showed strong chromatin association (Supplementary Figure S4A–C), suggesting that the WH1 domain alone is incapable of binding to chromatin. This observation is in

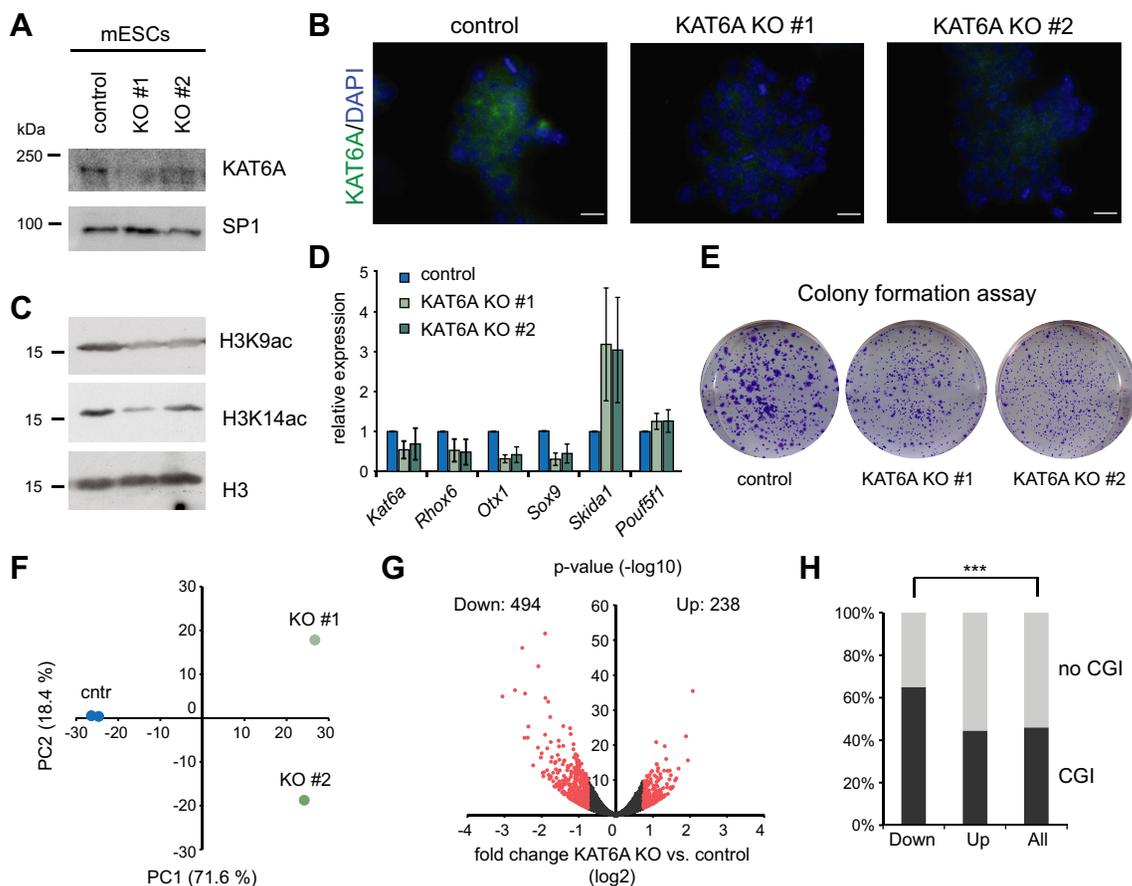


Figure 4. (A) Western blot of KAT6A KO mouse ES cell clones. (B) Immunofluorescence of endogenous KAT6A in control and KAT6A KO cells. Scale bar = 20 μm. (C) Western blot of H3K9ac and H3K14ac upon KAT6A KO. (D) Gene expression changes of known KAT6A target genes upon KAT6A deletion. (E) Colony formation assay of control and KAT6A KO mouse ES cells. (F) Principal component analysis (PCA) of RNA-Seq data. (G) Volcano plot of gene expression changes upon KAT6A deletion in mouse ES cells. Genes in red indicate genes with \log_2 -fold change >0.75 and $P < 0.01$. (H) Distribution of CGI-containing genes in the up- and downregulated genes, compared to all genes. Significance was determined using a hypergeometric test. *** $P < 0.001$.

line with the relatively low DNA binding affinity of the WH1 domain (Figure 2G) and is similar to the WH domain of SAMD1, which is also not able to bind to chromatin in isolation (36). To address, which additional regions of KAT6A are required for chromatin binding, we investigated KAT6A proteins that were extended after the WH1 domain (Supplementary Figure S4D–F). We observed that addition of the WH2 and the DPF domain stepwise increased the chromatin binding of KAT6A, suggesting that each of these domains may contribute to the chromatin binding of KAT6A (Supplementary Figure S4E). A protein comprising the WH1, WH2 and the DPF binds to chromatin at a similar level as the full-length protein, suggesting that this region is the main chromatin binding module of KAT6A. Interestingly however, deletion of WH2 or the DPF from the full-length KAT6A led to only subtly reduced chromatin binding of KAT6A, which contrasts to mutation of the WH1 (Supplementary Figure S4G–I). This suggests that the WH1 is the most crucial domain for the association of KAT6A with CGIs, while both the WH2 and the DPF domain are likely also involved but less essential for this function of KAT6A. Possibly, also regions outside of the globular domains, or proteins that bind to

these regions, may contribute to the chromatin binding of KAT6A.

Together we conclude that the WH1 is necessary but not sufficient for the chromatin binding of KAT6A.

Overexpression of mutant KAT6A has a dominant-negative effect on histone acetylation

Next, we assessed the consequence of overexpressing the KAT6A WH1 mutant on the endogenous KAT6A and on histone acetylation. For this purpose, we used a KAT6A antibody that recognizes both the endogenous and the ectopically expressed KAT6A (Supplementary Figure S5A), and antibodies against H3K9ac and H3K14ac, the main enzymatic targets of KAT6A (21,26,41). We found that overexpression of wild-type KAT6A did not significantly increase the total KAT6A level on chromatin (Figure 5E). This observation suggests that the KAT6A chromatin binding is already saturated and cannot be further enhanced. Consistently, we observed only a minor influence on histone acetylation (Figure 5E). Interestingly, however, we observed a decreased level of chromatin bound KAT6A when overexpressing the KAT6A WH1 mutant (Figure 5E). This find-

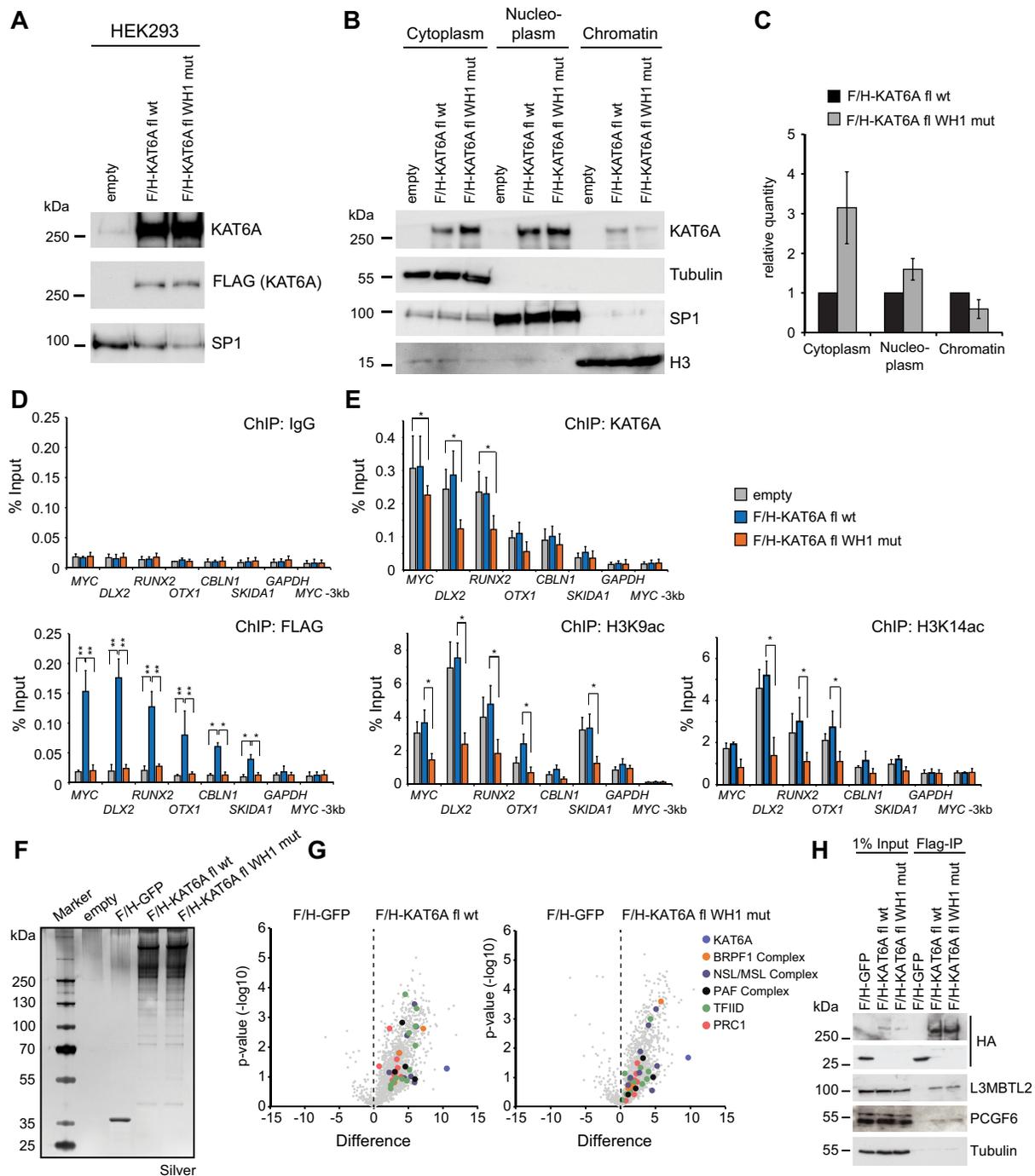


Figure 5. (A) Western blotting of ectopically overexpressed KAT6A wild-type and WH1 mutant in HEK293 cells. (B) Representative Western blot of cellular fractionation experiments of cells from (A). (C) Quantification of experiments in (B) ($n = 3$). (D) ChIP-qPCR of cells from (A) using a FLAG-antibody. IgG is shown as control. (E) ChIP-qPCR using antibodies against KAT6A, H3K9ac and H3K14ac. (F) Silver staining of FLAG-purified KAT6A wild-type and WH1 mutant from HEK293 cells. GFP was used as a control. (G) Volcano plot of the proteins identified by mass spectrometry. Putative interaction partners of KAT6A are marked by colored dots. (see Supplementary Figure S6B and Table S4 for details). (H) Coimmunoprecipitation confirming the interaction of KAT6A wild-type and WH1 mutant with two members of the PRC1 complex. In (D) and (E), data represent the mean \pm s.d. of two biological replicates. The significance was evaluated via a two-tailed unpaired Student's *t*-test, comparing the results to wild-type KAT6A. * $P < 0.05$, ** $P < 0.01$.

ing cannot be explained by an altered protein level of endogenous KAT6A, which remained constant (Supplementary Figure S5A). Instead, it suggests an impairment of the chromatin binding of the endogenous KAT6A. In line with this observation, we also observed a reduction in H3K9ac and H3K14ac (Figure 5E), suggesting that overexpressing the KAT6A WH1 mutant has a dominant negative effect. A similar dominant negative effect can be seen when overexpressing an enzymatically inactive KAT6A HAT domain mutant (Q654E/G657E, ‘HAT mut’) (41), which binds as efficiently to chromatin as the wild-type protein (Supplementary Figure S5B, C). These results suggest that overexpression of either the WH1 mutant or the HAT mutant diminishes the function of endogenous KAT6A.

KAT6A interacts with multiple chromatin regulatory protein complexes in a DNA binding independent manner

Why does endogenous KAT6A chromatin binding decrease upon overexpression of the KAT6A WH1 mutant? Given its abrogated DNA binding, the WH1 mutant KAT6A is unlikely to interfere with the chromatin binding of the endogenous KAT6A, which can therefore not explain the dominant negative effect. An alternative explanation is that the overexpressed KAT6A WH1 mutant protein competitively displaces the endogenous KAT6A from its interaction partners. Consequently, the endogenous KAT6A has fewer interacting partners, which could impair its chromatin association. To assess the interactome of the mutant KAT6A, compared to wild-type KAT6A, we ectopically expressed both proteins in HEK293 cells and performed FLAG-immunoprecipitation. Silver staining of the precipitated material showed that both proteins pulled down many proteins (Figure 5F). Consistent with silver staining, we found by mass spectrometry that numerous proteins were enriched with both the KAT6A wild-type and WH1 mutant (Figure 5G, Supplementary Table S4). Notably, in addition to known interaction partners such as the BRPF1 complex (BRPF1, MEAF1, MLLT1 (ENL)), we also identified members of the PRC1 complex (Figure 5G). This association has previously been proposed in *Drosophila* (29,30), suggesting a conserved interplay of the KAT6 histone acetyltransferase proteins with the Polycomb system. Via coimmunoprecipitation, we confirmed that KAT6A wild-type and WH1 mutant interact with members of the human PRC1 complex in HEK293 cells (Figure 5H). Additionally, other putative interaction partners, such as the PAF1 and the NSL complex, showed a similar enrichment with wild-type and WH1 mutant KAT6A in the mass-spectrometry data (Figure 5G, Supplementary Figure S6B), suggesting that the mutation of the WH1 domain does not strongly influence the capacity of KAT6A to associate with its interaction partners.

Thus, this finding is in line with the idea that the ectopic overexpression of the KAT6A WH1 mutant can competitively replace the endogenous KAT6A from its interaction partners. The endogenous KAT6A could therefore have fewer chromatin binding interaction partners, leading to the observed reduced chromatin association. Furthermore, our mass-spectrometry results suggest that the reduced chromatin association of the KAT6A WH1 mutant

is likely not caused by an altered interactome but mainly by the abrogated DNA binding capacity.

KAT6A localizes genome-wide to unmethylated CpG islands and gene bodies

To address the consequences of mutating the WH1 domain of KAT6A at a genome-wide level, we performed ChIP-Seq experiments. For this purpose, we overexpressed KAT6A wild-type, WH1 mutant, and HAT mutant and precipitated the proteins using a FLAG antibody. To assess the effect on histone acetylation, we investigated the H3K9ac mark. As a negative control, we used HEK293 cells transfected with an empty vector.

First, we analyzed the binding pattern of the wild-type KAT6A, for which we identified 20,903 significant peaks. Consistent with the hypothesis that KAT6A is recruited to CpG islands, we found a strong overlap of these peaks with CGIs (Figure 6A). Given that most CGIs are at promoters, we found a strong enrichment of KAT6A at promoters (Figure 6B, C). Motif analysis of the KAT6A bound locations revealed the enrichment of CpG-rich motifs (Figure 6D), which are similar to the motifs identified via the protein binding microarray (Figure 1G).

To explore whether the sequence preference of the isolated WH1 domain contributes to KAT6A localization, we systematically examined for all possible four nucleotide sequences (4-mers) the correlation of *in vitro* binding of the isolated WH1 domain to synthetic sequences on PBMs with *in vivo* binding of the tagged full-length protein to genomic sequence, controlling for CpG content of the tested sequences (see Methods). The association of a given 4-mer with *in vitro* and *in vivo* binding is significantly correlated (Pearson's $R = 0.31$, $P < 0.001$, Figure 6E), and the 4-mer most highly correlated with binding *in vitro* (GCCG) is likewise mostly highly correlated with binding *in vivo*. Moreover, instances of this 4-mer are significantly biased toward the summit of KAT6A peaks (Figure 6F), consistent with direct targeting of this sequence *in vivo*. Our PBM and structural analysis further suggests that KAT6A can bind to the palindromic CCGNCGG motif as a dimer (Figures 1G and 3). To investigate whether the occurrence of such a motif may enhance the chromatin binding of KAT6A, we analyzed the binding profile of KAT6A around these motifs. Indeed, we found that at CCGNCGG motifs KAT6A levels are locally enriched, compared to motifs that have the same CG content but a different sequence or spacing (Supplementary Figure S7A–C). This observation suggests that KAT6A has a binding preference for this palindromic sequence. However, ~40% of KAT6A bound CGIs do not possess such a motif (Supplementary Figure S7D), suggesting that the presence of a dimeric binding motif is advantageous but not essential for KAT6A binding.

Using publicly available CpG island recovery assay (CIRA)-Seq data (11) and whole genome bisulfite sequencing (WGBS) (56), we found that KAT6A mainly binds to unmethylated CGIs, while the KAT6A-unbound CGIs are mostly methylated (Figure 6G). Thus, consistent with the EMSA experiment (Figure 2F), these results support that KAT6A is repelled by methylated DNA. To investigate whether removing DNA methylation from methylated

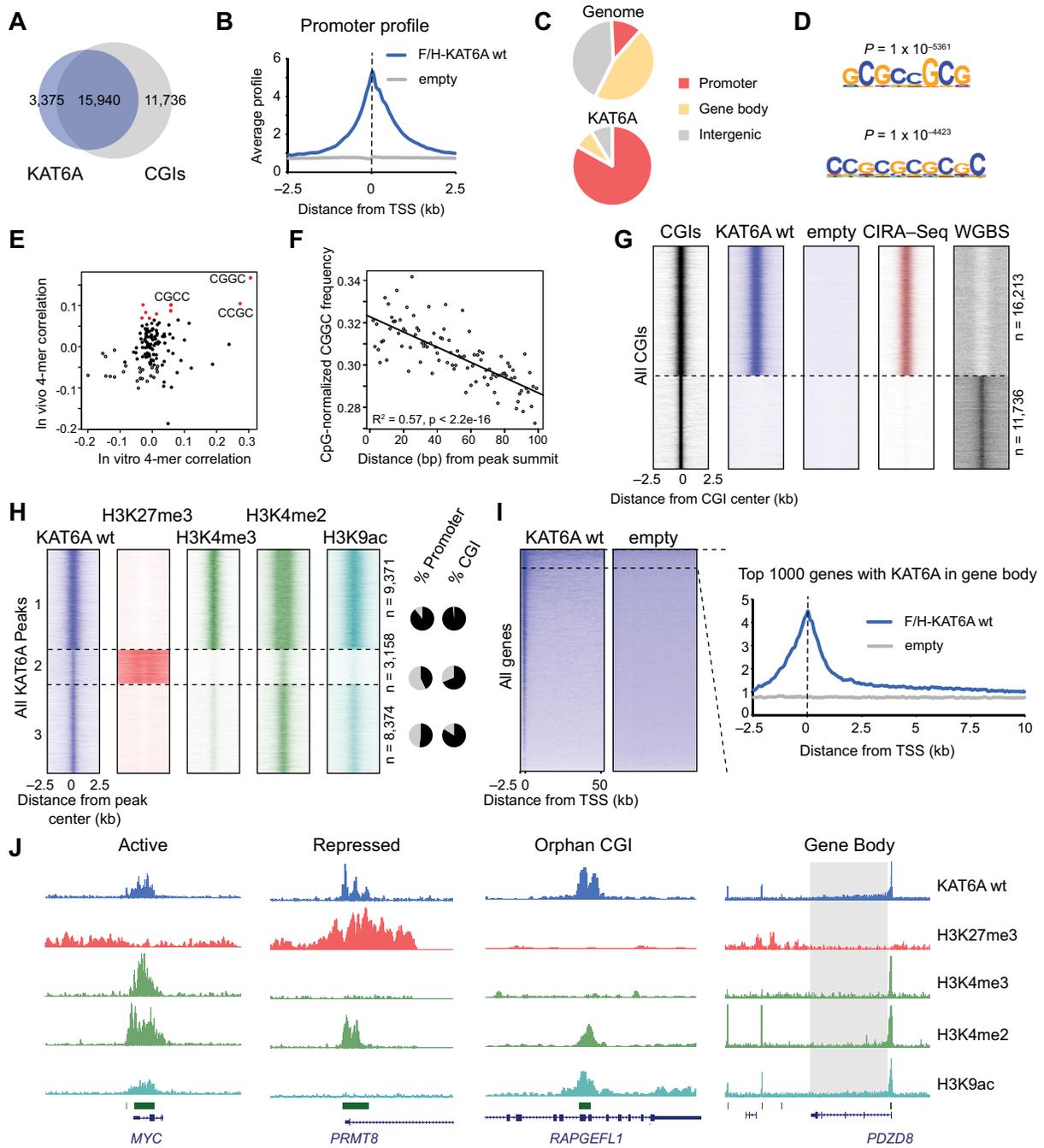


Figure 6. (A) Venn diagram showing the overlap of KAT6A peaks with CGIs. (B) Promoter profile of KAT6A compared to the empty vector control. (C) Genomic distribution of KAT6A bound locations compared to the entire genome. (D) Enriched motifs at KAT6A bound locations. Compare with Figure 1G. (E) Correlation of 4 nt sequences (4-mers) with in vitro and in vivo binding. The 4-mer CGGC is most correlated in both analyses. (F) Density of the most enriched 4-mer CGGC is significantly correlated with proximity to peak summit (Pearson's $R = 0.86, P < 0.001$). (G) Heatmaps of all CGIs separated by KAT6A binding, in comparison to CpG island recovery assay (CIRA)-Seq (11) and whole-genome bisulfite sequencing (WGBS) (56). (H) Clusters of KAT6A bound locations, grouped by enrichment for either H3K4me3, H3K27me3 or neither. (I) Heatmap and profile showing KAT6A enrichment in the gene bodies of a subset of genes. (J) Examples of KAT6A bound locations.

CGIs leads to a de novo recruitment of KAT6A, we used the DNMT1 inhibitor GSK-3484862, which has been described to reduce the level of DNA methylation genome-wide (43). Based on available DNA methylation data (56) and our KAT6A ChIP-Seq, we selected several KAT6A-unbound gene promoters with methylated CGIs, and investigated those genes via ChIP-qPCR (Supplementary Figure S8A). After treating the cells with the inhibitor for 3 days, we found that the levels of KAT6A remained constantly high at CGIs which were already unmethylated under untreated conditions (Supplementary Figure S8B). At CGIs which were methylated under untreated conditions, we observed a significant increase of KAT6A upon DNMT1 inhibition in a dose dependent manner (Supplementary Figure S8B). This result suggests that at those CGIs a reduced DNA methylation, due to DNMT1 inhibition, leads to the de novo recruitment of KAT6A. This finding further supports that KAT6A's binding ability to CGIs is dependent on their DNA methylation status.

Consistent with the preference for unmethylated CGIs, most KAT6A bound locations are decorated by the active H3K4me3 and H3K9ac histone marks (group 1; $n = 9351$), suggesting that KAT6A typically plays a role at actively transcribed promoters (Figure 6H, J). The genes associated with this location are linked to general cellular processes, such as ribosome biogenesis and metabolism (Supplementary Figure S9A). Another subset of KAT6A-targeted loci (group 2; $n = 3158$) is decorated by repressive H3K27me3 (Figure 6H, J), supporting that KAT6A may also play a role at locations that are in a repressed state. We also identified a subset of KAT6A-targeted locations that were decorated by neither H3K4me3 nor H3K27me3 (group 3; $n = 8374$) (Figure 6H, J). Although many of these locations are not at promoters, they still possess CGIs. Since these locations are often decorated by the enhancer mark H3K4me2, KAT6A may bind to orphan CGIs, which have been proposed to function as enhancers (68). These locations are also enriched for H3K9 acetylation (Figure 6H, J). Genes associated with group 2 and group 3 are mostly linked to processes related to kidney development (Supplementary Figure S9A), suggesting the involvement of these genes in developmental processes. Of course, kidney development specifically reflects the origin of HEK293 from embryonic kidneys. Interestingly, consistent with the mass spectrometry results, we found a strong overlap of KAT6A with members of the PRC1 in HEK293 cells (Supplementary Figure S6C, D) (60), raising the possibility that these proteins cooperate to regulate transcription in these cells.

Further inspection of the data showed that the wild-type KAT6A was also enriched in the gene body of a small subset of genes (Figure 6I). Closer inspection of these gene bodies suggests that KAT6A binding positively correlates with the presence of the histone elongation marks H3K4me1 and less strongly with H3K36me3 and RNA Polymerase II (Supplementary Figure S9C, D). Consistently, the associated genes are highly expressed (Supplementary Figure S9B). Furthermore, gene ontology analysis demonstrated that these genes are linked to the cell cycle (Supplementary Figure S9A), suggesting that the cell cycle and active transcription may play a role for the recruitment of KAT6A to the gene bodies. Interestingly, KAT6A bound gene bod-

ies are characterized by a lower GC content (Supplementary Figure S9C,D), making it likely that KAT6A is recruited to these loci in a WH1 independent manner. Possible mechanisms that may contribute to the association of KAT6A with gene bodies could include the interaction with transcription elongation factors, such as the PAF complex or ENL (Supplementary Figure S6B) (41), interaction of KAT6A with specific histone marks via the double PHD finger (61,62), or DNA binding functions of KAT6A's HAT domain (69) and KAT6A's interaction partner BRPF1 (70). Given that we used ectopically expressed KAT6A, it will be interesting to assess whether endogenous KAT6A also binds to gene bodies.

In summary, our investigations of the genome-wide binding pattern of KAT6A indicate that KAT6A is mostly recruited to unmethylated CpG islands in the genome.

Mutation of the WH1 domain impairs KAT6A recruitment to CGIs but not to gene bodies

Based on our ChIP-qPCR experiments mutating the WH1 domain of KAT6A almost completely abrogated chromatin binding at the investigated genes (Figure 5D). On the contrary, cellular fractionation experiments suggested that the mutated KAT6A still bound to some extent to chromatin (Figure 5B). To investigate whether the KAT6A WH1 mutant has some residual chromatin binding or may be recruited to noncanonical locations, we investigated the chromatin association of overexpressed KAT6A wild-type and WH1 mutant at the genome-wide level.

Consistent with the hypothesis that the WH1 domain is required for the recruitment of KAT6A to CGIs, we found that the KAT6A WH1 mutant is completely absent from CGIs (Figure 7A, B, E). In contrast, mutation of the HAT domain does not strongly influence the chromatin binding pattern of KAT6A (Figure 7A, B, E). Unexpectedly, although the WH1-mutated KAT6A does not bind to CGIs, it is still enriched in the gene bodies (Figure 7C–E), suggesting that KAT6A recruitment to gene bodies is independent of the DNA binding of the WH1. A direct comparison of the KAT6A levels in the gene bodies shows that the WH1 mutant KAT6A even has a slightly increased occupancy in these regions compared to the wild-type KAT6A (Figure 7D). Possibly, this increase can be explained by an elevated concentration of the KAT6A mutant in the nucleoplasm (Figure 5B, C), leading to an enhanced association with locations that can still be targeted by the mutant.

Thus, these results suggest that the WH1 domain is required for the chromatin recruitment of KAT6A to CGIs but not to gene bodies.

Overexpression of mutant KAT6A leads to reduced H3K9 acetylation genome-wide

In ChIP-qPCR experiments, we observed that ectopic overexpression of the KAT6A wild-type only minimally increased histone acetylation (Figure 5E), suggesting that KAT6A overexpression is not able to further enhance the histone acetylation levels in HEK293 cells. In contrast, overexpression of the KAT6A WH1 mutant and the HAT mutant led to reduced levels of histone acetylation at the

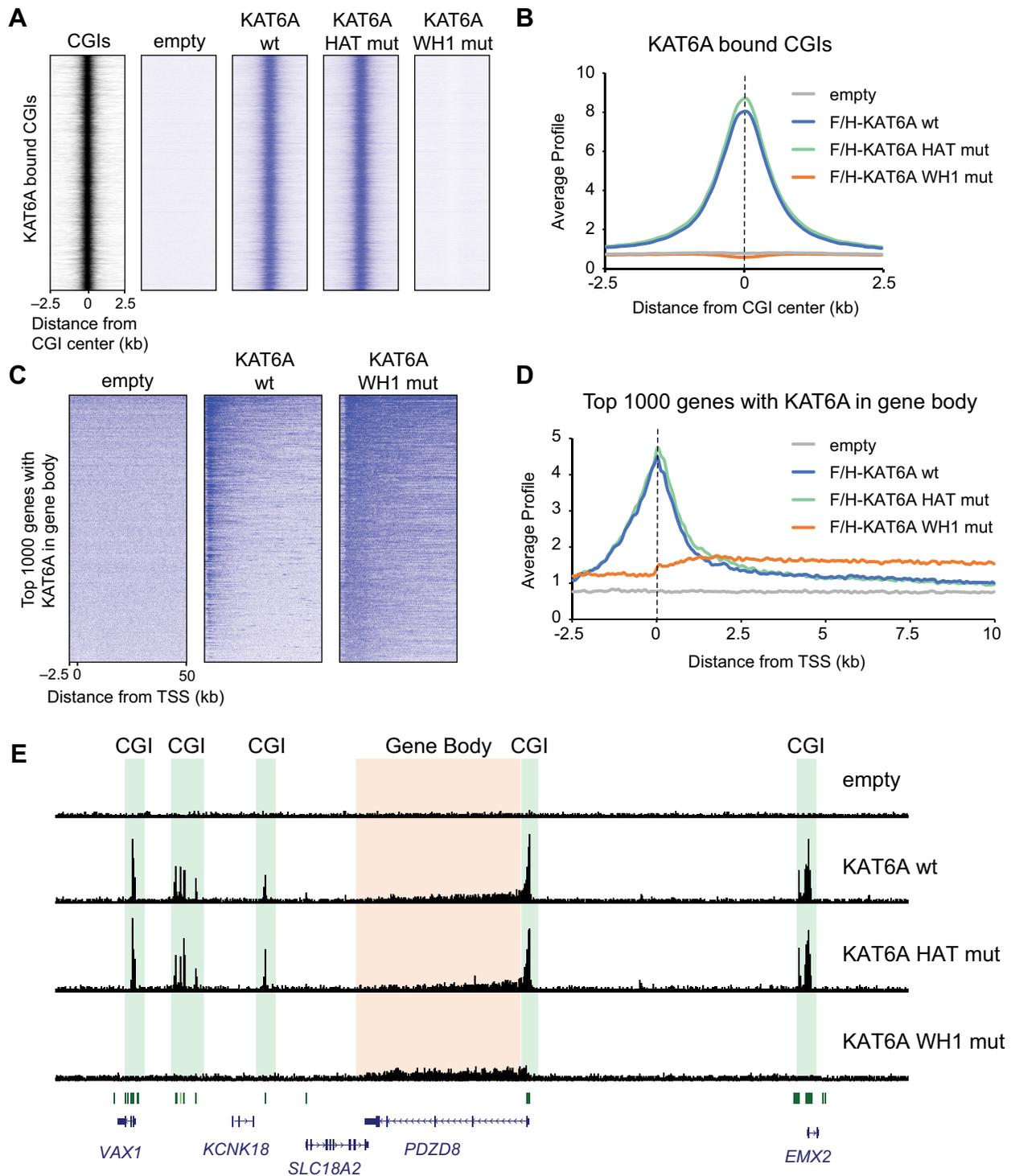


Figure 7. (A) Heatmaps of the three KAT6A variants (wild-type, WH1 mut, HAT mut) at KAT6A bound CGIs. (B) Profile of the three KAT6A variants at KAT6A bound CGIs. (C) Heatmaps of KAT6A wild-type and WH1 mutant at all genes sorted after KAT6A level in the gene body. (D) Profile of the three KAT6A variants at the top 1000 genes with KAT6A in the gene body. (E) Genome browser view showing ChIP-Seq results of the three KAT6A variants compared to the empty vector control. At CGIs (green) the binding of KAT6A is lost upon mutation of the WH1 domain. Binding to gene bodies (orange) is not affected.

investigated target genes (Figure 5E, Supplementary Figure S5C), likely due to dominant negative effects. Importantly, we confirmed these results at a genome-wide level via ChIP-Seq experiments for H3K9ac, although the effects were overall subtle (Supplementary Figure S10A–E). The reduction in the H3K9ac level was more severe upon overexpression of the KAT6A HAT domain mutant (Supplementary Figure S10B–D), while the KAT6A WH1 mutant had only mild effects. Consistent with the relatively gentle consequences on histone acetylation levels, we observed no substantial changes in the gene expression of KAT6A target genes (Supplementary Figure S10F). Interestingly, although the KAT6A WH1 mutant became more strongly enriched in the gene bodies compared to the wild-type protein (Figure 7D), we did not observe an increased level of H3K9ac in these regions (Supplementary Figure S10E). Possible explanations for this observation are that KAT6A affects other histone acetylation marks not investigated here, that mutation of the WH1 domain also impairs the histone acetyl transferase activity of KAT6A or that KAT6A facilitates histone acetylation-unrelated processes in the gene body.

Using our H3K9ac data, we further investigated whether the local enrichment of KAT6A at the palindromic CCGNCGG sequence may also influence histone acetylation levels. Indeed, similar to the enrichment of KAT6A, we also found a local enrichment of H3K9ac at this motif (Supplementary Figure S7A, E). This observation supports that the dimeric binding of KAT6A to the CCGNCGG motifs within unmethylated CGIs could be relevant to regulating the local chromatin status.

Taken together, while overexpression of wild-type KAT6A in HEK293 cells minimally affects genome-wide H3K9ac levels, overexpression of WH1 or HAT mutated KAT6A leads to a mild reduction of H3K9ac, likely due to dominant negative effects.

DISCUSSION

Histone acetylation, established by histone acetyltransferases (HATs), such as KAT6A, KAT6B, p300, CBP and PCAF, plays a fundamental role in the regulation of chromatin compaction and transcriptional activity (71). Most histone acetyltransferases are recruited to their target genes by specific transcription factors or other chromatin binding factors (72), to decompact the chromatin and to activate gene transcription. The main substrates of KAT6A are H3K9 (26), H3K14 (73) and H3K23 (12), where KAT6A not only establishes histone acetylation but also catalyzes further lysine acylation reactions, such as propionylation (12). The N-terminal region of KAT6A, previously termed NEMM, has been implicated in the chromatin association of KAT6A (19), but a specific chromatin binding mechanism of this region has not yet been identified.

Here, we revealed that the NEMM of KAT6A consists of two winged-helix domains, and we demonstrated that the more N-terminal WH1 domain of KAT6A is required for the recruitment of KAT6A to unmethylated CpG islands (Figure 8A). The structure and the DNA binding mode of the WH1 domain of KAT6A are very similar to those of the WH domain of SAMD1 (36). As shown in Figure 2H, the

overlay structure of the DNA bound WH1 of KAT6A superimposed well with that of the DNA bound WH domain of SAMD1. In particular, the C-terminal end of the $\alpha 1$ helix of both proteins was inserted into the CpG-containing DNA major groove, indicating that both proteins recognize the unmethylated CpG motif through the same structural elements. In contrast, SAMD1 contacted the DNA minor groove through the W1 loop, while KAT6A contacted it through the W2 loop, suggesting that the minor groove recognition mechanisms are not conserved between both proteins. This may explain why the minor groove recognition by the SAMD1 WH but not by the KAT6A WH1 contributes significantly to the DNA binding (36).

The WH1 domains of KAT6A and KAT6B are highly similar, and the essential amino acids important for binding the DNA are conserved between KAT6A and KAT6B (Figure 1E). EMSA experiments suggested that the WH1 domain of KAT6B also binds to CpG-rich DNA (Figure 1F). Thus, although we did not investigate the WH1 domain of KAT6B in further detail, it is likely that this WH1 domain fulfills a similar functional role as the WH1 domain of KAT6A. Since the genomic binding pattern of KAT6B has not yet been determined, it will be of interest to assess whether KAT6B is also mainly recruited to unmethylated CpG islands.

The WH1 of KAT6A functions not only as monomer at DNA but our work demonstrated that it can also bind as a dimer to the palindromic sequence CCGNCGG. Our genome-wide data support that KAT6A prefers this sequence at CGIs. It is possible that a dimeric binding of KAT6A to such a motif may lead to an alternative functionality, which could be important for the biological function of KAT6A. Further research will be necessary to clarify the biological relevance of this dimerization ability, and whether mutations/polymorphisms that destroy or create CCGNCGG motifs at specific promoters may have an KAT6A-dependent biological impact.

The DNA binding affinity (K_D) of the WH1 to a single CpG motif is in the micromolar range, consistent with the binding affinities of the CXXC domain of CFPI (74), and the WH1 domains of human PCLs (34) to CpG-motifs. Previous studies suggested that the N-terminal part of KAT6A, comprising the WH1, the WH2 and the DPF domain, is involved in chromatin binding (1). Interestingly, deletion of either the WH2 or the DPF has unexpectedly minor consequences on the chromatin binding capacity of KAT6A (Supplementary Figure S4H). In contrast, mutation of the DNA binding WH1 domain alone is sufficient to completely abrogate the chromatin association of KAT6A with unmethylated CGIs, genome-wide (Figures 5 and 6). This suggests that the WH1 is the most crucial domain for the chromatin association of KAT6A. Consequently, we conclude that the DNA sequence composition and the DNA methylation status are the decisive determinants for KAT6A recruitment. The relevance of the DNA methylation status is supported by the observation that inhibition of the DNA methyltransferase DNMT1 leads to de novo recruitment of KAT6A to methylated CGIs (Supplementary Figure S8B). Nonetheless, the recruitment of the KAT6A WH1 mutant to gene bodies suggests that KAT6A can also be recruited to certain loci without the DNA binding func-

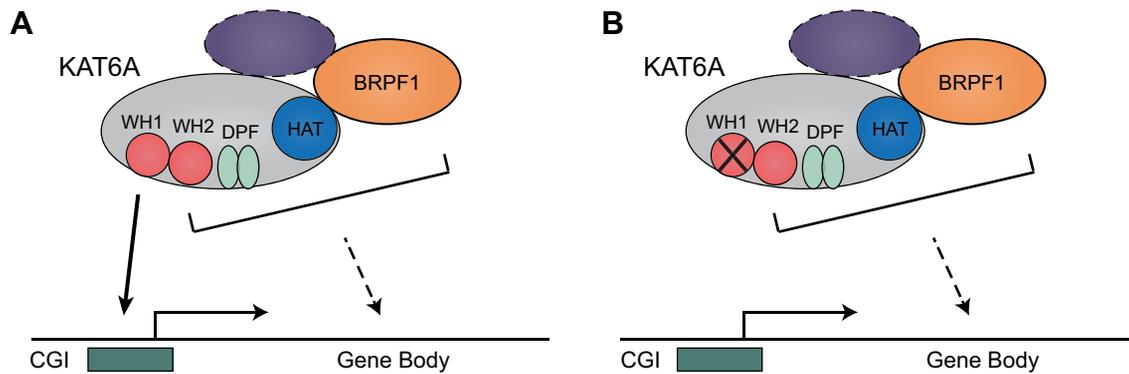


Figure 8. (A) CGI binding of KAT6A strongly depends on the function of the WH1. The recruitment to the gene bodies may be dependent on other domains of KAT6A or on specific interaction partners, such as BRPF1 and others (21). (B) Upon mutating the WH1, KAT6A is no longer recruited to CGIs, but can still bind to gene bodies.

tion of the WH1 domain (Figure 8B), and independently from CGIs.

Interestingly, the domain composition of the N-terminal chromatin binding region is similar between KAT6A, KAT6B and the *Drosophila* homolog Enok (Supplementary Figure S1). Since the *Drosophila* genome does not contain CpG islands, it raises the question of whether Enok's WH1 may have an alternative chromatin binding function. Indeed, closer inspection revealed several differences compared to KAT6A. First, the distance between the two WH domains is much shorter in Enok (5 a.a. versus 25 a.a.) (Supplementary Figure S1). Second, the amino acid composition of the WH1 has substantially diverged during evolution. Although the key DNA binding amino acids are conserved between KAT6A/B and Enok (Figure 1E), other parts of the domain are rather distinct with only about 50% amino acid identity (Supplementary Figure S1G). These sequence differences may lead to another structural conformation, which could influence the DNA binding ability. Thus, it would be interesting to assess whether WH1 of Enok binds to DNA at all, and if yes, whether it has a certain specificity that may contribute to the chromatin association of Enok. The WH2 shows even stronger sequence differences between all three proteins, with only ~30–50% identity between Enok, KAT6A and KAT6B (Supplementary Figure S1G), raising the possibility that the WH2 may have specific functions in each of the three proteins. The histone binding DPF domain (4,62) is similar between KAT6A and KAT6B, but is highly diverged in Enok, with only 20% identity (Supplementary Figure S1G), suggesting that it could also have another function in Enok. Future work will be necessary to clarify how the WH1, WH2 and the DPF domain work together for chromatin binding of the various KAT6 homologs.

More than 300 mutations of KAT6A and KAT6B have been described that lead to intellectual disability (15,17,75–77). Most of these mutations lead to truncated proteins that leave the N-terminal part intact (15,78). It is possible that these truncated proteins may still be recruited to CGIs, which is supported by our observation that the N-terminal part of KAT6A, comprising the region that include WH1, WH2 and the DPF, is sufficient for efficient chromatin binding (Supplementary Figure S4D, E). If the truncated pro-

teins are still recruited, one could speculate that the lack of the C-terminal part of the protein may prevent the interaction with important interaction partners, such as RUNX1 (19). Additionally, the C-terminal region of KAT6A has been implicated in gene activation (19). Thus, a shortened KAT6A or KAT6B protein at CGIs may lead to aberrant gene transcription, which may contribute to the developmental defects observed in patients with truncated KAT6A or KAT6B (17,75,76).

In addition, the KAT6A and KAT6B genes have been recurrently found in acute myeloid leukemia (AML) to be translocated to other histone acetyltransferases, such as CBP and p300 (2,79,80). The resulting proteins contain two HAT domains, and are therefore considered as 'super-HATs' (21). Given that the N-terminal parts of KAT6A or KAT6B, including the WH1 domain, are still present in these fusion proteins, it is likely that these super-HATs are recruited to genomic loci similar to those of the wild-type KAT6A and KAT6B proteins, respectively. It has been proposed that these fusion proteins are hyperactive, leading to aberrant chromatin acetylation patterns and consequently to dysregulated gene expression, which may contribute to AML (21,81). In addition to these translocations, high expression of KAT6A has also been linked to AML progression (41), further supporting that overactivation of KAT6A can lead to cancer. Since the removal of DNA methylation using a DNMT1 inhibitor leads to de novo recruitment of KAT6A (Supplementary Figure S8), it further suggests that aberrant DNA methylation patterns in cancer (82) could potentially lead to erroneous chromatin recruitment of KAT6A, which may also contribute to cancer progression. More work will be required to address whether targeting the WH1 domain may be an alternative strategy to inhibit aberrant functions of KAT6A in diseases (41,73).

In summary, in our work we identified and characterized a DNA binding winged helix domain at the very N-terminus of KAT6A. We demonstrated that this domain is most essential for the recruitment of KAT6A to unmethylated CpG islands in the genome, providing new insights into the targeting mechanisms of KAT6A. This study not only establishes KAT6A as a histone acetyltransferase with a specific DNA binding function but also sheds light on the chromatin regulatory mechanisms at CGIs.

DATA AVAILABILITY

The following websites/tools were used:

AlphaFold Database: <https://alphafold.ebi.ac.uk/>
 Bowtie: <http://bowtie-bio.sourceforge.net/index.shtml>
 Cistrome Data Browser: <http://cistrome.org/db/#/>
 Cistrome Data Analysis Platform: <http://cistrome.org/ap/deeptools>
 deepTools: <https://deeptools.readthedocs.io/>
 Galaxy Europe: <https://usegalaxy.eu/>
 GENCODE: <https://www.gencodegenes.org/>
 GREAT: <http://great.stanford.edu/>
 GSEA: <https://www.gsea-msigdb.org/>
 HOMER: <http://homer.ucsd.edu/homer/motif/>
 UCSC Browser: <https://genome.ucsc.edu/>
 UniProt: <https://www.uniprot.org/>

ChIP-Seq (HEK293) and RNA-Seq (mESCs) have been deposited into the GEO repository under the accession numbers GSE206272 and GSE206273, respectively. Atomic coordinates and structure factors for the reported crystal structure have been deposited with the Protein Data Bank under accession number 7Y43 and 8H7A. Mass spectrometry data were uploaded to ProteomeXchange with the accession number PXD034833.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the staff from BL18U1 and BL19U1 beamlines at Shanghai Synchrotron Radiation Facility (SSRF) in China for assistance during data collection. We acknowledge the Protein Analytics Unit at the Biomedical Center, Ludwig-Maximilians University Munich, for providing services and assistance with data analysis. We thank Andres Blanco, University of Pennsylvania, for providing a KAT6A construct with HAT domain mutations.

FUNDING

R.L. acknowledges support from the German Research Foundation (DFG) [109546710, 416910386]; Fritz Thyssen Foundation [10.20.1.005MN]; Z.W. was supported by funds from the National Natural Science Foundation of China (NSFC) [32071204, 31870725, 32125008]; M.L.B acknowledges funding from the National Institutes of Health (NIH) [R01 HG010501]. Funding for open access charge: Deutsche Forschungsgemeinschaft [109546710].

Conflict of interest statement. None declared.

REFERENCES

- Yang, X.J. (2015) MOZ and MORF acetyltransferases: molecular interaction, animal development and human disease. *Biochim. Biophys. Acta*, **1853**, 1818–1826.
- Borrow, J., Stanton, V.P. Jr, Andresen, J.M., Becher, R., Behm, F.G., Chaganti, R.S., Civin, C.I., Distche, C., Dube, I., Frischauf, A.M. *et al.* (1996) The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB-binding protein. *Nat. Genet.*, **14**, 33–41.
- Champagne, N., Bertos, N.R., Pelletier, N., Wang, A.H., Vezmar, M., Yang, Y., Heng, H.H. and Yang, X.J. (1999) Identification of a human histone acetyltransferase related to monocytic leukemia zinc finger protein. *J. Biol. Chem.*, **274**, 28528–28536.
- Dreveny, L., Deeves, S.E., Fulton, J., Yue, B., Messmer, M., Bhattacharya, A., Collins, H.M. and Heery, D.M. (2014) The double PHD finger domain of MOZ/MYST3 induces alpha-helical structure of the histone H3 tail to facilitate acetylation and methylation sampling and modification. *Nucleic Acids Res.*, **42**, 822–835.
- Rokudai, S., Laptenko, O., Arnal, S.M., Taya, Y., Kitabayashi, I. and Prives, C. (2013) MOZ increases p53 acetylation and premature senescence through its complex formation with PML. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 3895–3900.
- Yu, B., Luo, F., Sun, B., Liu, W., Shi, Q., Cheng, S.Y., Chen, C., Chen, G., Li, Y. and Feng, H. (2021) KAT6A acetylation of SMAD3 regulates myeloid-derived suppressor cell recruitment, metastasis, and immunotherapy in triple-negative breast cancer. *Adv. Sci. (Weinh.)*, **8**, e2100014.
- Katsumoto, T., Aikawa, Y., Iwama, A., Ueda, S., Ichikawa, H., Ochiya, T. and Kitabayashi, I. (2006) MOZ is essential for maintenance of hematopoietic stem cells. *Genes Dev.*, **20**, 1321–1330.
- Thomas, T., Corcoran, L.M., Gugasyan, R., Dixon, M.P., Brodnicki, T., Nutt, S.L., Metcalf, D. and Voss, A.K. (2006) Monocytic leukemia zinc finger protein is essential for the development of long-term reconstituting hematopoietic stem cells. *Genes Dev.*, **20**, 1175–1186.
- Thomas, T., Voss, A.K., Chowdhury, K. and Gruss, P. (2000) Querkopf, a MYST family histone acetyltransferase, is required for normal cerebral cortex development. *Development*, **127**, 2537–2548.
- Perez-Campo, F.M., Costa, G., Lie-a-Ling, M., Kouskoff, V. and Lacaud, G. (2013) The MYSTERIOUS MOZ, a histone acetyltransferase with a key role in haematopoiesis. *Immunology*, **139**, 161–165.
- Miyamoto, R., Okuda, H., Kanai, A., Takahashi, S., Kawamura, T., Matsui, H., Kitamura, T., Kitabayashi, I., Inaba, T. and Yokoyama, A. (2020) Activation of cpg-rich promoters mediated by MLL drives MOZ-rearranged leukemia. *Cell Rep.*, **32**, 108200.
- Yan, K., Rousseau, J., Machol, K., Cross, L.A., Agre, K.E., Gibson, C.F., Goverde, A., Engleman, K.L., Verdin, H., De Baere, E. *et al.* (2020) Deficient histone H3 propionylation by BRPF1-KAT6 complexes in neurodevelopmental disorders and cancer. *Sci. Adv.*, **6**, eaax0021.
- Crump, J.G., Swartz, M.E., Eberhart, J.K. and Kimmel, C.B. (2006) Moz-dependent hox expression controls segment-specific fate maps of skeletal precursors in the face. *Development*, **133**, 2661–2669.
- Miller, C.T., Maves, L. and Kimmel, C.B. (2004) moz regulates Hox expression and pharyngeal segmental identity in zebrafish. *Development*, **131**, 2443–2461.
- Kennedy, J., Goudie, D., Blair, E., Chandler, K., Joss, S., McKay, V., Green, A., Armstrong, R., Lees, M., Kamien, B. *et al.* (2019) KAT6A syndrome: genotype-phenotype correlation in 76 patients with pathogenic KAT6A variants. *Genet. Med.*, **21**, 850–860.
- Tham, E., Lindstrand, A., Santani, A., Malmgren, H., Nesbitt, A., Dubbs, H.A., Zackai, E.H., Parker, M.J., Millan, F., Rosenbaum, K. *et al.* (2015) Dominant mutations in KAT6A cause intellectual disability with recognizable syndromic features. *Am. J. Hum. Genet.*, **96**, 507–513.
- Arboleda, V.A., Lee, H., Dorrani, N., Zadeh, N., Willis, M., Macmurdo, C.F., Manning, M.A., Kwan, A., Hudgins, L., Barthelemy, F. *et al.* (2015) De novo nonsense mutations in KAT6A, a lysine acetyl-transferase gene, cause a syndrome including microcephaly and global developmental delay. *Am. J. Hum. Genet.*, **96**, 498–506.
- Champagne, N., Pelletier, N. and Yang, X.J. (2001) The monocytic leukemia zinc finger protein MOZ is a histone acetyltransferase. *Oncogene*, **20**, 404–409.
- Kitabayashi, I., Aikawa, Y., Nguyen, L.A., Yokoyama, A. and Ohki, M. (2001) Activation of AML1-mediated transcription by MOZ and inhibition by the MOZ-CBP fusion protein. *EMBO J.*, **20**, 7184–7196.
- Pelletier, N., Champagne, N., Stifani, S. and Yang, X.J. (2002) MOZ and MORF histone acetyltransferases interact with the Runt-domain transcription factor runx2. *Oncogene*, **21**, 2729–2740.
- Klein, B.J., Lalonde, M.E., Cote, J., Yang, X.J. and Kutateladze, T.G. (2014) Crosstalk between epigenetic readers regulates the MOZ/MORF HAT complexes. *Epigenetics*, **9**, 186–193.

22. Surapureddi, S., Yu, S., Bu, H., Hashimoto, T., Yeldandi, A. V., Kashireddy, P., Cherkaoui-Malki, M., Qi, C., Zhu, Y. J., Rao, M. S. *et al.* (2002) Identification of a transcriptionally active peroxisome proliferator-activated receptor alpha-interacting cofactor complex in rat liver and characterization of PRIC285 as a coactivator. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 11836–11841.
23. Yang, X. J. and Ullah, M. (2007) MOZ and MORF, two large MYSTic HATs in normal and cancer stem cells. *Oncogene*, **26**, 5408–5419.
24. Sheikh, B. N., Downer, N. L., Phipson, B., Vanyai, H. K., Kueh, A. J., McCarthy, D. J., Smyth, G. K., Thomas, T. and Voss, A. K. (2015) MOZ and BMI1 play opposing roles during hox gene activation in ES cells and in body segment identity specification in vivo. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 5437–5442.
25. Laue, K., Daujat, S., Crump, J. G., Plaster, N., Roehl, H. H., Kimmel, C. B., Schneider, R. and Hammerschmidt, M. (2008) The multidomain protein brpf1 binds histones and is required for hox gene expression and segmental identity. *Development*, **135**, 1935–1946.
26. Voss, A. K., Collin, C., Dixon, M. P. and Thomas, T. (2009) Moz and retinoic acid coordinately regulate H3K9 acetylation, hox gene expression, and segment identity. *Dev. Cell*, **17**, 674–686.
27. Vanyai, H. K., Garnham, A., May, R. E., McRae, H. M., Collin, C., Wilcox, S., Smyth, G. K., Thomas, T. and Voss, A. K. (2019) MOZ directs the distal-less homeobox gene expression program during craniofacial development. *Development*, **146**, dev175042.
28. Aranda, S., Mas, G. and Di Croce, L. (2015) Regulation of gene transcription by polycomb proteins. *Sci. Adv.*, **1**, e1500737.
29. Strubbe, G., Popp, C., Schmidt, A., Pauli, A., Ringrose, L., Beisel, C. and Paro, R. (2011) Polycomb purification by in vivo biotinylation tagging reveals cohesin and trithorax group proteins as interaction partners. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 5572–5577.
30. Kang, H., Jung, Y. L., McElroy, K. A., Zee, B. M., Wallace, H. A., Woolnough, J. L., Park, P. J. and Kuroda, M. I. (2017) Bivalent complexes of PRC1 with orthologs of BRD4 and MOZ/MORF target developmental genes in drosophila. *Genes Dev.*, **31**, 1988–2002.
31. Deaton, A. M. and Bird, A. (2011) CpG islands and the regulation of transcription. *Genes Dev.*, **25**, 1010–1022.
32. Thomson, J. P., Skene, P. J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A. R., Deaton, A., Andrews, R., James, K. D. *et al.* (2010) CpG islands influence chromatin structure via the cpg-binding protein cfp1. *Nature*, **464**, 1082–1086.
33. Farcas, A. M., Blackledge, N. P., Sudbery, I., Long, H. K., McGouran, J. F., Rose, N. R., Lee, S., Sims, D., Cerase, A., Sheahan, T. W. *et al.* (2012) KDM2B links the polycomb repressive complex 1 (PRC1) to recognition of CpG islands. *Elife*, **1**, e00205.
34. Li, H., Liefke, R., Jiang, J., Kurland, J. V., Tian, W., Deng, P., Zhang, W., He, Q., Patel, D. J., Bulyk, M. L. *et al.* (2017) Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature*, **549**, 287–291.
35. Fischer, S., Weber, L. M. and Liefke, R. (2022) Evolutionary adaptation of the polycomb repressive complex 2. *Epigenetics Chromatin*, **15**, 7.
36. Stielow, B., Zhou, Y., Cao, Y., Simon, C., Pogoda, H. M., Jiang, J., Ren, Y., Phanor, S. K., Rohner, I., Nist, A. *et al.* (2021) The SAM domain-containing protein 1 (SAMD1) acts as a repressive chromatin regulator at unmethylated CpG islands. *Sci. Adv.*, **7**, eabf2229.
37. Stielow, B., Simon, C. and Liefke, R. (2021) Making fundamental scientific discoveries by combining information from literature, databases, and computational tools - An example. *Comput. Struct. Biotechnol. J.*, **19**, 3027–3033.
38. Simon, C., Stielow, B., Nist, A., Rohner, I., Weber, L. M., Geller, M., Fischer, S., Stiewe, T. and Liefke, R. (2022) The CpG island-binding protein SAMD1 contributes to an unfavorable gene signature in hepg2 hepatocellular carcinoma cells. *Biology (Basel)*, **11**, 557.
39. Berger, M. F., Philippakis, A. A., Qureshi, A. M., He, F. S., Estep, P. W. 3rd and Bulyk, M. L. (2006) Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat. Biotechnol.*, **24**, 1429–1435.
40. Sanjana, N. E., Shalem, O. and Zhang, F. (2014) Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods*, **11**, 783–784.
41. Yan, F., Li, J., Milosevic, J., Petroni, R., Liu, S., Shi, Z., Yuan, S., Reynaga, J. M., Qi, Y., Rico, J. *et al.* (2022) KAT6A and ENL form an epigenetic transcriptional control module to drive critical leukemogenic gene-expression programs. *Cancer Discov.*, **12**, 792–811.
42. Volkel, S., Stielow, B., Finkernagel, F., Stiewe, T., Nist, A. and Suske, G. (2015) Zinc finger independent genome-wide binding of Sp2 potentiates recruitment of histone-fold protein Nf-y distinguishing it from Sp1 and sp3. *PLoS Genet.*, **11**, e1005102.
43. Azevedo Portilho, N., Saini, D., Hossain, I., Sirois, J., Moraes, C. and Pastor, W. A. (2021) The DNMT1 inhibitor GSK-3484862 mediates global demethylation in murine embryonic stem cells. *Epigenetics Chromatin*, **14**, 56.
44. Nelson, J. D., Denisenko, O. and Bomsztyk, K. (2006) Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nat. Protoc.*, **1**, 179–185.
45. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S. L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
46. Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. and Manke, T. (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.*, **42**, W187–W191.
47. Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W. *et al.* (2008) Model-based analysis of chip-Seq (MACS). *Genome Biol.*, **9**, R137.
48. Yu, G., Wang, L. G. and He, Q. Y. (2015) ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, **31**, 2382–2383.
49. McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., Wenger, A. M. and Bejerano, G. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.*, **28**, 495–501.
50. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. and Glass, C. K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and b cell identities. *Mol. Cell*, **38**, 576–589.
51. Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
52. Liu, T., Ortiz, J. A., Taing, L., Meyer, C. A., Lee, B., Zhang, Y., Shin, H., Wong, S. S., Ma, J., Lei, Y. *et al.* (2011) Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol.*, **12**, R83.
53. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T. R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
54. Love, M. I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
55. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, A. S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 15545–15550.
56. Piunti, A., Smith, E. R., Morgan, M. A. J., Ugarenko, M., Khaltyan, N., Helmin, K. A., Ryan, C. A., Murray, D. C., Rickels, R. A., Yilmaz, B. D. *et al.* (2019) CATAComb: an endogenous inducible gene that antagonizes H3K27 methylation activity of polycomb repressive complex 2 via an H3K27M-like mechanism. *Sci. Adv.*, **5**, eaax2887.
57. Liu, W., Ma, Q., Wong, K., Li, W., Ohgi, K., Zhang, J., Aggarwal, A. and Rosenfeld, M. G. (2013) Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release. *Cell*, **155**, 1581–1595.
58. Morgan, M. A. J., Rickels, R. A., Collings, C. K., He, X., Cao, K., Herz, H. M., Cozzolino, K. A., Abshiru, N. A., Marshall, S. A., Rendleman, E. J. *et al.* (2017) A cryptic tudor domain links BRWD2/PHIP to COMPASS-mediated histone H3K4 methylation. *Genes Dev.*, **31**, 2003–2014.
59. Liang, K., Smith, E. R., Aoi, Y., Stoltz, K. L., Katagi, H., Woodfin, A. R., Rendleman, E. J., Marshall, S. A., Murray, D. C., Wang, L. *et al.* (2018) Targeting processive transcription elongation via SEC disruption for MYC-Induced cancer therapy. *Cell*, **175**, 766–779.
60. Stielow, B., Finkernagel, F., Stiewe, T., Nist, A. and Suske, G. (2018) MGA, L3MBTL2 and E2F6 determine genomic binding of the non-canonical polycomb repressive complex PRC1.6. *PLoS Genet.*, **14**, e1007193.
61. Qiu, Y., Liu, L., Zhao, C., Han, C., Li, F., Zhang, J., Wang, Y., Li, G., Mei, Y., Wu, M. *et al.* (2012) Combinatorial readout of unmodified H3R2 and acetylated H3K14 by the tandem PHD finger of MOZ

- reveals a regulatory mechanism for HOXA9 transcription. *Genes Dev.*, **26**, 1376–1391.
62. Xiong, X., Panchenko, T., Yang, S., Zhao, S., Yan, P., Zhang, W., Xie, W., Li, Y., Zhao, Y., Allis, C.D. *et al.* (2016) Selective recognition of histone crotonylation by double PHD fingers of MOZ and DPF2. *Nat. Chem. Biol.*, **12**, 1111–1118.
 63. Ullah, M., Pelletier, N., Xiao, L., Zhao, S.P., Wang, K., Degerny, C., Tahmasebi, S., Cayrou, C., Doyon, Y., Goh, S.L. *et al.* (2008) Molecular architecture of quartet MOZ/MORF histone acetyltransferase complexes. *Mol. Cell. Biol.*, **28**, 6828–6843.
 64. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A. *et al.* (2021) Highly accurate protein structure prediction with alphafold. *Nature*, **596**, 583–589.
 65. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A. *et al.* (2022) AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.*, **50**, D439–D444.
 66. Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M. and Iyer, L.M. (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol. Rev.*, **29**, 231–262.
 67. Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M. and Helin, K. (2007) The polycomb group protein *suz12* is required for embryonic stem cell differentiation. *Mol. Cell. Biol.*, **27**, 3769–3779.
 68. Pachano, T., Sanchez-Gaya, V., Ealo, T., Mariner-Fauli, M., Bleckwehl, T., Asenjo, H.G., Respuela, P., Cruz-Molina, S., Munoz-San Martin, M., Haro, E. *et al.* (2021) Orphan CpG islands amplify poised enhancer regulatory activity and determine target gene responsiveness. *Nat. Genet.*, **53**, 1036–1049.
 69. Holbert, M.A., Sikorski, T., Carten, J., Snowflack, D., Hodawadekar, S. and Marmorstein, R. (2007) The human monocytic leukemia zinc finger histone acetyltransferase domain contains DNA-binding activity implicated in chromatin targeting. *J. Biol. Chem.*, **282**, 36603–36613.
 70. Klein, B.J., Cox, K.L., Jang, S.M., Cote, J., Poirier, M.G. and Kutateladze, T.G. (2020) Molecular basis for the PZP domain of BRPF1 association with chromatin. *Structure*, **28**, 105–110.
 71. Yang, X.J. and Seto, E. (2007) HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. *Oncogene*, **26**, 5310–5318.
 72. Legube, G. and Trouche, D. (2003) Regulating histone acetyltransferases and deacetylases. *EMBO Rep.*, **4**, 944–947.
 73. Baell, J.B., Leaver, D.J., Hermans, S.J., Kelly, G.L., Brennan, M.S., Downer, N.L., Nguyen, N., Wichmann, J., McRae, H.M., Yang, Y. *et al.* (2018) Inhibitors of histone acetyltransferases KAT6A/B induce senescence and arrest tumour growth. *Nature*, **560**, 253–257.
 74. Xu, C., Bian, C., Lam, R., Dong, A. and Min, J. (2011) The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain. *Nat. Commun.*, **2**, 227.
 75. Kraft, M., Cirstea, I.C., Voss, A.K., Thomas, T., Goehring, I., Sheikh, B.N., Gordon, L., Scott, H., Smyth, G.K., Ahmadian, M.R. *et al.* (2011) Disruption of the histone acetyltransferase MYST4 leads to a Noonan syndrome-like phenotype and hyperactivated MAPK signaling in humans and mice. *J. Clin. Invest.*, **121**, 3479–3491.
 76. Trinh, J., Huning, I., Yuksel, Z., Baalman, N., Imhoff, S., Klein, C., Rolfs, A., Gillesen-Kaesbach, G. and Lohmann, K. (2018) A KAT6A variant in a family with autosomal dominantly inherited microcephaly and developmental delay. *J. Hum. Genet.*, **63**, 997–1001.
 77. Simpson, M.A., Deshpande, C., Dafou, D., Vissers, L.E., Woollard, W.J., Holder, S.E., Gillesen-Kaesbach, G., Derks, R., White, S.M., Cohen-Snuijff, R. *et al.* (2012) De novo mutations of the gene encoding the histone acetyltransferase KAT6B cause genitopatellar syndrome. *Am. J. Hum. Genet.*, **90**, 290–294.
 78. Zhang, L.X., Lemire, G., Gonzaga-Jauregui, C., Molidpere, S., Galaz-Montoya, C., Liu, D.S., Verloes, A., Shillington, A.G., Izumi, K., Ritter, A.L. *et al.* (2020) Further delineation of the clinical spectrum of KAT6B disorders and allelic series of pathogenic variants. *Genet. Med.*, **22**, 1338–1347.
 79. Chaffanet, M., Gressin, L., Preudhomme, C., Soenen-Cornu, V., Birnbaum, D. and Pebusque, M.J. (2000) MOZ is fused to p300 in an acute monocytic leukemia with t(8;22). *Genes Chromosomes Cancer*, **28**, 138–144.
 80. Panagopoulos, I., Fioretos, T., Isaksson, M., Samuelsson, U., Billstrom, R., Strombeck, B., Mitelman, F. and Johansson, B. (2001) Fusion of the MORF and CBP genes in acute myeloid leukemia with the t(10;16)(q22;p13). *Hum. Mol. Genet.*, **10**, 395–404.
 81. Zhou, C., Liu, W. and Duan, Y. (2020) MOZ/KAT6A: a promising target for acute myeloid leukemia therapy. *Future Med Chem*, **12**, 759–761.
 82. Ehrlich, M. (2002) DNA methylation in cancer: too much, but also too little. *Oncogene*, **21**, 5400–5413.