

Aus dem Institut für Molekularbiologie und Tumorforschung  
Geschäftsführender Direktor: Prof. Dr. Alexander Brehm  
des Fachbereichs Medizin der Philipps-Universität Marburg

*Characterization of histone acetyltransferase KAT6A and  
transcriptional repressor SAMD1 – two proteins possessing a  
novel class of unmethylated DNA-binding winged helix domains*

Inaugural-Dissertation zur Erlangung  
des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)  
dem Fachbereich Medizin der Philipps-Universität Marburg  
vorgelegt von

**Lisa Marie Weber**

aus Siegen

Marburg, 2023

Angenommen vom Fachbereich Medizin der Philipps-Universität Marburg  
am: 21.09.2023

Gedruckt mit Genehmigung des Fachbereichs Medizin

Dekanin: Prof. Dr. Denise Hilfiker-Kleiner

Referent: PD Dr. Robert Liefke

Korreferentin: Prof. Dr. Sandra Hake

## TABLE OF CONTENTS

Table of contents .....	1
List of abbreviations.....	2
1. Introduction .....	6
1.1 CpG islands and their role in chromatin remodelling.....	6
1.2 Chromatin binding proteins and their DNA recruitment mechanisms.....	7
1.3 DNA-binding of winged helix domain proteins .....	8
1.4 SAMD1-like family of winged helix domains.....	10
2. The histone acetyltransferase KAT6A is recruited to unmethylated CpG islands via a DNA binding winged helix domain .....	11
2.1 Introduction .....	11
2.1.1 Lysine acylation activity of the KAT6A complex .....	11
2.1.2 KAT6A implications in development and disease.....	13
2.2 Synopsis of results .....	16
2.3 Discussion.....	24
2.3.1 Substantial effects of a DNA-binding WH domain on genome-wide aspects .....	24
2.3.2 A newly identified connection of KAT6A to PRC1.....	26
2.3.3 Perspectives of KAT6A WH1 involvement in cancer and developmental disorders .....	29
2.4 Contribution statement .....	31
3. Investigation of SAMD1 ablation in mice .....	32
3.1 Introduction .....	32
3.2 Synopsis of results .....	34
3.3 Discussion.....	40
3.4 Contribution statement .....	43
4. Summary.....	44
5. Zusammenfassung.....	45
6. References .....	47
Appendix .....	55

## LIST OF ABBREVIATIONS

%	Percent
Δ	Delta/deletion
5mC	5- methylcytosine
α	Alpha-
β	Beta-
Å	Ångström
ac	Acetyl
AML	Acute myeloid leukaemia
BRPF1	Bromodomain-PHD finger protein 1
C-	Carboxy-
CAP	Catabolite gene activator protein
CBP	CREB-binding protein
CD31	Cluster of differentiation 31
CFP1	CXXC-type zinc finger protein 1
CFU	Colony-forming unit
CGI	Cpg island
ChIP	Chromatin immunoprecipitation
CIRA	Cpg island recovery assay
CNC	Cranial neural crest cells
CpG	Cytosine-phosphate-guanine
CREB	Cyclic AMP-responsive element-binding protein
d	Drosophila-
Dlx	Distal-less homeobox
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DP	Dimerization partner
DPF	Double PHD finger
ds	Double-stranded
ds	Double stranded
E	Embryonic day
E2F	Transcription factor E2F
E2F4-DP2	E2F4 and dimerization partner 2
EBD	ENL binding domain
EBs	Embryoid bodies
ECs	Endothelial cells
EMSA	Electrophoretic mobility shift assay
EPc	Enhancer of Polycomb
ER	Estrogen receptor
et al.	Et alii (and others)
fanChIP	Fractionation-assisted native chip
FGF-2	Fibroblast growth factor 2
FKH	Forkhead
fl	Full-length

FLAG	DYKDDDDK octapeptide
Foxa2	Hepatocyte nuclear factor 3-beta
GBM	Glioblastoma
Gfap	Glial fibrillary acidic protein
GSEA	Gene set enrichment analysis
H	Histone
h	Human-
H&E	Haematoxylin and eosin
HA	Hemagglutinin
HAT	Histone acetyltransferase
HBO1	HAT bound to ORC1
HCC	Hepatocellular carcinoma
HDL	High-density lipoprotein
HEK293	Human embryonic kidney 293
HET	Heterozygous
HFD	High-fat diet
HNF3 $\gamma$	Hepatocyte nuclear factor 3-gamma
Hox	Homeobox
HTH	Helix-turn-helix
ICM	Inner cell mass
ID	Intellectual disability
ING4/5	Inhibitor of growth 4/5
IP	Immunoprecipitation
ITC	Isothermal titration calorimetry
Kac	Lysine acetylation
KAT6A/B	Lysine acetyltransferase 6A/B
kb	Kilobase
Kbu	Lysine butyrylation
Kcr	Lysine crotonylation
kDa	Kilodalton
KDM1A	Lysine-specific demethylase 1A
KDM2A/B	Lysine-specific demethylase 2A/B
KO	Knockout
Kpr	Lysine propionylation
L3MBTL3	Lethal (3) malignant brain tumour (L(3)mbt)-like 3
LDL	Low-density lipoprotein
LIF	Leukaemia inhibitory factor
m	Mouse-
MBD	Methyl-binding domain
MBT	Malignant brain tumour
mCpG	Methylated cpG-DNA
me	Methyl
MEAF6	MYST/Esa1-associated factor 6
MEFs	Mouse embryonic fibroblasts
mESCs	Mouse embryonic stem cells

miR	MicroRNA
MLL2	Mixed-lineage leukaemia protein 2
MORF	MOZ related factor
MOZ	Monocytic leukaemia zinc finger protein
mRNA	Messenger RNA
MS	Mass spectrometry
mut	Mutation
Myh6	Myosin-6
MYST	MOZ-Ybf2-Sas2-TIP60 domain
N-	Amino-
NEMM	N-terminal part of Enok, MOZ or MORF
Nes	Nestin
nM	Nanomolar
NPCs	Neural progenitor cells
OGTT	Oral glucose tolerance test
Olig1	Oligodendrocyte transcription factor 1
P	Postnatal day
p53	Cellular tumor antigen p53
PBM	Protein binding microarray
PcG	Polycomb group
PCL	Polycomb-like
PE	Poised enhancer
PHD	Plant homeodomain-linked
PRC	Polycomb repressive complex
qPCR	Quantitative polymerase chain reaction
RA	Retinoic acid
RBCs	Red blood cells
RFX1	Regulatory factor X1
RNA	Ribonucleic acid
RNAP2	RNA polymerase II
RT	Real-time
Runx1	Runt-domain transcription factor 1
SAM	Sterile alpha motif
SAMD1	SAM domain-containing protein 1
Seq	Sequencing
SINE	Short interspersed elements
SM domain	Serine and methionine-rich domain
SNP	Single nucleotide polymorphism
SRA	SET and RING-associated
TF	Transcription factor
trxG	Trithorax group
TSS	Transcription start sites
Tubb3	Tubulin beta-3 chain
UTR	Untranslated region
VEGFR2	Vascular endothelial growth factor receptor 2

VSMCs	Vascular smooth muscle cells
WES	Whole exome sequencing
WGBS	Whole genome bisulfite sequencing
WH	Winged helix
WT	Wild-type
ZMYND11	Zinc finger MYND domain-containing protein

Abbreviations of amino acids and nucleotides follow the standard single-letter code.

## 1. INTRODUCTION

### 1.1 CpG islands and their role in chromatin remodelling

Chromatin is composed of condensed heterochromatic and open euchromatic areas that are associated with silent and active transcription, respectively. Transcriptional regulation of genes is mainly governed by factors of the chromatin environment, including DNA methylation and histone modifications.

DNA methylation is directly linked to gene expression and epigenetically defines cell types and lineages through genomic stability and transcription control (Bhootra et al., 2023). It is defined as the addition of a methyl group to a cytosine or adenine base. DNA methylation in mammals occurs almost exclusively at cytosines located in CpG dinucleotides at the carbon-5 position on the cytosine ring (methylcytosine/5mC). Approximately 60-90% of CpG sites are methylated, typically associated with reduced gene expression (Bird, 1986; Hu et al., 2013; Varriale & Bernardi, 2010). Methylation patterns across the genome are established by maintenance and *de novo* methyltransferases as well as demethylases. Maintenance DNA methyltransferase 1 (DNMT1) preserves tissue-specific methylation by recognizing and modifying hemimethylated sites that arise during DNA replication and copying the parental methylation pattern to the daughter strand (Bhootra et al., 2023; Caiafa & Zampieri, 2005). These patterns influence the recruitment and binding of transcription factors (TFs), methyl-binding proteins and linker histones to DNA by altered steric effects and therefore have an impact on chromatin structure (Bhootra et al., 2023; Davey et al., 1997). Minor alterations in the methylation patterns of a single gene, individual chromosome or whole genome can be an important part of normal development, while aberrant gene expression due to *de novo* methylation or disruption of DNA methylation can be the cause of several diseases, including cancer (Bhootra et al., 2023).

Chromatin remodelling is the regulatory mechanism of dynamic changes between transcriptionally active or repressed sites, defined by distinct selections of local histone modifications. A mark generally associated with transcriptional activation is the methylation of histone H3 at lysine 4 (H3K4me). H3K4me3 is facilitated by the histone methyltransferase complex MLL2 (mixed-lineage leukaemia protein 2) and recognized by the PHD (plant homeodomain-linked) finger domain of ING4, a subunit of the, H3K4me3 inhibits the activity of DNMTs, all together supporting an active chromatin configuration (Bell & Vertino, 2017; Deaton & Bird, 2011; Hung et al., 2009; Ku et al., 2008). Transcriptional repression is mainly linked to the methylation of histone H3 at lysine 27 (H3K27me), a mark deposited by the catalytic subunit EZH2 of the Polycomb repressive complex 2 (PRC2). The underlying recruitment mechanism of PRC2 is reciprocal interplay with PRC1 and its deposition of monoubiquitylation of histone H2A at lysine 119 (H2AK119ub) (Deaton & Bird, 2011; Owen & Davidovich, 2022; Stielow et al., 2018). Specific cells, such as mouse embryonic stem cells (mESCs), were shown to possess promoters in a bivalent chromatin state, carrying both the active H3K4me3 mark and the repressive H3K27me3 mark. These genes are in a poised state between active transcription and stable transcriptional repression (Bryan et al., 2021; Deaton & Bird, 2011).



A particular feature of the vertebrate genome is CpG islands (CGIs). They are hypomethylated, CpG-rich regulatory elements associated with a euchromatic and transcriptionally active state (Bell & Vertino, 2017; Bird, 1986). CpG-rich sequences are comparably rare motifs in vertebrate DNA and occur at one-fifth of the expected frequency. The human genome contains approximately 30,000 CGIs, stretching over the length of 0.5 - 2 kb of DNA (Jeziorska et al., 2017). Due to their accessible chromatin conformation, CGIs are largely connected to promoter sites. However, only 50 - 65% of all annotated CGIs in mouse and human genomes are associated with classical promoter regions, whereas the remaining so-called orphan CGIs often lie in gene body regions. Orphan CGIs are suggested to act as active enhancer sites throughout different tissues, to represent novel promoter sites, or to become methylated during developmental processes and cell differentiation (Deaton & Bird, 2011; Jeziorska et al., 2017; Pachano et al., 2021).

Although their enrichment in CpG nucleotides provides an attractive substrate for DNA methyltransferases, CGIs remain largely unmethylated. Coinciding with their methylation status, the transcription of genes linked to CGIs is elevated as long as they remain unmethylated but is inhibited upon methylation (Caiafa & Zampieri, 2005). Indeed, promoters without or with sparse CGI frequency were found to be more susceptible to methylation than promoters high in CGI frequency (Shen et al., 2007). The presence of H3K4me3 repels DNMT binding and thus prevents the hypermethylation of CGIs within promoters and enhancers and the concomitant recruitment of histone deacetylases and other chromatin remodelers that lead to the repression of linked genes. In addition to H3K4me3 and DNA hypomethylation, promoter and enhancer CGIs share other common features, such as decoration with acetyl-H3 lysine 9/27 (H3K9ac/H3K27ac) marks and general or cell type-specific transcription factor binding (Bell & Vertino, 2017). Likewise, Polycomb group (PcG) proteins belonging to the two types of Polycomb repressive complexes are targeted to CpG islands in a cell type-specific manner (Owen & Davidovich, 2022).

Other than CGIs, there are additional DNA regions rich in CpG, such as repetitive B elements in mice and Alu elements, a family of SINEs, in humans. Compared to CGIs, these elements exhibit higher methylation levels, are slightly smaller (< 0.5 kb) and show a lower CpG content (< 55%) (Caiafa & Zampieri, 2005). Additionally, B and Alu elements have been proposed to act as methylation centres, able to initiate spreading of *de novo* methylation to neighbouring regions in the absence of a boundary (Caiafa & Zampieri, 2005; Deaton & Bird, 2011). Interestingly, neighbouring regions of methylated promoter CGIs are often Alu sites, and nearly one-third of the human genome mCpGs are located within Alu repeats (Deaton & Bird, 2011).

## 1.2 Chromatin binding proteins and their DNA recruitment mechanisms

In addition to covalent chromatin modifications, chromatin binding proteins are a major contributor to the preservation of either repressed or open chromatin configurations during development. One of the ubiquitous chromatin binding proteins is linker histone H1. In mammals, eleven linker histones are known, grouped into five ubiquitously expressed variants, namely,

H1.1-5. According to their name, they bind to the linker DNA between nucleosomes and promote DNA packaging (Behrends & Engmann, 2020). Linker histones are larger than core histones and less tightly bound to DNA. Their binding affects the nucleosome repeat length as well as the stability of higher-order chromatin structures. H1-like linker histones bind to the linker DNA amid nucleosomes in an 'on-dyad' position, mediated by the electrostatic attraction between the negatively charged phosphate backbone of DNA and specific basic residues of the N-terminus, the  $\alpha 3$  helices, and L1 loops of the globular domains within H1. Meanwhile, the C-terminus of linker histones take part in determining the proximity of linker DNA between nucleosomes (Hergeth & Schneider, 2015; Saha & Dalal, 2021).

Several proteins have the capacity to specifically bind to methylated CpG and are involved in establishing or maintaining the repressive heterochromatin state of DNA. This is achieved by their recruitment to mCpG and their concomitant association with effector enzymes, which in turn catalyse chromatin modifications. To date, nine mCpG-binding proteins have been identified in mammals that can be divided into three main families: the methyl-binding domain (MBD) family, SRA domain proteins and Kaiso and Kaiso-like proteins (Defossez & Stancheva, 2011). While some of the mCpG binding zinc-finger proteins of the Kaiso family are associated with histone deacetylase activity and the establishment of heterochromatin, members of the MBD family are known to bind to hypermethylated DNA and maintain the silent chromatin state (Bogdanović & Veenstra, 2009).

On the other hand, proteins have been identified that specifically bind to unmethylated CpG. Their presence is generally associated with gene activation and the opening of adjoining condensed chromatin. For example, CFP1 (CXXC-type zinc finger protein 1, also called CXXC1 or CGBP), part of the SETD1 H3K4 methyltransferase complex, as well as the H3K36 demethylases KDM2A and KDM2B (lysine-specific demethylase 2A/B, or CXXC8/CXXC2, respectively)(Marx et al., 2013; Shin Voo et al., 2000; Thomson et al., 2010). The combined feature of these and more proteins is a cysteine-rich zinc-binding CXXC domain, mediating their recruitment to unmethylated promoter CGIs. In mammals, twelve CXXC zinc finger proteins have been identified thus far (X. Xiong et al., 2019). Furthermore, the ubiquitously expressed transcription factors Sp1 and Sp3 can bind to the GC box motif (5'-GGGGCGGGG-3') within CGIs and (co-)occupy promoter regions, thereby protecting the targeted genes from cytosine methylation and maintaining the open chromatin state (Lee et al., 2001; Völkel et al., 2015). Of note, some transcription factors, such as KLF4, were found to bind to both methylated and unmethylated DNA motifs via two functionally distinct domains, suggesting specific and independent roles in different biological contexts (Hu et al., 2013).

### 1.3 DNA-binding of winged helix domain proteins

A recurring structure involved in protein-DNA interactions is the winged helix (WH) domain. WH domains interact with double-stranded (ds) DNA and are present in one-third of all DNA-binding proteins (Harami et al., 2013). The WH domain was initially described as a helix-turn-helix

(HTH) domain, characterized by three core alpha helices forming a right-handed bundle in a partly open conformation with a sharp turn between  $\alpha 2$  and  $\alpha 3$  helices. A loop between the  $\alpha 1$  and  $\alpha 2$  helices allows for higher conformational variance and modifications in further classes of HTH domains, as well as additional N- or C-terminal extensions (Aravind et al., 2005). The protruding  $\alpha 3$  helix is defined as the canonical recognition helix, typically involved in direct and base-specific contact with the major groove of DNA (Brennan, 1993). In contrast to HTH domains, WH domains are characterized by an additional C-terminal beta sheet that may take different conformations, such as a two-strand or a three-strand sheet, and two characteristic loops (also called 'wings'), which often provide a second interface for substrate contact with the minor groove of DNA (Aravind et al., 2005; Harami et al., 2013).

Winged helix domain-containing proteins associate with a variety of different complexes. They bind to DNA in different conformations, such as monomers, homodimers or heterodimers, rendering the WH-domain an extremely versatile motif (Gajiwala & Burley, 2000). They can be categorized according to their binding mechanism into a canonical and a non-canonical group, the main difference being the orientation and involvement of the  $\alpha 3$  helix in DNA recognition at the major groove.

The first identified WH domain recognizing dsDNA is part of the hepatocyte nuclear factor 3-gamma (HNF3 $\gamma$  or FOXA3) protein, and its structure defines canonical winged helix domains (HNF3 $\gamma$ -like). Canonical WH domains bind to the major groove of DNA via their  $\alpha 3$  helix. This domain is found in transcription factors such as the FOX family, E2F4 and dimerization partner 2 (E2F4-DP2), and catabolite gene activator protein (CAP) (Harami et al., 2013). Members of the forkhead (FKH) winged helix family of transcription factors, such as FOXO1, FOXA2 and FOXP3, facilitate direct interaction with DNA via their FKH domain. These canonical winged helix domains were found to preferentially bind to DNA at the FKH consensus sequence (5'-GTAAACA-3') and act either alone or in cooperation with other DNA-binding domains (J. Li et al., 2017). Moreover, E2F is a family of transcription factors that often heterodimerizes with members of the dimerization partner (DP) protein family and regulates genes involved in DNA replication and growth. Although missing loop W2, both E2F4 and DP2 are canonical WH proteins due to the DNA binding of each  $\alpha 3$  helix to half of a palindromic target sequence (5'-CGCGCG-3') (Gajiwala & Burley, 2000).

In contrast to canonical WH domains, non-canonical WH domains have multiple modes of interacting with dsDNA, for example, by major groove interaction of the W1 loop, as is the case for the transcriptional activator regulatory factor X1 (RFX1) (Harami et al., 2013). Here, the  $\alpha 3$  helix is still involved in DNA interaction but binds with only one lysine side chain to bases in the minor groove, while the W1 loop of RFX1 initiates most contacts to DNA with its basic amino acid residues at the major groove (Gajiwala & Burley, 2000). Similarly, the globular domains of linker histones H1 are winged helix domains, and their positively charged loop W1 surfaces bind to DNA (Gajiwala & Burley, 2000; Widom, 1998). Other non-canonical WH domains occur in Polycomb-like (PCL) proteins, PHF1, MTF2 or PHF19, which are associated with the mammalian PRC2.1 subcomplex and link PRC2 to CpG islands. In addition to a Tudor domain and two PHD fingers,

PCLs possess a winged helix domain that binds in an atypical fashion to DNA, with multiple contacts of the W1 loop to bases of CGG sequences. The role of the  $\alpha 3$  helix within the WH domain of PCL proteins is still unclear and discussed in the literature. PCL WHs were proposed to bind to DNA in a sequence-independent manner with a preference for unmethylated CpGs (Fischer et al., 2022; H. Li et al., 2017; Owen & Davidovich, 2022).

#### 1.4 SAMD1-like family of winged helix domains

Of particular interest for this work are members of the newly defined SAMD1-like family of WH domains. These belong to the group of non-canonical winged helix domains that were first described by Stielow and colleagues in sterile alpha motif (SAM) domain-containing protein 1 (SAMD1; alternative name Atherin) to directly recognize and bind to unmethylated CpG motifs. The WH domain of SAMD1 is a globular, N-terminally localized domain. Highly homologous regions are also found in the transcriptional regulators KAT6A, KAT6B, and ZMYND11. Functionally, KAT6A (also known as MOZ and MYST3) and its paralog KAT6B (MORF and MYST4) are human lysine acetyltransferases, and ZMYND11 binds to trimethylated lysine 36 of H3.3 (H3.3K36me3) (Wen et al., 2014; Yang, 2015). Due to their related sequences, these four proteins form a novel class of WH domains (Stielow, Zhou, et al., 2021). Similar to PCL proteins, SAMD1 was shown to interact with unmethylated CpG islands, and by interacting with chromatin regulators such as KDM1A and L3MBTL3, it modulates the chromatin landscape and has a transcriptionally repressive effect on its target genes (H. Li et al., 2017; Stielow, Zhou, et al., 2021). Experimentally, the SAMD1 WH domain was found to preferentially contact 5'-GCGC-3' motifs of dsDNA by simultaneously binding to the identified motif in the major groove of DNA with the C-terminal end of its  $\alpha 1$  helix and adjacent loop, as well as to the minor groove with two residues of the W1 loop, connecting the two-stranded  $\beta$ -sheet. This DNA-binding conformation differs not only from PCL proteins but also from all other known WH domains (Stielow, Zhou, et al., 2021). In the following, this thesis emphasizes the role of the newly identified family of SAM1-like winged helix domain proteins and their binding capacity to unmethylated CpG islands. In particular, SAMD1 and KAT6A are presented with their implications in different functional and biological contexts.

## 2. THE HISTONE ACETYLTRANSFERASE KAT6A IS RECRUITED TO UNMETHYLATED CPG ISLANDS VIA A DNA BINDING WINGED HELIX DOMAIN

Lisa Marie Weber, Yulin Jia, Bastian Stielow, Stephen S. Gisselbrecht, Yinghua Cao, Yanpeng Ren, Iris Rohner, Jessica King, Elisabeth Rothman, Sabrina Fischer, Clara Simon, Ignasi Forné, Andrea Nist, Thorsten Stiewe, Martha L. Bulyk, Zhanxin Wang and Robert Liefke

*Nucleic Acids Res.* 2023 Jan 25; 51(2): 574–594, doi: 10.1093/nar/gkac1188

### 2.1 Introduction

#### 2.1.1 Lysine acylation activity of the KAT6A complex

In 1996 KAT6A was first described by Borrow et al. and named “monocytic leukaemia zinc finger protein” (MOZ) due to a conserved C2HC-type zinc finger. This domain, together with adjacent sequences of high homology in human MOZ, yeast Ybf2 (Sas3), yeast Sas2 and human TIP60, is referred to with the acronym “MYST domain”. Five human MYST family members were identified, which share the feature of intrinsic lysine acetyltransferase capacity and underwent various renaming processes over the years: KAT5 (TIP60: HIV Tat-interacting 60 kDa protein), KAT6A (MOZ; MYST3), KAT6B (MORF: MOZ related factor; MYST4), KAT7 (HBO1: HAT bound to ORC1; MYST1) and KAT8 (hMOF: human males-absent-on-the-first; MYST2) (Borrow et al., 1996; Yang, 2015).

Acetyl CoA-binding motifs and intrinsic HAT activities of both human KAT6A and KAT6B were first described by Champagne and colleagues in 2001 and 1999, respectively (Champagne et al., 1999, 2001). In addition to the MYST domain, KAT6A and KAT6B share a double PHD finger (DPF) domain and an N-terminal region of high homology with the *Drosophila* HAT Enok. The DPFs can target and engage with the first 15 residues of histone H3, thus directing the acetylation activity of KAT6A/B to lysine residues within histone 3. In particular, DPFs of KAT6A bind to H3 tails with unmodified H3R2 and acetylated H3K14 but are repelled by H3K4me3 decorated tails, a mark that hinders H3 acetylation (Dreveny et al., 2014; Qiu et al., 2012). The N-terminally localized homologous region of KAT6A was initially called the NEMM (N-terminal part of Enok, MOZ or MORF) domain and has only recently been described to contain two different winged helices, with WH1 being a SAMD1-like WH domain able to directly interact with unmethylated CpG-rich DNA, first characterized by our group and Becht and colleagues for KAT6A and KAT6B, respectively, and the WH2 resembling a linker histone H1/5 WH (Becht et al., 2023; Weber et al., 2023; Yang & Ullah, 2007). The C-terminal part of KAT6A and KAT6B consists of an acidic region with high levels of glutamate and aspartate and a domain rich in serine and methionine (SM domain). Both of these domains can be found in zebrafish orthologues but are absent in dEnok and a related sea urchin protein, indicating vertebrate specificity of these regions (Wiesel-Motiuk & Assaraf, 2020; Yang, 2015). The acidic region has several implications in leukaemia and

developmental disorders, which will be further elucidated in later paragraphs. The SM domain was characterized as a potent transcriptional activation domain, a feature that has already been described for other acetyltransferases, such as p300 and CBP. Although KAT6A and KAT6B have an overall sequence identity of 60% and sequence similarity of 66%, they display differential substrate specificities and subtle distinctions in the activation capacity of the SM domain (Champagne et al., 1999, 2001).

KAT6A/B are the catalytic subunits of a multisubunit complex, which also contains the tumour suppressor family member ING5 (inhibitor of growth 5) or its paralog ING4, MEAF6 (MYST/Esa1-associated factor 6) and BRPF1/2/3 (bromodomain-PHD finger protein 1/-2 or-3) (Ullah et al., 2008; Zu et al., 2022). BRPF proteins directly interact with the MYST domains of KAT6A and KAT6B, mainly via two conserved N-terminal subdomains, acting as scaffolding proteins and mediating their connection to ING5 and MEAF6 by an enhancer of Polycomb (EPC) homology region. In this context, BRPF1 and ING5 were shown to stimulate the HAT domain activities of KAT6A and KAT6B to acetylate histones H3 and H4, highlighting their role as coactivators (Ullah et al., 2008; Zu et al., 2022). BRPF1 is a multidomain protein containing a bromodomain that interacts with acetylated lysines, PHD fingers that bind to methylated H3K4 and a PWWP domain with H2A- and H2B-binding capacity. This interaction is maintained even throughout mitosis, leading to recruitment of KAT6A and subsequent acetylation of histones H3 and H2A (Laue et al., 2008; Zu et al., 2022). Interestingly, in addition to H3 lysine-14 acetylation (Kac), the DPF domains of KAT6A also recognize other acylation marks, such as lysine-14 crotonylation (Kcr), as well as H3K14 butyrylation (Kbu) and propionylation (Kpr), suggesting a regulatory mechanism by which KAT6A controls transcription depending on the availability of corresponding acyl-CoA metabolites (X. Xiong et al., 2016). In addition to the recognition of a broader spectrum of acylation marks, KAT6A was also implicated in facilitating such marks and is thus also described as a propionyltransferase. In a complex with coactivator BRPF1, KAT6A MYST domain activity acetylates and propionylates H3K23 (K. Yan et al., 2020).

In addition to the various histone tail decorations, KAT6A was also found to facilitate acetylation of non-histone substrates. The ubiquitously expressed protein p53 is widely known as a tumour suppressor protein, and by interacting with its SM and HAT domains, KAT6A promotes acetylation of p53 lysines K120 and K382, leading to transcriptional induction of p53 target genes and subsequent premature senescence (Rokudai et al., 2013). Recently, SMAD3, a complex-bound transcriptional activator governed by various posttranslational modifications, was added to this short list. KAT6A acetylates K20 and K117 of SMAD3, promoting SMAD3 binding to the KAT6A-acetylated H3K23 reader TRIM24, an oncogenic chromatin modifier, resulting in closer chromatin association of this complex and enhanced signalling activation (B. Yu et al., 2021).

Overall, the KAT6A-ING5-MEAF6-BRPF1 complex has been widely studied and described to facilitate epigenetic modifications at a number of substrates in various organisms, including acetylation at histones H3K9, H3K14, H3K23, H2AK5, H4K5, H4K8, H4K12 and H4K16, propionylation at H3K23, and acetylation of nonhistone proteins p53 and SMAD3 (Laue et al., 2008; Rokudai et al., 2013; X. Xiong et al., 2016; B. Yu et al., 2021; Zu et al., 2022).

### 2.1.2 KAT6A implications in development and disease

KAT6A has a variety of genomic targets and plays an activating role in their transcription. Among the main high-confidence targets are the hematopoietic regulator *Runx1* (Runt-domain transcription factor 1; or *Aml1/Cbfa2*) and the transcription factor *Runx2* (*Aml3/Cbfa1*), which are involved in osteoblast differentiation and skeletal formation (Kitabayashi et al., 2001; Pelletier et al., 2002). KAT6A was shown to localize to the promoters of the homeobox gene *Hoxa9*, as well as to the oncogene *Myc*, both of which are known to be involved in the induction of leukemogenesis upon dysregulation (Becht et al., 2023; Camós et al., 2006; Qiu et al., 2012).

The involvement of KAT6A in leukaemia-associated translocations and its interaction with *Runx1* suggest a role of KAT6A in the regulation of haematopoiesis. Investigations in mice reveal that *Kat6a*-deficient embryos die at day E15.5 and display various phenotypical defects. They fail to oxygenate blood due to aortic arch defects and show facial deformations, including a cleft palate (Katsumoto et al., 2006; Thomas et al., 2006). Morphological analyses at day E14.5 reveal failures in liver development, haemorrhaging and subcutaneous oedema, originating from a reduced population of hematopoietic stem cells (Katsumoto et al., 2006). Heterozygous *Kat6a* knockout foetuses are viable but show significant dysgenesis of the thymus and spleen. An overall deficient development of the hematopoietic system is indicated by reduced numbers of lineage-committed precursor cells (Thomas et al., 2006). In addition, *Kat6a* is necessary for the maintenance of adult hematopoietic stem cells and their role in the long-term repopulation of bone marrow cells (Sheikh et al., 2016). Furthermore, the expression of *Kat6a* is essential for facilitating the active H3K9ac mark at *Hoxa* and *Hoxb* gene clusters and the correct development and identity specification of 19 body segments of the neck, trunk and nervous system in E10.5 mouse embryos (Voss et al., 2009). Since *Hox* gene activation is crucial for embryogenesis, their transcriptional expression is poised and dependent on dynamic interactions of chromatin modifiers. In concert with the transcriptional activator KAT6A, the canonical PRC1 member PCGF4 (BMI1) was identified as an opposing force during development and is involved in the transcriptional repression of *Hox* gene activation (Sheikh et al., 2015). Similarly, *moz*, the zebrafish orthologue of KAT6A, was shown to regulate *Hox* genes by maintaining their expression in the hindbrain and cranial neural crest cells (CNC), thereby determining pharyngeal segment identity (Miller et al., 2004). In particular, *Hoxa2b* and *Hoxb2a*, which play a role in second segment skeletal development of the jaw and gill cover, are under the influence of *moz* (Crump et al., 2006). Hence, due to the transcriptional maintenance of *moz* activity on *Hox* genes, *moz* was proposed to be a trithorax group (trxG) factor that regulates *Hox* gene expression in a similar way and is defined to genetically counterbalance PcG mutant phenotypes (Miller et al., 2004).

KAT6A is involved in various developmental processes, such as cardiogenesis and facial as well as pharyngeal development. Here, KAT6A catalyses acetylation of H3K9 at the *Tbx1* gene, a T-box DNA-binding transcription factor, thereby facilitating its expression. *In vivo* experiments showed a phenotype resembling DiGeorge syndrome after KAT6A depletion, with particular emphasis on its role in palate formation and the development of facial structures, the thymus and the

cardiovascular system (Voss et al., 2012). Deeper insight into the underlying mechanisms of craniofacial development revealed that the family of distal-less homeobox (Dlx) genes was particularly affected and less abundantly expressed upon *Kat6a* depletion (Vanyai et al., 2019).

Mutations of the *Kat6a* gene in patients with developmental delay and intellectual disability (ID) were first discovered in 2015 by clinically applied diagnostic whole-exome sequencing (WES). The described *de novo* mutations led to the expression of truncated versions of the KAT6A protein due to frameshifts caused by nucleotide insertions or deletions. Phenotypic consequences in affected individuals are hypotonia, cardiac defects, microcephaly, facial dysmorphism, oromotor dysfunction and speech impairment (Arboleda et al., 2015; Millan et al., 2016; Tham et al., 2015). Additional characteristics can involve strabismus, short stature, pituitary anomalies and forms of gastrointestinal complications (Kennedy et al., 2019; Trinh et al., 2018; Zwaveling-Soonawala et al., 2017). According to the initial discoverers, this form of ID was first referred to as Arboleda-Tham syndrome but is also called KAT6A syndrome. Most of the characterized mutations occur in the C-terminal half, particularly in the acidic domain of KAT6A. Accordingly, a correlation of the severity of developmental delay and the increased occurrence of other forms of the syndromic features was observed with C-terminally truncated pathogenic variants in the last two exons of KAT6A. These truncations result in a loss of the SM domain accompanied by SM domain interactors and potentially reduced p53 acetylation and dysregulated downstream signalling (Arboleda et al., 2015; Kennedy et al., 2019). To date, in 2019, the literature has described 61 different – mostly *de novo* – mutations in 78 individuals (Kennedy et al., 2019), and according to the homepage of the KAT6 foundation, in 2022, 350 patients were diagnosed with KAT6A syndrome worldwide (KAT6 Foundation, 2023).

Apart from developmental disorders, KAT6A was also implied to play an oncogenic role in different types of cancer. High expression levels of KAT6A were found in clinical glioblastoma (GBM) patients, promoting H3K23 acetylation, cell proliferation and tumour growth and correlating with a poor prognosis (Lv et al., 2017). Similarly, high levels of KAT6A in 15% of breast cancers, including ER+ and luminal breast cancer, correlate with poor survival. In aggressive forms of ER-positive breast cancer, KAT6A activates the estrogen receptor (ER)  $\alpha$  promoter (L. Yu et al., 2017).

Enhanced transcriptional activity can be caused by the dysregulation of chromatin modifiers such as histone acetyltransferases or demethylases. KAT6A was shown to be involved in such events by translocating to other HATs, thereby increasing intrinsic acetylation potential by the fusion of two HAT domains in one hyperactive protein, leading to hyperacetylation of target genes. Three different KAT6A translocations have thus far been discovered, all of which have been identified in patients suffering from acute myeloid leukaemia (AML). First, the translocation t(8;16)(p11;p13) describes KAT6A rearranged with CREB-binding protein (CBP) occurring mainly in acute myelomonocytic or monocytic leukaemia, subtypes of AML (Borrow et al., 1996). Second, t(8;22)(p11;q13) is a KAT6A to HAT p300 in-frame fusion protein that is rearranged in acute monocytic leukaemia (Chaffanet et al., 2000). Third, the pericentric inversion inv(8)(p11q13)



leads to a KAT6A-TIF2 fusion. TIF2 is a transcriptional coactivator that additionally contains a CBP/p300 interaction domain, which is retained in the fusion protein, thus recruiting CBP and p300 to the nucleosome and promoting AML (Deguchi et al., 2003; Yang & Ullah, 2007). KAT6A-related translocations are found in approximately 2% of AMLs, with the entire N-terminus of KAT6A being included in these rearranged fusion proteins, a breakpoint in the acidic region, and a replaced C-terminus with the respective fusion partners' including their respective HAT domain (Katsumoto et al., 2008; Zhou et al., 2020). They define a distinct subtype of leukaemia, displaying an aberrated gene expression profile in comparison to other myelomonocytic or monocytic AMLs (Camós et al., 2006; Murati et al., 2009; F. Yan et al., 2022).

Leukaemias with KAT6A fusion proteins have a very poor prognosis and a mean survival of  $\leq 5$  months, with a large proportion of patients represented by children. To date, haematopoietic stem cell transplantation is the only possible treatment option in these cases but is considered a high-risk procedure, emphasizing the need for other, more effective therapies and medical treatments (Zhou et al., 2020). Two KAT6A inhibitors against intrinsic MYST domain activity are applied in research that compete with acetyl-CoA binding. WM-8014 inhibits KAT6A with an  $IC_{50}$  of 8 nM but also interacts with other MYST family proteins (KAT6B,  $IC_{50}$  28 nM; KAT5,  $IC_{50}$  224 nM; KAT7,  $IC_{50}$  342). More efficient and specific is the WM-8014 derivative WM-1119, which was already effective in reducing the tumour burden of lymphoma in mice *in vivo* experiments (Baell et al., 2018).

In addition to KAT6A, other leukaemogenic translocations have been reported; for example, ENL (MLLT1) is a fusion protein with MLL (KMT2A/ALL1) in MLL-rearranged leukaemia. MLL and its fusion proteins are known to target unmethylated CpG islands at transcriptionally active promoter sites by their CXXC domain, similar to KAT6A and its WH1 domain. ENL was described as a histone acylation reader protein that binds to H3K9/18/27ac in active chromatin via its YEATS domain and sustains oncogenic gene transcription in AML. Due to its selectivity towards H3K9ac, ENL was found to cooperate with KAT6A in a histone acetylation “writer-reader” manner, driving MYB and MYC expression and promoting AML progression (F. Yan et al., 2022). A direct interaction of KAT6A with ENL was ascertained through multiple motifs contained in the KAT6A residues 781-1116, parts of the acidic domain further subdivided into an additional part called the ENL binding domain (EBD). Reciprocally, ENL binds KAT6A through its YEATS domain. ENL reportedly associates with several other enzymatically active proteins, such as histone methyltransferase DOT1L, MLL fusion protein AF4 (MLLT2) and components of the Polycomb repressive complex 1, via its ANC1 homology domain. Furthermore, a basic stretch between the DPF and MYST domains of KAT6A facilitates interaction with MLL, providing a dual direct link of KAT6A to MLL-ENL translocations (Miyamoto et al., 2020).

Overall, KAT6A plays a role in various developmental processes, especially in haematopoiesis and craniofacial growth. Aberrant expression of KAT6A has oncogenic effects at the onset of acute myeloid leukaemia as well as glioblastoma and breast cancer formation. Furthermore, pathogenic C-terminally truncated protein forms are the sole cause of KAT6A syndrome, a rare developmental disorder.

## 2.2 Synopsis of results

Since its first discovery nearly four decades ago, KAT6A has been viewed as a transcriptional coactivator, binding to histones and other proteins that are directly bound to DNA and chromatin, thereby influencing chromatin accessibility and acetylation progression. In a previous study, our group characterized the CpG island-binding capacity of SAMD1 by its N-terminally located winged helix domain and discovered a highly homologous – at that time – unannotated winged helix domain at the very N-terminus of KAT6A and KAT6B (Stielow, Zhou, et al., 2021). In the present study, we reveal that this KAT6A WH1 domain has a DNA-binding motif, directly recruiting KAT6A to unmethylated CpG islands, providing new and more detailed insights into KAT6A activity and chromatin association.

Structural predictions obtained with the AlphaFold database (Varadi et al., 2022) indicated two distinguishable globular domains contained in the NEMM region of KAT6A (Figure 1A). Each of these domains is composed of three alpha helices and two beta sheets, a classical conformation of WH domains (Harami et al., 2013), connected by a 25 amino acid-long linker (Figure 1B). Thus, we labelled these domains WH1 and WH2 (Figure 1C). Analyses of the human KAT6A paralog KAT6B and *Drosophila* orthologue Enok displayed homologous domain structures to KAT6A and reinforced the NEMM domain composition of two individual WH domains (Supplementary figure S1A-F). Minor variations occur in the length of the linker element between the WH1 and WH2 domains, comprising 30 amino acids in KAT6B (Supplementary figure S1C) and 5 amino acids in addition to a predicted beta strand proximal to the WH2 in Enok (Supplementary figure S1F). Amino acid sequence comparisons of the distinct WH1, WH2 and DPF domains between KAT6A, KAT6B and Enok (Supplementary figure S1G) revealed overall high similarities between the three WH1 domains (> 50%), while the three concerned WH2 domains share less identity (< 50%). The highest similarities were observed between KAT6A and KAT6B in the WH1 and DPF domains, with 81.3% and 86.1% shared identity, respectively. The pronounced variations in sequence homology, with particular emphasis on the differences between Enok and KAT6A/B, indicate evolutionary adaptation of the three domains to varying extents. Phylogenetic analysis with a focus on WH1 (Figure 1D) revealed homologues in zebrafish (*Danio rerio*) and the freshwater polyp (*Hydra vulgaris*), suggesting an evolutionarily ancient common ancestor. Related KAT6A/B WH1 domains were found in human and zebrafish SAMD1 and ZMYND11 (Figure 1D, E), composing a new group of SAMD1-like WH domains (Stielow, Zhou, et al., 2021). Since the SAMD1 WH domain was already shown to bind to CpG islands (Stielow, Zhou, et al., 2021) and published genome-wide data indicated a recruitment of KAT6A to CpG-rich sequences (Miyamoto et al., 2020), we hypothesized that KAT6A and KAT6B WH1 domains would also bind specifically to CpG-rich DNA motifs due to their sequence and structural homology. Indeed, electrophoretic mobility shift assays (EMSAs) showed that the KAT6A and KAT6B WH1 domains bind exclusively to CpG-rich DNA and not to AT-rich DNA, while the ZMYND11 WH domain indicates no such binding capacity (Figure 1F). Furthermore, the KAT6A WH2 domain was analysed and did not exhibit CpG binding in EMSA experiments (Supplementary figure S2A). To elucidate the particular DNA motif KAT6A

WH1 binds to, we applied an unbiased protein binding microarray (PBM) approach (Supplementary Table S3). Two motifs were highly enriched, the 5'-GCGCG-3' motif and the palindromic 5'-CCGNGGC-3' motif, indicated by an E-score greater than 0.45 in 96 and 125 octamers (8-mers), respectively (Figure 1G). In accordance with the EMSA results, PBM experiments elucidating the function of the WH2 domain revealed no significant binding motif when WH2 alone was analysed (Supplementary figure S2B) and overlapping motifs to the WH1 analysis when WH1-WH2 were combined (Supplementary figure S2C), rendering WH1 as the driving force in CpG recognition and binding.

To elucidate the molecular mechanisms of the KAT6A WH1 domain interaction with CpG-rich DNA, protein-DNA complex crystallization was performed. The KAT6A N-terminal amino acids 1-85 comprising the WH1 domain were crystallized with double-stranded DNA containing a 5'-GCGC-3' motif, and the structure was solved at a resolution of 1.5 Å (Supplementary table S1). As predicted, the complex structure reveals a folding of WH1 into three alpha helices ( $\alpha$ 1-3) followed by two beta sheets ( $\beta$ 1/2), with two loops (W1/2), one interspersing  $\beta$ 1 and  $\beta$ 2, and the other located C-terminal to both  $\beta$  sheets (Figure 2A). The main interactor with the presented sequence is  $\alpha$ 1, inserting its C-terminus into the major groove of DNA and thus facilitating sequence-specific contact (Figure 2B). The most important amino acids in this context are Gln23, Lys24 and Gln25, which form direct hydrogen bonds with the base pairs of the 5'-GCGC-3' binding motif (Figure 2B, C). In addition to the major groove interaction, W2 contacts the minor groove with its Arg79 residue (Figure 2C). To verify C and G as the essential base pairs for the KAT6A WH1-DNA interaction, we performed further EMSA experiments, exchanging one key base component at a time to one of the other three possible bases (Figure 2E). Furthermore, the binding affinity of WH1 to methylated CpG-DNA (mCpG) was probed (Figure 2F). The displayed results reinforce the binding specificity of KAT6A WH1 to unmethylated CpG-rich DNA motifs. To reciprocally elucidate the essential DNA-recognizing amino acids of KAT6A WH1, we mutated sequence-interacting amino acids found in the structural complex analyses and performed isothermal titration calorimetry (ITC) measurements (Figure 2G). Lys24 proved to be the most crucial amino acid by weakening the binding affinity of WH1 to DNA by approximately 30-fold when mutated to Ala (K24A), aided by Gln23, which negatively influences binding about 3-fold when mutated (Q23A). Comparing KAT6A WH1 with its homologous SAMD1 WH domain, both exhibit protein-DNA interactions in a similar fashion, as apparent in the superimposed structures of the complexes (Figure 2H). The SAMD1-like WH domains represent a new type of non-canonical WH domain, since they depend on the protrusion of their  $\alpha$ 1 helix into the major groove to allow motif recognition, instead of their  $\alpha$ 3 helix, as is categorized for canonical WH domains (Brennan, 1993).

A second palindromic motif was recognized in the PBM analysis of KAT6A WH1 binding motifs (Figure 1G). Considering the undefined base (N) in the centre of the palindromic motif, we conjectured that the molecular binding mechanisms might differ between the sequences. To elucidate this point, we repeated the protein-DNA complex crystallization described above with double-stranded DNA containing a 5'-CCGTCCG-3' motif and solved the structure at a resolution

of 1.93 Å (Supplementary table S1). Interestingly, KAT6A WH1 bound to this sequence in a dimeric form, exhibiting a head-to-head conformation (Figure 3A). In addition to the previously observed protein-DNA interactions, this complex is further connected by protein-protein interactions between Glu30 with Ser28 and Glu29 of the opposite WH1 molecule (Figure 3B). To test the stability and likelihood of occurrence of the two different complexes, gel-filtration analysis was implemented, assessing strong interactions of both the monomeric and dimeric KAT6A WH1-DNA complexes (Figure 3C). Superimposing the monomeric and dimeric structures showed a high correlation, varying only slightly in the minor groove protruding beta sheet, particularly at W1 and W2 (Figure 3D), a connection not essential for DNA binding of the WH1 domain. In contrast, the interaction of the key amino acids Gln23, Lys24 and Gln25 with the CGG bases of the palindromic motif is further emphasized in this conformation, complemented by additional hydrogen bonds to the phosphate backbone of the adjoining amino acid Arg26 (Figure 3E, F). Overall, both the monomeric and dimeric complexes form all essential connections between the C-terminus of each WH1  $\alpha$ 1 helix with the major groove of the DNA and the located CGC/CGG recognition motifs.

Due to the various implications of KAT6A in embryonic development with a particular role in haematopoiesis (Katsumoto et al., 2006; Sheikh et al., 2016; Voss et al., 2009), we set out to further investigate its function in gene regulation in mESCs. Applying the CRISPR/Cas9 technique (Ran et al., 2013), we implemented KAT6A knockout (KO) cells using two different gRNA constructs and validated the obtained single clones by Western blotting (Figure 4A), immunofluorescence microscopy (Figure 4B) and Sanger sequencing (Supplementary figure S3A). Concurrent with the histone acetyltransferase function of KAT6A, the acetylation levels at its targets H3K9 and H3K14 were slightly decreased upon KAT6A depletion (Figure 4C). Furthermore, RT-qPCR analysis confirmed the transcriptional dysregulation of multiple known target genes (Figure 4D) (Vanyai et al., 2019). Proliferation assays (Supplementary figure S3B), brightfield microscopy (Supplementary figure S3C) and colony formation assays (Figure 4E) of the KAT6A KO mESCs demonstrated increased proliferation, distinct morphology and reduced clonogenicity, respectively, in comparison to control cells.

To study the effect of KAT6A on genome-wide transcription, we performed RNA-Seq with the described knockouts. Principal component analysis revealed strong distinctions between the KO and control samples (Figure 4F). Gene expression was downregulated at 494 targets, a majority of which were localized to CpG islands (Figure 4H), and upregulated at 238 targets ( $\log_2$ -fold change  $> 0.75$  and  $P < 0.01$ ) (Figure 4G), concurring with a loss of the transcriptional activating capacity of KAT6A as a histone acetyltransferase in the KO cells. Applying gene set enrichment analysis (GSEA) to the obtained data, gene set categories relating to development (red) and signalling (green) were particularly negatively affected by KAT6A depletion, while gene sets of neural development (blue) were upregulated (Supplementary figure S3D). An interesting resemblance of affected gene sets upon KAT6A KO with SUZ12 target genes was observed in the GSEA, including both down- and upregulated targets (Supplementary figure S3D, E). To further

elucidate this overlap, we investigated SUZ12 expression at the mRNA (Supplementary figure S3F) as well as protein level (Supplementary figure S3G) and observed a reduction in both, concurring with the KAT6A KO. Additionally, the recruitment of the SUZ12 interactor and PRC2 complex member EZH2 to specific target genes was substantially reduced in KAT6A-depleted cells (Supplementary figure S3H). Overall, the observed effects of KAT6A deletion on the transcriptomic profile in mESCs correlated well with the implications of KAT6A in patients with the developmental syndrome (Kennedy et al., 2019). We concluded that a combination of direct and indirect consequences led to the dysregulation of targets, first, by compromised histone acetyltransferase activity of KAT6A and second, by altered expression of other affected genes, such as the chromatin modifier SUZ12.

Since we encountered technical difficulties in the immunoprecipitation of endogenous KAT6A and ectopic expression of KAT6A constructs in mESCs to study genome-wide binding sites and to further investigate the role of WH1 in chromatin recruitment and transcription control, we decided to utilize a different cell model. Henceforth, we conducted our research in HEK293 (human embryonic kidney 293) cells, which are more amenable to plasmid transfection. Ectopic expression of FLAG-HA (F/H)-tagged KAT6A full-length (fl) wild-type (wt) and WH1 mutant (K24A/Q25A, “WH1 mut”) constructs was successfully introduced into these cells (Figure 5A, Supplementary figure S5A, Supplementary figure S6A) to investigate the effects of abrogated DNA binding of KAT6A WH1. Employing cellular fractionation experiments, we found divergent enrichment of KAT6A wt and WH1 mut in the three fractions (Figure 5B). The chromatin interaction of mutated KAT6A WH1 was reduced, while a higher abundance was measured in the nucleoplasm and cytoplasm fractions in comparison to wild-type KAT6A (Figure 5C). Due to the indications that KAT6A WH1 might preferentially interact with CpG-rich DNA *in vivo*, we consulted previously published ChIP-Seq data (Miyamoto et al., 2020) and selected promising target genes for ChIP-qPCR experiments. While exogenous KAT6A wt specifically bound to CGI-containing promoters, exogenous KAT6A WH1 mut displayed no such recruitment (Figure 5D). To elucidate the degree of WH1 autonomy in KAT6A chromatin recruitment and binding, we cloned various constructs with mutated, depleted or truncated KAT6A domains (Supplementary figure S4A, D, G) and transiently transfected these constructs into HEK293 cells (Supplementary figure S4C, F, I). ChIP-qPCR experiments revealed that the WH1 wt domain alone is not sufficient for chromatin recruitment and barely differs from the WH1 mut fl or truncated construct (Supplementary figure S4AB). Gradually expanding the WH1 domain by each of the following annotated domains led to a stepwise increase in chromatin recruitment, which was already improved in the WH1-WH2 construct and restored to the KAT6A fl wt extent in the WH1-DPF construct (Supplementary figure S4E). Furthermore, KAT6A full-length constructs with deletions of either the WH2 ( $\Delta$ WH2) or DPF ( $\Delta$ DPF) domain did not abrogate chromatin binding (Supplementary figure S4AH). Thus, the WH1 domain of KAT6A is essential for chromatin recognition and recruitment but is, on its own, not sufficient in this function and requires the contribution of the WH2 and DPF domains.

To investigate the effect of KAT6A wt and WH1 mut overexpression on endogenous KAT6A expression in HEK293 cells, we used an antibody that bound to both forms. While the overexpression of wild-type KAT6A did not increase KAT6A recruitment to chromatin, the overexpression of WH1-mutated KAT6A significantly decreased KAT6A recruitment to chromatin (Figure 5E). This led to the conclusion that endogenous KAT6A binding to chromatin is already in a saturated state and therefore not enhanced by ectopic expression. In contrast, KAT6A WH1 mut likely has a competitive advantage and displaces endogenous KAT6A, thus reducing chromatin recruitment, since the global expression levels of both endogenous and exogenous constructs remain constant (Supplementary figure S5A). Considering that KAT6A has intrinsic histone acetylation activity, we included the H3K9ac and H3K14ac targets in our ChIP-qPCR analyses of the KAT6A overexpression constructs. Similar to general KAT6A recruitment to chromatin, we observed no significant alteration of the histone acetylation marks upon overexpression of KAT6A wt but a distinct reduction in acetylation upon overexpression of KAT6A WH1 mut (Figure 5E). We included an enzymatically inactive KAT6A fl HAT domain mutant construct (Q654E/G657E, “HAT mut”) (F. Yan et al., 2022) to ensure a better comparability of the effects on histone acetylation levels. Indeed, the HAT mut construct displayed as good a recruitment to CGI-containing promoters as the wild-type (Supplementary figure S5B), but H3K9ac and H3K14ac levels at the investigated targets were reduced to a strikingly similar extent as with the KAT6A WH1 mut construct (Supplementary figure S5C). Recapitulating these results, we concluded that the expression of ectopic KAT6A WH1 mut or HAT mut displaced or diminished the activity of endogenous KAT6A – a dominant negative effect.

To verify the hypothesis that KAT6A WH1 mut replaces endogenous KAT6A in its complex formation and other protein-protein interactions, we performed FLAG-immunoprecipitation with subsequent mass spectrometry (MS) to assess the interactomes of ectopically expressed KAT6A wt and WH1 mut. The immunoprecipitation (IP) displayed a multitude of interactors (Figure 5F), and the characterization of these proteins revealed closely resembling distributional patterns between both overexpression constructs (Figure 5G, Supplementary table S4). Already known interaction partners of the BRPF1 complex were confirmed in the MS findings, and putative interactors such as the NSL complex or PRC1 were enriched in both KAT6A wild-type and WH1 mut samples (Figure 5G, Supplementary figure S6B). To date, a direct interaction and coregulatory function for the KAT6A/B orthologue Enok with PRC1 has only been described in *Drosophila* (Bryan et al., 2021; Kang et al., 2017). With coimmunoprecipitation, we validated the interaction of ectopic KAT6A wt and WH1 mut with PRC1 subunits, in particular with non-canonical PRC1.6 (Figure 5H). In line with these results, ectopic KAT6A wt and WH1 mut interactomes strongly overlap and can competitively displace endogenous KAT6A in protein-protein interactions, thus influencing their chromatin association.

To determine if the observed alternating patterns of exogenous KAT6A wt, WH1 mut and HAT mut in chromatin binding at selected promoter sites were representative of genome-wide targets, we subsequently conducted ChIP-Seq experiments. Initially focussing on the wild-type form, we identified 20,903 significant binding sites in HEK293 cells, a majority of which

overlapped with CpG islands (Figure 6A). Correlating, most of these KAT6A-bound locations were promoter sites (Figure 5C) with a distinct accumulation of KAT6A wt at their transcription start sites (TSS) (Figure 6B). A motif analysis of these locations displayed CpG-rich motifs (Figure 6D), similar to the previously conducted PBM array (Figure 1G). A comparison of the KAT6A WH1 preference towards four nucleotide DNA motifs (4-mer) was applied to the obtained data of these *in vivo* and *in vitro* experiments, which correlated significantly (Pearson's  $R = 0.31$ ,  $P < 0.001$ ) with the 5'-GCCG-3' 4-mer as the most likely motif in both cases (Figure 6E). Furthermore, this motif is localized with higher frequency approaching the peak summit of the KAT6A binding profiles (Figure 6F). Considering that the PBM array and subsequent complex crystallization (Figure 3A-F) revealed a dimeric KAT6A WH1 binding state with a seven nucleotide palindromic CpG-rich sequence, we reviewed the occurrence of such a motif in the *in vivo* data. Indeed, comparing the palindromic motif to other non-palindromic motifs (Supplementary figure S7A) or palindromic motifs with partly exchanged CpG bases or increased spacer length (Supplementary figure S7B, C) at KAT6A binding profiles, we observed a distinct enrichment at the very peak summit. Therefore, KAT6A seems to prefer the palindromic motif, which is present at 58% of all CGIs (Supplementary figure S7D).

As already indicated by EMSA, KAT6A seems to have a preference for unmethylated CpG-rich DNA (Figure 2F). For that reason, we correlated our KAT6A wt CHIP-Seq data with published CpG island recovery assay (CIRA)-Seq (Miyamoto et al., 2020) and whole genome bisulfite sequencing (WGBS) data (Piunti et al., 2019). Evidently, KAT6A binds to unmethylated CGIs and is repelled from methylated CGIs (Figure 6G). To investigate this further and assess whether the removal of methylation from DNA at CGIs leads to *de novo* recruitment of KAT6A, we applied the DNMT1 inhibitor GSK-3484862 (Azevedo Portilho et al., 2021). We selected representative unmethylated and methylated CGI-containing promoter sites (Supplementary figure S8A) and observed changes in KAT6A recruitment upon treatment with different concentrations of GSK-3484862 in HEK293 cells. While the binding levels of ectopically expressed KAT6A did not vary at unmethylated target genes, methylated CGIs displayed *de novo* recruitment of FLAG-tagged KAT6A, positively correlating with increasing concentrations of the applied DNMT1 inhibitor (Supplementary figure S8B). These findings further support the specificity of KAT6A WH1 in binding to unmethylated DNA.

By analysing the genome-wide distribution of KAT6A and its colocalization with various histone marks, we identified three different groups of CpG islands: active, repressed and orphan CGIs (Figure 6H). Each of these groups is characterized by a unique combination of histone modifications, group 1 ( $n=9351$ ) by H3K4me2/3 and H3K9ac at active sites, group 2 ( $n=3158$ ) by H3K27me3 at repressed sites, and group 3 ( $n=8374$ ) by H3K4me2 and H3K9ac at orphan CGIs (Figure 6H, J). Additionally, a fourth and rather unexpected group of KAT6A-bound genes was identified, with KAT6A enrichment beyond the TSS, in downstream gene body regions (Figure 6I, J). A gene ontology analysis of the four identified clusters revealed that group 2 and group 3, in particular, are involved in cell-type specific developmental processes (Supplementary figure S9A). While the decoration of group 3 with H3K4me2 suggests an association of the KAT6A-bound

orphan CGIs with enhancer function (Pachano et al., 2021), group 2 strongly accumulates with repressive H3K27me<sub>3</sub>, a mark facilitated by Polycomb group proteins (Fischer et al., 2022). In accordance with our MS data and coimmunoprecipitation experiments, we found a high correlation between KAT6A and PRC1 components in ChIP-Seq data (Supplementary figure S6C, D), indicating a joint regulatory function in HEK293 cells. Closer studies of the gene body-linked KAT6A sites demonstrated high transcriptional expression of the associated genes (Supplementary figure S9B), correlating to an overlap with the histone elongation marks H3K4me<sub>1</sub> and H3K36me<sub>3</sub>, as well as polymerase II (Supplementary figure S9C, D, E, S6C). Including the ectopically expressed KAT6A WH1 mut and HAT mut constructs in our analyses, we aimed to further elucidate the role of the WH1 and HAT domains in genome-wide KAT6A chromatin recruitment and acetylation progression, respectively. In agreement with the previous ChIP-qPCR data (Supplementary figure S5B), WH1 mut completely abrogated KAT6A recruitment to CGIs at a genome-wide level, whereas the HAT mut showed no apparent difference in recruitment compared to wild-type (Figure 7A, B, E). However, in distinction from the binding mechanisms at CpG islands, KAT6A WH1 mut recruitment to gene body regions is not abrogated; on the contrary, it is slightly increased as opposed to wild-type and HAT mut (Figure 7C, D, E). Furthermore, the GC content of the observed gene body regions was distinctly lower than the general genomic GC content and correlated negatively with the binding of KAT6A WH1 mut (Supplementary figure S9C, D). This suggests that although the WH1 domain of KAT6A is essential for the recognition of and binding to unmethylated DNA at CGIs, other mechanisms recruit KAT6A to gene body regions of lower GC content, independent of WH1.

In summary, KAT6A is recruited to four different chromatin regions, each characterized by other histone modifications, and takes part in various cellular processes by influencing transcription control. A majority of these target sites is comprised of unmethylated CGIs, a direct interaction facilitated by the WH1 domain of KAT6A, while a smaller fraction is located in gene bodies, by recruitment mechanisms independent of WH1. Moreover, the inhibition of DNMT1 leads to a *de novo* recruitment of KAT6A to methylated CGIs.

In view of KAT6A histone acetyltransferase activity, we further employed genome-wide experiments with one of its main enzymatic targets, H3K9ac (Laue et al., 2008). Beforehand, we observed a minimal increase in H3K9/K14ac upon ectopic expression of KAT6A wt and a distinct reduction upon WH1 mut and HAT mut expression at specific CGI promoter target sites (Figure 5E, Supplementary figure S5C). Consequently, we correlated genome-wide H3K9ac occurrence with KAT6A-bound CGIs and observed changes in acetylation between KAT6A wild-type overexpression and both mutants (Supplementary figure S10A, B). Although the reduction of H3K9ac upon KAT6A WH1 mut and HAT mut overexpression was measurable compared to wild-type and control, a clear difference was observed between both mutants, illustrating the HAT mut as more effective in abrogating acetylation at CGIs than WH1 mut (Supplementary figure S10B, C, D). Furthermore, HAT mut-expressing cells displayed tendencies of a slightly lower abundance of a small number of KAT6A target genes in RT-qPCR experiments, correlating with the mild reduction of H3K9ac in these samples (Supplementary figure S10F). Similar to the enrichment of



KAT6A at CGIs containing the palindromic 5'- CCGNGGC-3' motif, H3K9ac was locally enriched at this motif (Supplementary figure S7E), echoing the involvement of WH1 in monomeric and dimeric CGI binding and its associated function in chromatin remodelling. However, taking another closer look at KAT6A-bound gene bodies, no apparent influence on H3K9ac was noticeable in these regions as a consequence of exogenous KAT6A expression (Supplementary figure S10E), even though KAT6A WH1 mut was prominently enriched at gene bodies (Figure 7D). In summary, slightly decreased levels of H3K9 acetylation were observed at KAT6A-bound CGIs genome-wide upon ectopic expression of KAT6A WH1 mut and HAT mut in comparison to KAT6A wild-type, supporting the role of both domains in histone acetylation.

Overall, we characterized a novel winged helix domain at the N-terminal end of KAT6A as a distinct feature of the previously described NEMM domain. Moreover, we developed a model of how this WH1 domain of KAT6A is involved in DNA motif recognition at unmethylated CpG islands and chromatin recruitment of KAT6A to genome-wide target sites to facilitate histone acetylation and enforce transcription control (Figure 8A). In addition, other unidentified mechanisms contribute to KAT6A recruitment to gene body regions of target genes, independent of WH1 function (Figure 8B).

## 2.3 Discussion

### 2.3.1 Substantial effects of a DNA-binding WH domain on genome-wide aspects

In the present study, we characterized a newly identified (Stielow, Zhou, et al., 2021) winged helix domain at the very N-terminus of the histone acetyltransferase KAT6A that is responsible for the direct recruitment of KAT6A to unmethylated CpG-rich DNA sequences. This is the first indication that KAT6A is not exclusively dependent on cofactors such as its complex partner BRPF1 for chromatin association (Zu et al., 2022). Nevertheless, interaction partners and cofactors of KAT6A might still play a role in processes such as complex association and stabilization of chromatin interactions, since KAT6A protein-protein interactions hardly differ between the wild-type and the WH1 mut version (Figure 5 F-H, Supplementary figure S6B-D). It would be interesting to investigate which targets are directly bound by the WH1 domain and which by cofactors such as BRPF1 that indirectly tether KAT6A to chromatin and how they overlap or differ, with the implications that arise for the developmental processes that KAT6A takes part in as a self-sufficient chromatin regulator or in concert with other complex members.

Interestingly, one source stated that initial *in vitro* experiments indicated that the KAT6A HAT domain itself might have DNA-binding properties, likely mediated through an N-terminal zinc finger region and a C-terminal HTH motif, therefore having both chromatin-targeting and enzymatic activities. These DNA-binding properties of the KAT6A HAT domain seemed not to be conserved in other members of the MYST family of acetyltransferases, therefore potentially playing a contributory role in AMLs caused by hyperacetylation due to KAT6A translocations (Holbert et al., 2007). Strategic chromatin binding experiments obtained by ChIP-qPCR of truncated or mutated KAT6A constructs were performed in the present study (Supplementary figure S4). Since no residual KAT6A chromatin recruitment was detectable in WH1 mut samples and a gradual extension of N-terminal domains showed no effect after the addition of the HAT domain compared to the WH1-WH2-DPF construct, a self-sufficient chromatin binding function of the KAT6A HAT domain seems unlikely. Moreover, Miyamoto and colleagues already identified KAT6A residues 1-84 to be particularly essential for KAT6A targeting to specific CpG-rich chromatin locations, such as the HOXA9 and MYC promoters, while analyses of the C-terminal part, including DPF and the acidic domain, showed low specificity and measurable chromatin enrichment only at sites colocalizing with H3K14ac, a known DPF target. They concluded a chromatin recognition motif contained in the very N-terminal region and a stabilization of this interaction by binding of this or further KAT6A domains to histones and other proteins. Since they copurified subunits of the poised RNA polymerase II (RNAP2) Ser5-P complex with this 84 aa long stretch, they suspected a direct interaction and possible recruitment mechanism to transcriptionally active genes and hence dubbed it the RNA polymerase II binding motif (RBM) (Miyamoto et al., 2020). Indeed, the study described here supports the overlap of KAT6A with RNAP2 at transcription start sites (Supplementary figure S6C, Supplementary figure S9), as does a second study (Becht et al., 2023). However, we have since discovered the structural integrity of

the WH1 domain and its DNA-binding function to unmethylated CpG-rich motifs, endorsing the direct interaction of KAT6A WH1 with chromatin independent of other cofactors but not excluding their role in further indirect KAT6A chromatin interactions. Additionally, the previously identified KAT6A linker H1/5-like WH domain (WH2) (Ullah et al., 2008) or the role of the KAT6A DPF domain was not included in the analyses of Miyamoto et al. (2020), whereas our study as well as Becht et al. (2023) acknowledge their contribution to chromatin recruitment and interaction stabilization. Therefore, a correlation of KAT6A with RNAP2 at transcription start sites might also coincide with the role of KAT6A as a transcriptional regulator and histone acetyltransferase in propagating histone acetylation, which induces an active chromatin conformation necessary for RNAP2 activity.

Aside from KAT6A and KAT6B, SAMD1 is the founding member of a newly defined conserved group of SAMD1-like winged helix domains (Stielow, Zhou, et al., 2021). Although all three proteins possess a very similar N-terminal winged helix domain binding to unmethylated CGIs, KAT6A/B, by default as histone acetyltransferases, have activating functions at target genes, while SAMD1 has an overall more repressive impact at target genes. Another considerable difference is that KAT6A WH1 can bind in a monomeric as well as dimeric structural formation at specific DNA motifs, while SAMD1, as far as we know, binds only in a monomeric form and to a distinctly smaller number of target genes at CGIs than KAT6A (Stielow, Zhou, et al., 2021). Furthermore, SAMD1 was not identified in our KAT6A MS data, as it was enriched in the KAT6A wt fraction compared to the KAT6A WH1 mut fraction (Figure 5G, Supplementary table S3), rendering a direct heterodimeric interaction of KAT6A WH1 with SAMD1 WH unlikely, in addition to the opposing roles of both proteins in cellular processes.

In the present study, we investigated the impact of KAT6A on H3K9ac at a genome-wide level (Supplementary figure S10) but did not include other known enzymatic targets, such as H3K14ac (X. Xiong et al., 2016) or H3K23ac (K. Yan et al., 2020), and therefore have limited insight into how other histone marks are affected by KAT6A WH1 or HAT mutations. Coincidentally, another research article concerning the KAT6A WH1 domain, among others, was published only a few weeks after our paper, which concentrated on H3K23ac as a histone mark facilitated by KAT6A as well as KAT6B (Becht et al., 2023). This study not only supports many of the results and WH1 characteristics we observed but also provides a good complementary addition to our own analyses, with nucleosome binding assays and *ex vivo* analyses of KAT6A/B fusion proteins.

Similar to our studies on KAT6A, Becht et al. (2023) focused on KAT6B and further characterized the previously identified N-terminal SAMD1-like winged helix domain - WH1 (Stielow, Zhou, et al., 2021). The double PHD finger domain of KAT6B associates with H3K14acyl, and next to H3K14ac, KAT6B primarily facilitates the H3K23ac mark, interactions and targets coinciding with KAT6A functionality. In EMSA experiments and ChIP-qPCR analyses with ectopically expressed C-terminally truncated KAT6B domain deletion constructs in K-562 lymphoblast cells, they observed KAT6B WH1-specific binding to unmethylated CGI-containing promoter sites (Becht et al., 2023), correlating with our data about homologous KAT6A WH1 (Figure 2E, F, Supplementary figure S4). Moreover, the depletion of KAT6B WH1 led to diminished

H3K14 and H3K23 acetylation levels at the investigated targets, as was the case upon KAT6A WH1 mutation with H3K9 and H3K14 acetylation (Supplementary figure S5C). Furthermore, Becht et al. (2023) identified the extranucleosomal linker DNA as the most likely interaction site for the KAT6B WH1 domain with CpG-rich DNA wrapped around euchromatic nucleosomes, matching well with the predicted binding of WH2 as a linker H1/5-like WH domain in an 'on-dyad' position (Saha & Dalal, 2021). Due to their high homology and overall conformity, this binding site specificity of KAT6B WH1 is most likely applicable to KAT6A WH1.

Interestingly, Becht et al. (2023) performed ChIP-Seq experiments with exogenously expressed FLAG-tagged truncated versions of KAT6A, as well as KAT6A and KAT6B WH1 domains alone, in human HEK293-T cells. Here, they confirmed that the WH1-dependent recruitment of KAT6B takes place at unmethylated CGIs genome wide, with an extension of the WH2-DPF domains further stabilizing this interaction, which they attributed to the H3K14ac recognition of the DPF. The observed binding of standalone KAT6A WH1 or KAT6B WH1 to genomic loci in ChIP-Seq data is opposing their former ChIP-qPCR analyses of KAT6B in K-562 cells where KAT6B WH1 was not recruited to selected CGIs at transcription start sites, as well as our ChIP-qPCR analyses in HEK293 cells where KAT6A WH1 alone was not sufficient for chromatin binding (Supplementary figure S4). A possible explanation for this phenomenon is the difference in the applied method. While our above-described ChIP-qPCR and ChIP-Seq experiments were conducted in a conventional way, the obtained ChIP-Seq data in (Becht et al., 2023) were acquired using the fanChIP (fractionation-assisted native ChIP) method. In fanChIP-Seq samples, protein-protein interactions of chromatin-bound complexes are not abrogated, essentially due to milder treatment conditions and the execution of subcellular fractionation instead of chromatin crosslinking (Miyamoto & Yokoyama, 2021). Therefore, considering that we detected the ability of KAT6A WH1 to bind in a dimeric form to a palindromic CpG-rich motif *in vitro* (Figure 3) and confirmed the accumulation of this motif at KAT6A peaks *in vivo* (Supplementary figure S7), the involvement of intact protein-protein interactions between dimeric KAT6A WH1 domains might provide a scaffolding factor in WH1 chromatin association when the fanChIP technique is employed.

Further emphasizing the parallels between the homologous KAT6A and KAT6B WH1 domains, genome-wide localization patterns of both to unmethylated CpG-rich promoters were highly similar and abrogated upon mutation of certain amino acids (K24A/R26A/K66A in KAT6B) or domain depletion (Becht et al., 2023).

### 2.3.2 A newly identified connection of KAT6A to PRC1

DNA methylation is a feature distinctly occurring in vertebrate genomes. Invertebrate species usually display only a low degree of genomic DNA methylation, and some organisms, such as *Drosophila melanogaster* or *Caenorhabditis elegans*, are nearly free of DNA methylation (Bogdanović & Veenstra, 2009). Importantly, CGIs represent the exception to the otherwise methylated genome in vertebrates and are associated with a high proportion of all promoters

(approximately 70%) (Deaton & Bird, 2011; Defossez & Stancheva, 2011). However, CpG islands can also be found in regions distal from transcription start sites, and these non-promoter CGIs are referred to as orphan CGIs and often show variability and tissue specificity (Defossez & Stancheva, 2011). The WH1-mediated recruitment of KAT6A to orphan CpG islands decorated with H3K4me2 and H3K9ac (Figure 6H, J) is an interesting new discovery, potentially linking KAT6A to enhancer functions (Owen & Davidovich, 2022) in addition to the known role in transcriptional activation at promoter sites. In mESCs, orphan CGIs were discovered to mediate and amplify the interaction of so-called poised enhancers (PEs) with their targets, usually promoters with large CGI clusters, by creating closer proximity. These poised enhancers have several characteristic features, including the binding of PcG complexes that contact respective target genes and take part in developmental processes, exemplarily in the induction of neural genes in mESC neural differentiation (Pachano et al., 2021). Potentially, in other species and cell types apart from murine ESCs, orphan CGIs provide a link between repressed Polycomb-bound regions and active CpG islands.

In addition, several potential connections between KAT6A and the PRC1 complex can be found in the published literature. Of particular importance, Enok, the KAT6A/B orthologue in *Drosophila*, has been confirmed to interact with dPRC1 (Bryan et al., 2021; Kang et al., 2017).

Evidently, KAT6A directly interacts with ENL, which in turn associates with components of the PRC1 complex (Miyamoto et al., 2020). Although colocalization studies of ectopically expressed ENL in HEK293T cells showed non-canonical PRC1 components such as PCGF4 and CBX8 at the transcription start sites of transcriptionally active KAT6A targets, including HOXA9 and MYC (Miyamoto et al., 2020), other PRC1 subcomplexes have not been investigated in this context. In contrast, experiments in mice suggest an antagonistic relationship between KAT6A and PCGF4 in regulating *Hox* gene expression in early mouse development, assisted by a bivalent chromatin state. In later developmental stages, KAT6A-controlled maintenance of HOX transcription supersedes (Sheikh et al., 2015). Furthermore, the depletion of KAT6A as well as its catalytic inhibition by treatment with WM-8014 resulted in the downregulation of EZH2, the enzymatically active subunit of the PRC2 complex (Baell et al., 2018; Sheikh et al., 2015). These observations coincide with our experiments in KAT6A KO mESCs, where global SU12 expression as well as EZH2 recruitment to genomic loci was impaired (Supplementary figure F-H). Moreover, the observed phenotype of KAT6A-KO mESCs in biological assays, the downregulation of SUZ12 and the correlation of transcriptomic changes with SUZ12 KO cells (Figure 4, Supplementary figure S3) are highly in accordance with previously published data, where SUZ12 was described as essential for proper differentiation but ineffective for proliferation (Pasini et al., 2007). Likewise, KAT6A-depleted mESCs did not display impaired cell viability (Supplementary figure S3B) but exhibited distinct morphological structures (Figure 4E, Supplementary figure S3C). Considering the involvement of KAT6A in developmental disorders and haematopoietic processes, a role in differentiation was already determined. SUZ12-depleted mESCs display abrogated levels of H3K27me3 and reduced H2K27me2 (Pasini et al., 2007), and it would be interesting to investigate whether H3K27me2/3 levels were indirectly affected upon KAT6A KO in mESCs as well.

Other studies explored KAT6B and observed an enrichment not only at active promoter sites but also at bivalent promoters in murine embryonic stem cells (mESCs). ING5, a complex member associated with either KAT6B or KAT6A, has a PHD finger domain able to bind to H3K4me<sub>3</sub>, therefore providing a recognition mode for one of the features represented at bivalent nucleosomes. However, apart from the known interaction of Enok with PRC1 in *Drosophila*, no evidence was presented that KAT6B or KAT6A directly interact with PRC1 in mice or other vertebrates as well (Bryan et al., 2021).

Our analyses seem to be the first to identify the PRC1.6 non-canonical subcomplex as a direct interactor of KAT6A (Figure 5H, Supplementary figure S6).

The majority of KAT6A target sites are composed of CpG islands located at gene promoters (Figure 6A, C, G). Concurrently, most PcG target genes overlap with CpG islands (Owen & Davidovich, 2022). However, not all PRC1 subcomplexes, canonical or non-canonical, have the same target site specificities, and cell type-specific differences in recruitment patterns occur. Canonical PRC1.2 and PRC1.4 strongly overlap with PRC2 and bind to the enzymatic product H3K27me<sub>3</sub>. Non-canonical PRC1.1, PRC1.3, PRC1.5 and PRC1.6 partly colocalize with their canonical counterparts but have additional target genes that are distinct between the subcomplexes (Owen & Davidovich, 2022; Stielow et al., 2018). Fittingly, these distinct targets of non-canonical PRC1 may carry active histone marks such as H3K4me<sub>3</sub> and H3K36me<sub>3</sub> and are therefore generally more transcriptionally active. PRC1.6-specific targets in particular are usually not bound by PRC2 or decorated with H3K27me<sub>3</sub>, and the PRC1.6 complex has several members with DNA-binding functions, additionally enabling chromatin interaction independent of RINGA/B (Owen & Davidovich, 2022; Stielow et al., 2018). Considering these aspects, it would be interesting to re-evaluate the genome-wide data of KAT6A and members of the PRC1.6 complex (Supplementary figure S6C) and assess their colocalization with the determined clusters according to histone modifications (Figure 6H). Nevertheless, a possible direct KAT6A interaction with PRC1.6 at active or bivalent CGIs does not exclude an indirect connection with PRC2 at repressed target sites and cooperation in transcription regulation, given the negatively affected expression of PRC2 members in KAT6A KO mESCs (Supplementary figure S3G-H). Furthermore, the histone binding protein L3MBTL2 is part of the PRC1.6 complex and has a conserved *Drosophila* orthologue, Sfmbt, that recruits dPRC1 to chromatin (Owen & Davidovich, 2022), thus providing a further link between PRC1.6 and KAT6A in their conserved ancestors Sfmbt and Enok.

Potentially, KAT6A has a similar relationship with PRC1 as was described for the KDM2A and KDM2B proteins, which possess CXXC domains binding to unmethylated CpG islands. Both KDM2A and KDM2B are bound to CGIs genome wide; they are strongly enriched at the centre of CGI-containing promoters but localize to CGIs remote from gene promoters as well. KDM2B, in particular, was shown to be additionally enriched at Polycomb-related CGIs (Farcas et al., 2012). According to Farcas et al. (2012), KDM2B directly interacts with PCGF1 and RING1B core complex members and therefore forms a PRC1.1 variant complex, tethering PRC1 to unmethylated CpG islands and allowing transcriptional silencing in this case. Functionality may differ in the case of a

KAT6A-associated PRC1.6 variant complex. We showed that the interactomes of KAT6A wt and WH1 mut were essentially unaffected (Figure 5 F-H, Supplementary figure S6B-D). Therefore, genome-wide ChIP-Seq analyses of PRC1 and PRC2 components underlying KAT6A WH1 mut and KAT6A KO constructs in comparison to KAT6A wild-type would be of interest to evaluate the tethering function of KAT6A WH1 for PRC1.6 and other Polycomb components and the general necessity of KAT6A as a linker of these factors to active chromatin, respectively.

### 2.3.3 Perspectives of KAT6A WH1 involvement in cancer and developmental disorders

In all three known KAT6A-rearranged fusion proteins, KAT6A-CBP, KAT6A-p300 and KAT6A-TIF2, the N-terminus of KAT6A is retained in its entirety, including all domains from WH1 to MYST, with breaking points in or around the acidic domain and translocations to the C-terminal parts of the respective partners, expressing both transcripts and creating large in-frame fusion proteins. For example, the breaking points in KAT6A and p300 occur at amino acids 1117 and 31, respectively (Chaffanet et al., 2000). Due to these translocations, gain-of function proteins arise, such as KAT6A-TIF2, which exhibits increased TIF2 transcription coactivation potential due to KAT6A intrinsic acetylation activity and TIF2-facilitated CBP/p300 recruitment, creating closer proximity of yet another histone acetyltransferase to chromatin, thus creating aberrant acetylation and contributing to an oncogenic myeloid transformation (Deguchi et al., 2003; Yang & Ullah, 2007). Chromatin association and recruitment of these hyperactive multidomain acetyltransferases to transcriptionally active CpG islands is ensured by the N-terminus of KAT6A. Therefore, the loss of DNA-binding domains of fusion partners, such as the basic HTH motif in TIF2, is likely compensated by the DNA-binding WH1 of KAT6A.

Indeed, experiments of transduced KAT6A/KAT6B-TIF2 fusion proteins with depleted or mutated WH1 or WH2 domains, respectively, in myeloid progenitor cells isolated from murine bone marrow cells were performed to determine their role in leukaemogenic activity (Becht et al., 2023). Cells with transduced KAT6A $\Delta$ WH1-TIF2 deletion or KAT6B $\Delta$ WH1-TIF2 deletion displayed drastically reduced clonogenicities in colony-forming unit (CFU) assays with semisolid media, while fusion protein-expressing cells with KAT6B WH2 mutation showed a slight impairment in the formation of colonies. The characteristic capacity to immortalize haematopoietic progenitors was retained in KAT6A/KAT6B-TIF2 wild-type cells. Additionally, both WH1-depleted fusion proteins failed to activate HOXA9 transcription, indicative of the role of KAT6A/B WH1 domains in binding to unmethylated CGIs and their significant contribution to the oncogenic activity of KAT6A/KAT6B translocated fusion proteins (Becht et al., 2023).

Further analysis of KAT6A, either in a monomeric or rearranged conformation, in human leukaemias would be highly interesting. Choosing the best cell lines as a model system for *in vitro* experiments is an essential step, however. A previous study found that KAT6A and MLL N-terminal parts in rearranged fusion proteins are functionally interchangeable due to their translocation or protein-protein interaction with either p300 or CBP, thus targeting similar chromatin sites

(Miyamoto et al., 2020). Human AML cell lines have diverse genetic alterations by default, such as MOLM-13 and NOMO-1 (MLL-AF9 translocation), MV4-11 (MLL-AF4 translocation), HL-60 (MYC amplified) and U937 (CALM-AF10 translocation) (F. Yan et al., 2022), which might already influence KAT6A target genes and should therefore be considered in future analyses.

Taking a closer look at the genomic data of KAT6A syndrome patients, one finds that approximately half of the characterized mutations occur in the acidic domain, and in general, most mutations are localized in the C-terminal half of KAT6A, leading to truncated versions of the protein (Kennedy et al., 2019). This allows the hypothesis that the N-terminal half, including its functional domains, WH1, WH2, DPF and MYST, is more important for normal development. Some patients whose mutations were analysed by WES and published in scientific journals exhibited amino acid variations due to nucleotide mutations in the WH1 or MYST domain, such as p.Ala16Pro, Asn65Lys and p.Arg79\* or p.Tyr607Leu and p.Asn643Ser, respectively (Kennedy et al., 2019; Millan et al., 2016; Zwaveling-Soonawala et al., 2017). Of note, none of the known mutations occur at the essential WH1 domain DNA-interacting amino acids (aa 23-25) or at the MYST domain acetyl CoA-binding site (aa 654-660), supporting the particular importance of these motifs for KAT6A chromatin recruitment and acetylation function. Only Arboleda et al. (2015) included histone marks in their study of three KAT6A syndrome patients and found decreased acetylation levels in fibroblasts of one patient only at various histones, including KAT6A targets H3K9, H3K14 and H2AK5, among others. Therefore, no conclusive data on the effects of KAT6A *de novo* mutations on general histone acetylation levels or aberrant chromatin compaction have been obtained from a larger cohort of patients. Furthermore, p53, as one of the main non-histone targets of KAT6A, was shown to be dysregulated due to mutations leading to premature stop codons and translocations omitting the C-terminal part of KAT6A. Since both the MYST domain and SM domain are required for p53 interaction and K120 and K382 acetylation, a truncation or loss of the SM domain might already suffice for reduced p53 acetylation and thus negatively influence p53 signalling, potentially promoting its oncogenic function in leukaemia and other cancers (Arboleda et al., 2015; Rokudai et al., 2013).

Considering that mutations in the KAT6A gene are the sole determining factor in the onset of developmental delay and other phenotypes that describe KAT6A syndrome, the essential role of histone acetyltransferase in the development and cellular coordination of transcription is emphasized. We characterized a novel function of the KAT6A WH1 domain in direct chromatin interaction and associated progression of histone acetylation. These implications of KAT6A WH1 function, also in aberrantly expressed KAT6A as translocations to other HATs and resulting hyperacetylation, provide new opportunities for the development of targeted inhibitors and potential treatment strategies for patients with KAT6A-rearranged acute myeloid leukaemia.



## 2.4 Contribution statement

This study was conceptualized by Dr. Robert Liefke, and he and I designed the molecular and cell biological experiments, a majority of which I conducted. In particular, I made the following contributions to this project:

- Molecular cloning of CRISPR-Cas9 plasmids targeting murine KAT6A and establishment of knockout cells in mESCs. Validation of acquired knockouts by Western blotting, immunofluorescence microscopy and genomic DNA PCR (Figure 4A, B, Supplementary figure S3A). KAT6A KO cell morphology and proliferation were assessed by biological assays and light microscopy (Figure 4E, Supplementary figure S3B, C). Execution of RT-qPCR and RNA-sequencing experiments addressing transcriptional deregulation upon KAT6A depletion and distribution of target genes to CpG islands (Figure 4D, F-H, Supplementary figure S3D, E), where sequencing was performed by Dr. Andrea Nist and bioinformatics analyses were performed by Dr. Robert Liefke. Additionally, I conducted ChIP-qPCR and Western blotting of KAT6A enzymatic targets H3K9/14ac and PRC2 components (Figure 4C, Supplementary figure S3G, H).
- Molecular cloning of human KAT6A wild-type and truncated or domain depleted overexpression constructs, as well as WH1 mut and HAT mut constructs. Here, Iris Rohner provided expertise and technical support. Establishment of stably expressing HEK293 cell lines for subsequent cell biological experiments. Validation of ectopic expression in Western blotting, immunofluorescence and subcellular fractionation assays of full-length KAT6A wild-type and WH1 mut proteins (Figure 5A-C, Supplementary figure S4, S6A). ChIP-qPCR analyses of endogenous KAT6A and exogenous wild-type, WH1 mut and HAT mut (Figure 5D, E, Supplementary figure S4, S5). Here, I was supported by Dr. Bastian Stielow. Furthermore, I performed expression analysis of KAT6A target genes by RT-qPCR upon overexpression of KAT6A constructs (Supplementary figure S10F).

All figures and tables included in this publication were compiled by Dr. Robert Liefke and me, with contributions from Prof. Zhanxin Wang (Figure 1F, 2, 3, Supplementary figure 2A, table S1) and Prof. Martha Bulyk (Figure 1G, 6E, F, Supplementary figure S2B, C, table S3). Furthermore, I contributed to writing the manuscript together with Dr. Robert Liefke.

### 3. INVESTIGATION OF SAMD1 ABLATION IN MICE

Bruce Campbell, [Lisa M. Weber](#), Sandra J. Engle, Terence R. S. Ozolinš, Patricia Bourassa, Robert Aiello and Robert Liefke

*Sci Rep. 2023; 13: 3000, doi: 10.1038/s41598-023-29779-3*

#### 3.1 Introduction

Sterile alpha motif domain-containing protein 1 (SAMd1) is a 56 kDa protein whose gene maps to the human chromosome 19p13.12 locus. The protein is conserved between mammals and consists of two annotated domains, an N-terminal winged helix domain and a name giving C-terminal SAM domain, interspersed by a central disordered region. In subcellular fractionation assays of murine cells, SAMd1 was shown to localize to all cellular compartments with particularly high abundance in the nucleus. Furthermore, SAMd1 exhibits chromatin recruitment and a direct interaction with DNA (Stielow, Zhou, et al., 2021; Tian et al., 2021). Linked to these findings, the winged helix domain of SAMd1 was identified as the first of its kind and the founding member of the SAMd1-like family of non-canonical WH domains, a homologous region that it shares with KAT6A, KAT6B and ZMYND11. SAMd1-WH was established as a DNA-binding winged helix specifically recruited to unmethylated CpG islands, which are typically associated with active promoter sites, where it preferentially binds to a 5'-GCGC-3' motif (Stielow, Zhou, et al., 2021). Consistently, several published mass spectrometry data and a single nucleotide polymorphism (SNP) interactome analysis support the association of SAMd1 with CpG-rich DNA and a repulsive force of methylated DNA (Perino et al., 2018; Prestel et al., 2019; Viturawong et al., 2013; J. Xiong et al., 2016). Moreover, the SAMd1-WH domain contributes to protein-protein interactions with the histone H3K4me2 demethylase KDM1A (lysine-specific demethylase 1A; further known as LSD1) (Stielow, Zhou, et al., 2021). In comparison, proteins containing SAM domains appear commonly in eukaryotes and are associated with a variety of cellular processes, including cytoplasmic scaffolding, transcriptional repression and lipid metabolism. They mediate the formation of protein-protein interactions by their capacity to homo- and hetero-oligomerize with other SAM domains, as well as interactions with non-SAM proteins, DNA or RNA (Knight et al., 2011; Lees et al., 2005; Meruelo & Bowie, 2009). Applying computational models, the SAM domain of SAMd1 was identified as having a high potential to polymerize, as was lethal (3) malignant brain tumour (L(3)mbt)-like 3 (L3MBTL3), a member of the malignant brain tumour (MBT) family of histone readers, among several others (Knight et al., 2011; Meruelo & Bowie, 2009). Recent crystallography experiments with SAMd1 led to analogous conclusions and demonstrated the self-association of SAMd1-SAM domains into pentameric circles. Additionally, the SAMd1-SAM domain contributes to the efficient chromatin recruitment of the SAMd1-WH domain and facilitates interactions with other SAM domain-containing proteins, such as L3MBTL3 and L3MBTL4 (Stielow, Zhou, et al., 2021).

While SAMD1 target sites are predominantly located at unmethylated CGI-containing promoter sites in published genome-wide studies, they vary significantly in an organism- and cell type-specific manner, depending on DNA methylation and CGI size, with only a small proportion of overlapping genomic loci (Simon et al., 2022; Stielow, Zhou, et al., 2021). Overall, SAMD1 was proposed to be a unique chromatin regulator and transcriptional repressor at target CGIs in interaction with the KDM1A demethylase complex and SAM domain polymerization partners, such as L3MBTL3 (Stielow, Simon, et al., 2021; Stielow, Zhou, et al., 2021; Tian et al., 2021).

SAMD1 was first mentioned under the name “Atherin” in 2005 within the context of a study deciphering the mechanisms underlying atherosclerosis (Lees et al., 2005). One appearance in the onset of atherosclerosis is the retention of low-density lipoprotein (LDL) at susceptible sites of the artery wall, initially resulting in the accumulation of leukocytes and platelets, followed by immune cells and the formation of macrophages. These macrophages, in turn, ingest lipids and evolve into foam cells. Together with vascular smooth muscle cells (VSMCs) and proteoglycans, the so-called atherosclerotic plaque is formed in the tunica intima, an elastic layer of the blood vessel touching the endothelium (Lees et al., 2005; Tian et al., 2021). In this connection, SAMD1 was identified as an arterial wall protein, focally appearing in atherosclerotic lesions of the intima and binding to LDL. Due to this exclusive expression of SAMD1 in lesions and not in healthy arteries and the extracellular localization in the atherosclerotic plaque, SAMD1 was attributed to be a secreted protein with a pathological role in atherogenesis (Lees et al., 2005).

Subsequently, the involvement of dysregulated microRNAs (miRs), short single-stranded noncoding RNAs, in the development of atherosclerosis was demonstrated. Screenings emphasized that miR-378c was significantly suppressed in atherosclerotic plaques, as well as in the blood of acute coronary syndrome patients, while SAMD1 expression was upregulated in these tissues. The hypothesis that SAMD1 is a downstream target of miR-378c was raised, and indeed, a direct interaction with the 3'UTR of the *SAMD1* gene was demonstrated *in vitro* (Tian et al., 2021). Furthermore, SAMD1 was identified as a relevant factor for VSMC proliferation and was demonstrably distributed in the cytoplasmic compartment, as well as in the nucleus and membrane fraction of these cells. In addition to its nuclear role in transcriptional regulation, SAMD1 located at the cell surface interacts with LDL and thus promotes its retention in atherosclerotic lesions and the formation of foam cells (Tian et al., 2021). The human family of miR-378 microRNAs possesses a common seed sequence for mRNA targeting and has been implicated in several diseases, such as the enhanced tumour invasiveness of glioma upon miR-378 decrease (Tian et al., 2021). Thus, it is likely that SAMD1 might also be involved in these processes.

Coincidentally, experiments deciphering the role of SAMD1 in carcinogenesis were performed, focussing on the liver cancer cell line HepG2, due to the high expression of SAMD1 in hepatocellular carcinoma (HCC). Here, a role in cell proliferation was also attributed to SAMD1, and the protein mainly localized to the nucleus, abundant in the nucleoplasm as well as in the chromatin-bound fraction (Simon et al., 2022). In HepG2 SAMD1 KO cells, specifically, a distinct downregulation of MYC target genes was observed, as well as an upregulation of PIK3IP1, a

negative regulator of PI3K signalling. Overall, the investigated gene expression changes upon SAMMD1 depletion resulted in a more favourable gene signature of HepG2 cells, supporting the oncogenic characteristics of SAMMD1 in hepatocellular cells (Simon et al., 2022).

Furthermore, the biological function of SAMMD1 in mESCs was assessed in undirected differentiation experiments. While the phenotype of SAMMD1 KO cells did not vary from the wild-type control, deregulations in gene expression patterns became obvious with increasing cell differentiation, correlating with a repressive function of SAMMD1 at its target genes. Interestingly, immune system- and metabolism-associated pathways were downregulated upon SAMMD1 KO, while neural development- and cardiac muscle contraction-associated pathways were upregulated, drawing a manifold picture for SAMMD1 function in biological processes (Stielow, Zhou, et al., 2021).

One consistent direct binding partner and target of SAMMD1 in a variety of investigated cell types and biological processes is L3MBTL3 (Prestel et al., 2019; Simon et al., 2022; Stielow, Zhou, et al., 2021; Viturawong et al., 2013). Upon SAMMD1 depletion, L3MBTL3 is significantly upregulated, supporting the notion that this occurs due to a derepression and potential compensatory effects of L3MBTL3 for the loss of its interaction partner SAMMD1 (Simon et al., 2022). Furthermore, the SAMMD1 interaction partner KDM1A was shown to be recruited to chromatin by L3MBTL3, particularly to Notch target genes, leading to their transcriptional repression (T. Xu et al., 2017). In this context, an N-terminal domain (aa 1-64) seems to be required for the direct recruitment of L3MBTL3 to chromatin, further mediated by the interaction with RBPJ, a Notch signalling regulator (T. Xu et al., 2017). Nevertheless, this does not exclude indirectly facilitated recruitment mechanisms of L3MBTL3 through hetero-oligomeric SAM domain interactions with SAMMD1, further supported by the observation of approximately 50% abrogated L3MBTL3 chromatin binding upon SAMMD1 KO in mESCs (Stielow, Zhou, et al., 2021). Furthermore, the interaction of SAMMD1 with L3MBTL3 and KDM1A and the ascertained recruitment mechanisms of KDM1A to chromatin most likely contribute to the observed dysregulation of H3K4me2 levels at CpG islands in SAMMD1 KO cells (Simon et al., 2022; Stielow, Zhou, et al., 2021; T. Xu et al., 2017).

Overall, SAMMD1 likely has multifaceted biological roles, from pluripotent cell differentiation to cancer development, differing strongly in chromatin recruitment between organisms and cell types, but is largely involved in the modulation of chromatin and transcriptional repression of target genes at unmethylated CGIs. Major interactors are the demethylase KDM1A and corepressor L3MBTL3 (Simon et al., 2022; Stielow, Zhou, et al., 2021; Tian et al., 2021).

### 3.2 Synopsis of results

In the publication presented here, we provide the first description of homozygous and heterozygous SAMMD1 knockout mice and their respective phenotypes in embryogenesis. Furthermore, wild-type and SAMMD1 knockout murine embryonic stem cells (mESCs) were used in neural differentiation experiments to gain insights into the role of SAMMD1 in neurogenesis.

Choosing a recombination-mediated genetic engineering approach, we deleted the genomic region containing all five coding exons of the SAMD1 gene and established C57BL/6J SAMD1 knockout mice (Figure 1a). SAMD1 gene expression was validated in mouse embryonic fibroblasts (MEFs) isolated from genotypic SAMD1 wild-type (WT), heterozygous (HET) and homozygous knockout (KO) mice by RT-qPCR (Figure 1c). Considering the expected genotypic Mendelian ratios (1:2:1) of a litter, crossing of two heterozygous mice resulted in deviant ratios, with visible numbers of embryonic lethal and resorbed KO embryos between E12.5 and E14.5 (Supplementary Table S1) and slightly reduced numbers of born HET fetuses (Supplementary Table S2). When WT mice were crossed with HET mice, similarly aberrant ratios of SAMD1 HET mice were observed (Supplementary table S3). Otherwise, HET mice were born alive and displayed no obvious phenotypic alterations in comparison with WT mice. Abnormal phenotypes and dead SAMD1 KO embryos were first detected at embryonic day (E) 14.5 (Figure 1b). Phenotypic analyses at day E14.5 revealed pink coloured amniotic fluid and a lack of blood vessels in the yolk sac of SAMD1 KO embryos in comparison to WT (Figure 1d). Furthermore, the knockout embryos were visibly smaller, with a paler surface, an evident oedema covering their entire backside, as well as a placenta less lined by blood vessels and weaker umbilical cord than their wild-type opposite (Figure 1e). SAMD1 KO embryos displayed a slight variety of phenotypes. One specimen at E15.5 had no cranial vault, exencephaly and a hypertrophic brain; additionally, it displayed several intact blood vessels but also signs of a failing circulatory system (Figure 1f). The collapse of lymphatic and blood vessels and resulting formation of subcutaneous oedema were consistently observed in the investigated SAMD1 KO embryos. Using public gene transcription data (Davis et al., 2018; Deng et al., 2014) in addition to isolated mRNA of the WT embryos presented here, we observed high SAMD1 expression at different stages of embryonic development and in various tissues, including after birth (Figure 1g, h, Supplementary figure S1). These findings emphasize the versatility of the biological contexts in which SAMD1 is involved during embryogenesis and in tissue-specific development.

The assessment of cellular defects in SAMD1 KO embryos at day E14.5 was largely obtained by histological stains, including haematoxylin and eosin (H&E) as well as immunohistochemistry. Sagittal sections of WT, HET and KO embryos were stained by H&E and displayed severe degradation of KO tissues, varying strongly from WT and HET, which were normally developed and indistinguishable from each other (Figure 2, Supplementary figure S2). Inspecting the heart, cardiac muscles had formed in the SAMD1 embryos, but a tissue structure was barely perceivable, and the atria, ventricle and other cardiac attributes appeared collapsed (Figure 2a). Similarly, the structures of the lung and liver were disintegrated in SAMD1 KO mice, bronchioles were decidedly smaller (Figure 2b) and although right and left lobes had formed (Figure 2c), their size was also smaller compared to WT. In none of these three organs were clusters of red blood cells (RBCs) in their respective blood vessels observable, contrary to WT stains. We therefore suggest that fragmentation and degradation of SAMD1 KO tissues occurred due to an impaired circulatory system.

Since the main morphological features of SAMD1 KO mice at E14.5 were their paleness of skin (Figure 1e) and lack of RBCs (Figure 2), we sought to investigate the appearance of blood and lymph vessels by immunohistochemistry stains of endothelial cells (ECs) using antibodies against endothelial marker proteins CD31 (cluster of differentiation 31; PECAM1) and VEGFR2 (vascular endothelial growth factor receptor 2) with haematoxylin counterstains. Sections of the skin revealed RBCs in defined capillaries (Figure 3a, b) and blood vessels (Figure 3a, c) in the dermis and subcutaneous tissue of WT embryos, respectively, while SAMD1 KO skin displayed apparently unorganized and scattered clusters of extravascular RBCs (Figure 3a-c), reminiscent of microhaemorrhages. While CD31 staining in WT tissue showed surface staining of adjoined cells indicative of capillaries (Figure 3d, e), staining in SAMD1 KO tissue showed fragmented lines and appeared rather cytoplasmic or nucleic, suggestive of necrosis or phagocytosis. In addition, the epidermis of SAMD1 KO skin was severed from the remaining dermis and subcutaneous tissue in these sections and therefore not included, very likely due to previously observed oedema formation (Figure 1e). However, the epidermis was retained at the skull and extremities; hence, we additionally stained the forepaws of SAMD1 KO mice. Here, too, only a few scattered clusters of RBCs and nucleic CD31 stains of formerly connected ECs were apparent (Supplementary figure S3). The discernible patterns of RBCs and ECs in SAMD1 KO mouse skin tissue suggest a previously intact blood vessel system that functionally transported RBCs to the observed sections but failed at a certain point of development and subsequently degraded.

Furthermore, we carried out a closer inspection of the blood vessel distribution in internal organs, with a focus on the lung (Figure 4a-c), liver (Figure 4d, e) and heart (Figure 5) in E14.5 SAMD1 KO specimen in comparison to WT. An overall lack of RBCs and ECs is evident in the three tissues of SAMD1 KO mice, as well as integral structural degradation. Notably, spindle-shaped VSMCs were visible adjacent to ECs in higher magnification microscopy pictures of lung and liver blood vessels of WT animals (Figure 4b, e; black arrows), which have previously been connected with SAMD1 and the formation of plaques in atherosclerotic lesions (Lees et al., 2005) and were absent in SAMD1 KO animals. Marginal CD31 and VEGFR2 staining in lung and liver tissue was mainly condensed to a nucleic distribution and therefore attributed to necrotic vessel wall ECs (Figure 4b, c, e; red arrows, Supplementary figure S4a). Furthermore, close-up sections of the hearts of SAMD1 KO mice revealed distinct degradation of cardiac muscle cells, epithelial cells and endothelial cells (Figure 5 b-e). Examples of healthy versus degraded ECs in WT and SAMD1 KO coronary vessels, respectively, are indicated by black arrows (Figure 5b, c, e). Phagocytosis of necrotic cells was also visualized by staining the heart tissue of KO embryos (Supplementary figure S4b). A pronounced difference was observed in sections of the pulmonary trunk, which completely lacked intact RBCs and showed no sign of VSMC development (Figure 5d, black arrow) in SAMD1 KO mice, suggesting that cardiac development was arrested before VSMC differentiation. Overall, the blood vessel system of the studied internal organs failed before E14.5 in SAMD1 KO mice, resulting in a lack of oxygen supply and degradation of tissues.

Analyses of bone and skeletal muscle structure of WT and SAMD1 KO mice were performed with rib tissue sections (Figure 6). While the general organization showed similarities (Figure 6a),

closer examinations revealed that rib bones of SAMD1 KO mice were clearly ossifying with a high percentage of hypertrophic chondrocytes in comparison to intact embryonic WT chondrocytes at E14.5 (Figure 6b, Supplementary figure S4c). Moreover, muscle fibres were detached and feeble (Figure 6c), in strong contrast to the WT situation. Where larger pools of RBCs were present in the WT tissue, SAMD1 KO displayed only a few nucleated RBCs and immune cells (Figure 6d). Overall, the observed defects are traceable to a failure of blood vessel maturation and RBC delivery throughout SAMD1 KO tissues and a resulting hypoxia. Although hypoxia is essential for the endochondral ossification of cartilage tissue in bone development (Schipani et al., 2001), here, we observed clear premature ossification in SAMD1 KO rib bones at E14.5. For a better overview, all detected phenotypic alterations in SAMD1 KO embryos were briefly summarized (Table 1).

To gain deeper insights into the underlying molecular mechanisms of the observed morphological defects in SAMD1-depleted embryos, we utilized previously published RNA-Seq data of undirected differentiation of WT and SAMD1 KO mESCs (Stielow, Zhou, et al., 2021). This approach relies on the function of the cytokine LIF (leukaemia inhibitory factor) in maintaining the pluripotency of mESCs, and upon deprivation, it drives their differentiation in an undirected manner, demonstrated by the formation of embryoid bodies (EBs) that contain cells of all three primary germ layers (Gao et al., 2014; Luckey et al., 2011). Multiple cellular pathways displayed transcriptional alterations upon SAMD1 deletion at day seven of differentiation (Supplementary figure S5). While pathways related to brain development, such as synaptic assembly and correlating genes, were derepressed in the course of differentiation in SAMD1 KO cells (Supplementary figure S5a, b), pathways connected with angiogenesis were severely downregulated, as well as indicative factors such as *Pecam1* and *Vegfa*, which were consistently lower in SAMD1 KO cells than in WT cells at all observed stages of differentiation (Supplementary figure S5c, d). Moreover, pathways involved in cardiac chamber morphogenesis and development and several connected transcription factors were upregulated upon SAMD1 KO mESC differentiation (Supplementary figure S5e, f). These analyses support a role for SAMD1 in the regulation of neurogenesis, angiogenesis and cardiac development. Considering the known function of SAMD1 as a transcriptional repressor, we conclude that direct as well as indirect effects must participate in cellular dysregulation upon SAMD1 depletion.

In addition to the transcriptional studies conducted here, other indications were presented that SAMD1 might play a role in the development and function of neural cells, as well as enhance the invasiveness of glioma (Stielow, Zhou, et al., 2021; Tian et al., 2021). Hence, we investigated the influence of SAMD1 on neurogenesis in a directed differentiation approach of mESCs.

During the last decades, various *in vitro* cell culture protocols have been established and published aiming for neural differentiation of cells extracted from embryonic brain tissue or stem cells to circumvent limited cell numbers and working with embryonic tissue *in vivo*. The pluripotency and self-renewing capacities of mouse embryonic stem cells (mESCs), initially isolated from the inner cell mass (ICM) of blastocysts, render them an ideal cell model for the

investigation of embryonic development (Gao et al., 2014; Luckey et al., 2011). Out of the ICM, three definitive multipotent germ layers are formed: mesoderm, endoderm, and ectoderm. The definitive ectoderm gives rise to precursor cells of the neural lineage (Shparberg et al., 2019). Eventually, the brain is composed of neuronal and glial cell types that mature either to neurons or oligodendrocytes and astrocytes, among others. These processes of neural differentiation, self-renewal and proliferation are highly specified and underlie epigenetic control (Salinas et al., 2020). We hypothesized that SAMD1 is likely involved in these mechanisms and participates in cellular alterations during differentiation due to its established role as a chromatin modulator in connection with KDM1A.

Here, we established a neural differentiation protocol from E14 mESC control and SAMD1 KO cells (Figure 7b, Supplementary figure S6) (Stielow, Zhou, et al., 2021) to further elucidate SAMD1 function and necessity during neurogenesis. In contrast to undirected differentiation, directed differentiation in general requires the further addition of chemicals and growth factors. Neural differentiation in particular is triggered by additives such as retinoic acid (RA; or all-trans retinoic acid, ATRA), fibroblast growth factor (FGF-2), N-2 and B-27 under serum-free conditions (Brewer et al., 1993; Gao et al., 2014; Kim et al., 2009; J. Xu et al., 2012). The applied three-step protocol (Figure 7a) was adapted from Bibel et al. (2004), Conti et al. (2005) and Spiliotopoulos et al. (2009) and comprised the following: first, the formation of three-dimensional EBs took place under LIF removal conditions; second, the development of neural progenitor cells (NPCs) by the addition of RA; and third, the maturation to neuronal cells under serum-free conditions with N-2 and B-27 supplemented media (N2B27). At the stage of EB formation, we observed a variety of dysregulated genes (Figure 7c, Supplementary table S4). Of the observed SAMD1 target genes, *L3mbtl3* was particularly derepressed when SAMD1 was absent, in accordance with previous publications (Simon et al., 2022; Stielow, Zhou, et al., 2021), whereas neural marker genes were slightly downregulated at this early time point of unspecified differentiation, as was also observed in the reanalysed transcriptional study (Supplementary figure S5b). Interestingly, genes related to the cardiovascular system, especially *Myh6* (Myosin-6), a marker of cardiac muscle cells, were strongly upregulated in SAMD1 KO cells, while pluripotency markers remained largely unaffected. The subsequently established NPCs displayed no distinct morphological features between control and SAMD1 KO cells (not shown), but a slight increase in proliferative capacity was measured (Figure 7d). During neuronal differentiation, proceeding from NPCs, we monitored SAMD1-associated target genes, as well as markers for embryonic development and neural cell types, at different time points (Figure 7e). Consistent with previous observations, the expression levels of *Samd1* and *L3mbtl3* continuously displayed an inverse relationship in control and SAMD1 KO cells. While *Foxa2* (hepatocyte nuclear factor 3-beta) and *Nes* (nestin), marker genes for embryonic endoderm development and for neural development and neural progenitor cells, respectively, decrease, neuron-specific marker gene *Tubb3* (tubulin beta-3 chain) and astrocyte marker *Gfap* (glial fibrillary acidic protein) strongly increase, indicating generally successful neuronal differentiation. In contrast to the neuron and astrocyte markers *Tubb3* and *Gfap*, the oligodendrocyte marker oligodendrocyte transcription factor 1 (*Olig1*) displays little alteration



during differentiation. While *Tubb3* mRNA expression was only mildly different between control and SAMD1 KO cells during neuronal differentiation (Figure 7e), representative immunofluorescence staining of the neuron marker Tuj1 (*Tubb3*) at day six of cellular transition clearly depicted an increase in the Tuj1-positive signal in SAMD1 KO cells (Figure 7f). Furthermore, immunofluorescence staining of differentiated SAMD1 KO and control cells for histone modification H3K4me2 revealed visibly elevated H3K4me2 levels in the KO situation (Figure 7f). Quantitative analysis of H3K4me2 levels at different stages of the neural differentiation procedure displays a significant upregulation of the histone mark in SAMD1 KO cells correlating with an increasing progression of cell development towards neurons (Figure 7g). Dysregulation of the H3K4me2 mark upon SAMD1 deletion is consistent with previous observations in undifferentiated mESCs and most likely results from an impairment of the direct SAMD1 interactor KDM1A (Stielow, Zhou, et al., 2021). Overall, the obtained data suggest that SAMD1 participates in neural differentiation and is required for the proper function of the demethylase KDM1A during this process.

Finally, we focussed on the effects of a heterozygous SAMD1 deletion in the investigated animal model. Although SAMD1 HET mice were born alive and morphological stains at E14.5 were indistinguishable from WT mice (Supplementary figure S2), statistical genotyping analyses at postnatal (P) day 21 revealed negatively diverging numbers of expected Mendelian ratios of SAMD1 HET to WT animals (Supplementary table S2, S3). To assess potential phenotypes related to nutrition and metabolism, we monitored key body measures of HET mice under several conditions. Already under regular diet conditions, HET mice generally weighed less than WT animals (Figure 8a). Additionally, they were less prone to gain weight under high-fat diet (HFD; 60% fat by kcal) conditions over the time frame of 15 weeks (Figure 8b) and developed less adipose tissue (Figure 8c) than WT animals. Accordingly, oral glucose tolerance tests (OGTTs) revealed more effective glucose disposal and decreased fasted glucose levels after HFD feeding in HET mice (Figure 8d, e), as well as distinctly reduced insulin levels during and after HFD feeding in HET mice compared to WT mice (Figure 8f, g). Analyses of cholesterol distributions in HET and WT animals exhibited a pronounced reduction in HDL (high-density lipoprotein) and a slight increase in VLDL (very low-density lipoprotein) and LDL in HET animals after HFD (Figure 8h). Related to these alterations in lipid metabolism, steroid hormone expression in HET mice was altered, as angiotensin II, testosterone, corticosterone and aldosterone were generally upregulated compared to WT mice (Figure 8i). It is likely that the reduced weight and development of fat depots arise due to aberrant steroidogenesis in HET animals. Overall, embryogenesis in heterozygous SAMD1 KO mice was not affected, but a phenotype of metabolism-related defects was observed in HET animals fed a high-fat diet, resulting in their impaired developmental capacity.

In conclusion, *in vivo* experiments attest to an essential function of SAMD1 for proper embryogenesis, and loss results in severe organ degradation prior to E14.5 due to hypoxia caused by functionally deficient blood vessel maturation and a failing circulatory system. Heterozygous

SAMD1 knockout mice were born alive but showed signs of aberrant metabolism and steroidogenesis. Furthermore, *in vitro* studies revealed impaired directed neural differentiation of mESCs upon SAMD1 deletion.

### 3.3 DISCUSSION

To date, only a limited number of publications have specifically investigated the role of SAMD1 in diverse biological contexts, leaving a wide range of potential future prospects in scientific research. Here, we present the first characterization of SAMD1 KO mice and their severe phenotypic defects during embryogenesis. The degraded morphology of tissues in SAMD1 KO embryos and a directed differentiation approach of mESCs indicate particular roles of SAMD1 in the developmental processes of angiogenesis, cardiogenesis and neurogenesis (Figures 3, 5, 7, Supplementary figure S5).

The presented characterization of SAMD1 KO mice mainly relies on histological tissue analyses. We report a definitive degradation and lack of endothelial cells and thereby directly connected failure and fragmentation of the circulatory system in several internal organs (Figures 3, 4, 5). Therefore, we attribute a developmental and regulatory role in angiogenesis and potentially also arteriogenesis to SAMD1, in addition to the previously known localization to atherosclerotic lesions and involvement in the formation of plaques (Lees et al., 2005). While the endogenous detection of SAMD1 using immunohistochemistry would have been a valuable addition to the presented data, the perfunctory observation that VSMCs were completely absent at distinctive sites such as the pulmonary trunk (Figure 5d) is in line with a previous publication that determined a role of SAMD1 in the proliferation of VSMCs (Tian et al., 2021). Inconveniently, no other cell-specific marker proteins were investigated in the presented histological study, except for the endothelial markers CD31 and VEGFR2, thus providing an incomplete picture of the underlying mechanisms of organ degradation, other than failure of the circulatory system and the resulting oxygen and nutrient deprivation.

Interestingly, features of the observed phenotype of SAMD1 KO embryos at day E14.5 strongly resemble previous reports of KAT6A KO embryos at the same developmental stage, in the appearance of subcutaneous oedema, haemorrhaging and deficient liver development (Figure 1d, e, 2c, 3a) (Katsumoto et al., 2006). In addition, abrogation of SAMD1 or KAT6A affects heart and blood vessel maturation, including observed defects in pulmonary trunk (Figure 2a, 3) and aortic arch development (Voss et al., 2012), respectively, implying vital roles in angiogenesis and development of the cardiovascular system for both proteins. In this context, it was previously demonstrated that the facilitation of H3K9ac through KAT6A is significant for the active transcription of the *Tbx1* gene locus, in contrast to other loci regulating heart development (Voss et al., 2012). Comparative ChIP data of cardiovascular SAMD1 KO and KAT6A KO cells at the *Tbx1* locus and other genes, such as *Gata4* or *Tbx5*, could provide insight into the collaborative or exclusive nature of SAMD1 and KAT6A function in the transcriptional regulation of cardiogenesis.

Although macroscale pictures of H&E staining and closer inspection of embryonic development in the heart, lung and liver were obtained in WT, heterozygous and homozygous SAMD1 KO mice at day E14.5 (Figure 2, Supplementary figure S2), a similar procedure was not undertaken with sections of the brain. Here, only brain sections of the cerebral cortex of HET animals were presented (Supplementary figure S2), which, like all other organs at this stage, were potentially indistinguishable from WT sections, but no comparable samples of homozygous KO brain tissues are shown. Considering the full-scale images of SAMD1 KO embryos at E14.5 and E15.5 (Figure 1d-f) and their depicted phenotype of exencephaly and craniofacial defects, among others, histological sections and H&E stains of different regions of the brain would have been another worthwhile addition. In addition to a morphological characterization of SAMD1 KO neuroanatomy, further elaboration of endothelial cells and blood vessel development in the brains of these animals would be of interest. Specifically, ECs in the brain have a particular function in the formation of the blood-brain barrier to prevent the passage of pathogens or potentially neurotoxic components (Z. Zhao et al., 2015). Therefore, other or more severe defects could arise in the brains of SAMD1 KO embryos in addition to the fragmentation and cellular degradation that was observed in other tissues and internal organs.

Differentiation of mESCs by the formation of embryoid bodies imitates embryonic development in an *in vitro* approach and has comparable features to murine blastocysts. They have the potential to differentiate in a directed manner towards mesoderm, endoderm or ectoderm lineages by the application of appropriate cytokines and other growth factors. Nevertheless, the development of such protocols took many years of extensive research, and optimization of methods for neural ectoderm differentiation was needed due to rather heterogeneous populations of neural cells and early cell death (Bibel et al., 2004; Keller, 2005). Here, we implemented a protocol established by Bibel et al. (2004) to achieve homogenous populations of neural precursor cells after embryoid body formation and treatment with RA that have a high capacity to become neurons. Subsequently, by applying a monolayer approach, we prompted neural commitment of these mESC-derived NPCs by supplying the specified medium with N-2 and B-27 as well as the growth factors EGF, FGF-2 and BDNF for optimal preservation of cell viability and maturation of neurons (Spiliotopoulos et al., 2009). Indeed, as was observable with the measurement of marker genes at various time points of neuronal differentiation (Figure 7e), the neuron marker *Tubb3* gradually increased over time, but even so and to a stronger extent, did the astrocyte marker *Gfap*, contrary to the expected ratios. Interestingly, differentiation towards oligodendrocytes was not triggered, indicating the formation of a mixture of neuronal and astroglial fates. Potentially, a decrease or removal of EGF would have resulted in purer populations of neuronal cells, as was observed by Conti et al. (2005) for NPCs derived from neurospheres. Although only slight differences in gene expression were observed between WT and SAMD1 KO cells, immunohistochemistry staining of neuronal differentiation at a later stage revealed a highly abundant signal in SAMD1 KO cells and visibly matured neurons in contrast to WT cells (Figure 7f). Therefore, SAMD1 may have a repressive or controlling function in neuronal differentiation, and a knockout leads to aberrant progression of neuronal maturation. A genome-

wide transcriptome analysis at this stage would certainly be interesting to further discern changes in gene expression upon SAMD1 KO and characterize its impact on neural cell fate and differentiation potential.

Considering the phenotype of SAMD1 KO embryos in heart development and deregulated pathways connected to angiogenesis and cardiac morphogenesis in transcriptional studies of differentiated mESCs (Figure 2a, 5, Supplementary figure S5c-f), another intriguing *in vitro* approach might be differentiation towards specialized cells of the cardiovascular system derived from the definitive mesoderm germ layer. For example, endothelial cells or cardiomyocytes have already been successfully differentiated from human or murine embryonic stem cells (Jiang et al., 2012; Yamashita, 2007). These investigations could further decipher the role of SAMD1 during cell differentiation processes and unravel cell type-specific differences and similarities in lineage commitment.

Determination of global H3K4me2 expression revealed increasingly upregulated levels in SAMD1 KO cells with progression of neuronal differentiation (Figure 7f, g). These observations strongly concur with previous observations in mESCs, where CGIs gain H3K4me2 signals upon SAMD1 deletion (Stielow, Zhou, et al., 2021). In contrast, a SAMD1 knockout in HepG2 cells displays negatively dysregulated H3K4me2 expression, although local effects vary at enhancer and promoter sites that are either up- or downregulated, respectively (Simon et al., 2022). Accordingly, SAMD1 deletion affects the chromatin landscape in an organism- and cell type-specific manner. Further explorations at which sites H3K4me2 is mainly affected in the SAMD1 KO neural cells and comparisons with ChIP-Seq data of mESCs would be interesting to determine changes in CGI occupancy during differentiation, as well as other histone marks like H3K4me3. The significance of histone methyltransferases SETD1 and MLL1/2 or EZH2 for the facilitation of H3K4me3 or H4K27m3, respectively, for cell differentiation in general and for proper neurogenesis in particular has already been determined (Berry et al., 2020; Salinas et al., 2020). Potentially, the histone demethylase KDM1A and, by association, SAMD1 have a similar function in chromatin modulation during differentiation through the regulation of H3K4me2 levels. Taking a closer look at SETD1, its complex member CFP1 (CXXC1) mediates recruitment to unmethylated CpG-rich chromatin sites, and a loss of CFP1 results in a significant decrease in H3K4me3 levels at CGIs (Thomson et al., 2010). This chromatin binding mode and indirect interference with the H3K4 methylation status of CFP1 seems reminiscent of SAMD1. Genome-wide data further support an association of SAMD1 with CFP1 and H3K4me3, among others, at active CpG islands (Stielow, Zhou, et al., 2021), suggestive of complementary or cooperative roles of CFP1 and SAMD1 at promoter sites.

In addition to the influence of SAMD1 on chromatin regulation and differentiation, proliferation assays revealed enhanced viability of SAMD1 KO NPCs compared to control cells (Figure 7d). As observed before, cell type-specific differences occur in the proliferation of SAMD1 KO cells between NPCs and HepG2 cells, which show slightly less proliferative tendencies than their respective controls (Simon et al., 2022). In contrast, undifferentiated SAMD1 KO mESCs display no aberrant proliferation (Stielow et al., 2021). Cell cycle analyses of undifferentiated and

differentiated SAMD1 KO mESCs by flow cytometry could provide a potential tool to further investigate the impact of SAMD1 on distinct cell cycle phases and the functional and developmental implications thereof.

Finally, it would be highly interesting to investigate the role of the two defined globular SAMD1 domains in cell differentiation and, more particularly, in the context of neural differentiation. While the SAMD1-WH domain was shown to bind to unmethylated CpG-rich DNA and facilitate interaction with KDM1A, the SAMD1-SAM domain has the ability to homo- and hetero-oligomerize with other SAM domains (Stielow, Zhou, et al., 2021). To discern their involvement in this developmental process and the dependency on their respective chromatin or protein-protein interactions thereof, one potential approach could be rescue experiments with ectopic overexpression of SAMD1  $\Delta$ WH and  $\Delta$ SAM domain constructs in mESC SAMD1 KO cells followed by neural differentiation.

Overall, SAMD1 has various associations with different diseases and developmental processes. While initial characterizations proposed a secreted extracellular localization of SAMD1 in atherosclerotic lesions (Lees et al., 2005), later studies proved an intracellular localization in all compartments from cytoplasm to membrane, nucleoplasm, and chromatin-bound fractions, with strong implications as a repressor of transcriptional regulation (Stielow, Zhou, et al., 2021; Tian et al., 2021). Although recent years have recorded substantially increased numbers of publications seeking to decipher the molecular structure and epigenetic function of SAMD1 in particular cellular contexts, many aspects remain to be uncovered. The above presented study provides the first description of SAMD1 abrogation in mice and demonstrates the fatal consequences on embryonic development and tissue integrity. Moreover, a directed differentiation approach of mESCs displays dysregulated aspects of neurogenesis upon SAMD1 depletion, further indicating the role of SAMD1 during cell development.

### 3.4 Contribution statement

I have made the following contributions to this project:

- Establishment of a neural differentiation protocol from mESCs to neural precursor cells (NPCs) using an embryoid body (EB) approach.
- Implementation of a protocol for the subsequent differentiation of NPCs into neuronal cells.
- Analysis of differentiated SAMD1 KO and control cells at distinct stages by immunofluorescence, western blot, proliferation assays and RT-qPCR, as well as the assessment of statistical significance, as presented in Figure 7.
- Preparation of the manuscript.

## 4. SUMMARY

This thesis addresses the functions of two transcriptional regulators; histone acetyltransferase and transcriptional activator KAT6A (MOZ), and transcriptional repressor SAMD1 (Atherin), in their respective molecular and cell type-specific roles.

Firstly, we identified a novel winged helix domain at the N-terminal end of KAT6A as a distinct feature of the previously described NEMM (N-terminal part of Enok, MOZ or MORF) domain, a homologous region with other histone acetyltransferases. This KAT6A WH1 domain possesses a DNA-recognition motif, directly recruiting KAT6A to unmethylated CpG-rich DNA. Analyses of genome-wide ectopic KAT6A binding sites in HEK293 cells revealed a high correlation with unmethylated CpG islands (CGIs), particularly with CGIs associated with active promoters, decorated with the histone marks H3K4me3 and H3K9ac. We demonstrated that KAT6A WH1 is necessary for the observed recruitment to CGIs, but by itself not sufficient and dependent on KAT6A WH2 and DPF domains for effective KAT6A-CGI interactions. Indeed, mutations of the KAT6A WH1 domain caused a complete abrogation of KAT6A binding to CGIs, but increased the binding to gene body regions at a subset of target genes, indicating that recruitment mechanisms to these sites are independent of KAT6A WH1. This study, for the first time, demonstrates a direct chromatin recruitment mechanism of a histone acetyltransferase. Moreover, it provides new and more detailed insights into KAT6A chromatin recognition and target site association to facilitate histone acetylation and enforce transcription control.

Secondly, we morphologically characterized homozygous and heterozygous SAMD1 knockout (KO) mouse embryos and their respective phenotypes in embryogenesis. The homozygous deletion of SAMD1 was embryonic lethal, while heterozygous SAMD1 KO mice were born alive. At embryonic day E14.5 severe degradation of internal organs and formation of subcutaneous oedema were visible in SAMD1 KO embryos, most likely due to disturbed blood vessel maturation and resulting hypoxia. The observed defects led to the conclusion that SAMD1 is required for functional angiogenesis and the development of the cardiovascular system. Furthermore, craniofacial defects upon SAMD1 KO indicated malfunctions in head and brain development. To gain further insight into the role of SAMD1 during neurogenesis, we established a directed neural differentiation approach of murine embryonic stem cells (mESCs). Transcriptional dysregulation and the global increase of H3K4me2 were observed in SAMD1 KO cells during this process, further supporting a functional role of SAMD1 in neural cell development.

Overall, this study sheds light on the underlying molecular mechanism of KAT6A binding to chromatin and emphasizes the similarities of KAT6A WH1 to the previously described homologous SAMD1 WH domain, which directly interacts with unmethylated CGIs and facilitates transcription regulation, as well. Both KAT6A and SAMD1 are involved in a variety of developmental processes and dysregulations are associated with cellular abnormalities and the onset of cancer.

## 5. ZUSAMMENFASSUNG

Diese Dissertation befasst sich mit den Funktionen zweier Transkriptionsregulatoren; Histonacetyltransferase und Transkriptionsaktivator KAT6A (MOZ) sowie Transkriptionsrepressor SAMD1 (Atherin) in ihren jeweiligen molekularen und zelltypspezifischen Zusammenhängen.

Zunächst identifizierten wir eine neuartige Winged-Helix Domäne am N-terminalen Ende von KAT6A als eigenständiges Merkmal der zuvor beschriebenen NEMM Domäne (N-terminaler Teil von Enok, MOZ oder MORF), einer homologen Region mit weiteren Histonacetyltransferasen. Diese KAT6A-WH1 Domäne besitzt ein DNA-Erkennungsmotiv, das KAT6A direkt an unmethylierte CpG-reiche DNA rekrutiert. Die Analyse genomweiter Bindestellen ektopischen KAT6As in HEK293-Zellen ergaben eine starke Korrelation mit unmethylierten CpG-Inseln (CGIs), insbesondere mit CGIs, die mit aktiven Promotoren assoziiert sind, gekennzeichnet von den Histonmodifizierungen H3K4me3 und H3K9ac. Wir haben gezeigt, dass KAT6A-WH1 für die beobachtete Rekrutierung zu CGIs zwar notwendig, aber alleine nicht ausreichend ist. Die KAT6A-WH2 und -DPF Domänen unterstützen die wirksame Interaktion von KAT6A-WH1 mit CGIs. Tatsächlich verursachten Mutationen der KAT6A-WH1 Domäne eine vollständige Aufhebung der KAT6A-Bindung an CGIs, erhöhten jedoch die Bindung an Genkörperregionen bei einer Untergruppe von Zielgenen, was darauf hindeutet, dass die Rekrutierungsmechanismen an diesen Stellen unabhängig von KAT6A-WH1 sind. Diese Studie zeigt erstmalig die direkte Bindung einer Histonacetyltransferase mit Chromatin auf, und bietet darüber hinaus neue und detailliertere Einblicke in die KAT6A Erkennung von Chromatin und dessen Zielgenassoziation, essentielle Mechanismen für die charakteristische Funktion von KAT6A in Histonacetylierung und Transkriptionskontrolle.

Zweitens haben wir homozygote und heterozygote SAMD1-Knockout (KO) Mausembryonen und ihre spezifischen Phänotypen während der Embryogenese morphologisch charakterisiert. Die homozygote Deletion von SAMD1 war embryonal letal, während heterozygote SAMD1 KO Mäuse lebend geboren wurden. Am Embryonaltag E14.5 waren bei SAMD1-KO Embryonen eine schwere Schädigung der inneren Organe und die Bildung subkutaner Ödeme sichtbar, höchstwahrscheinlich aufgrund einer gestörten Reifung der Blutgefäße und daraus resultierender Hypoxie. Die beobachteten Defekte führten zu der Schlussfolgerung, dass SAMD1 für die funktionelle Angiogenese und die Entwicklung des Herz-Kreislauf-Systems erforderlich ist. Darüber hinaus deuteten kraniofaziale Defekte bei SAMD1-KO Embryonen auf Fehlfunktionen der Kopf- und Gehirnentwicklung hin. Um weitere Einblicke in die Rolle von SAMD1 während der Neurogenese zu gewinnen, etablierten wir eine Methode zur gezielten neuronalen Differenzierung muriner embryonaler Stammzellen (mESCs). Während dieses Prozesses wurden in SAMD1-KO Zellen transkriptionelle Dysregulationen und der globale Anstieg von H3K4me2 beobachtet, was die funktionelle Rolle von SAMD1 während der Entwicklung neuronaler Zellen weiter unterstützt.

Insgesamt bietet diese Studie Einblicke in die zugrunde liegenden molekularen Mechanismen der KAT6A-Bindung an Chromatin und betont die Ähnlichkeiten von KAT6A-WH1 mit der zuvor

beschriebenen homologen SAMD1-WH Domäne, die ebenfalls direkt mit unmethylierten CGIs interagiert und zur Transkriptionsregulation beiträgt. Sowohl KAT6A als auch SAMD1 sind an einer Vielzahl von Entwicklungsprozessen beteiligt und Fehlregulationen werden mit zellulären Anomalien und der Entstehung von Krebs in Verbindung gebracht.



## 6. REFERENCES

- Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M., & Iyer, L. M. (2005). The many faces of the helix-turn-helix domain: Transcription regulation and beyond. In *FEMS Microbiology Reviews* (Vol. 29, Issue 2, pp. 231–262). Elsevier. <https://doi.org/10.1016/j.femsre.2004.12.008>
- Arboleda, V. A., Lee, H., Dorrani, N., Zadeh, N., Willis, M., Macmurdo, C. F., Manning, M. A., Kwan, A., Hudgins, L., Barthelemy, F., Miceli, M. C., Quintero-Rivera, F., Kantarci, S., Strom, S. P., Deignan, J. L., Grody, W. W., Vilain, E., & Nelson, S. F. (2015). De novo nonsense mutations in KAT6A, a lysine acetyl-transferase gene, cause a syndrome including microcephaly and global developmental delay. *American Journal of Human Genetics*, *96*(3), 498–506. <https://doi.org/10.1016/j.ajhg.2015.01.017>
- Azevedo Portilho, N., Saini, D., Hossain, I., Sirois, J., Moraes, C., & Pastor, W. A. (2021). The DNMT1 inhibitor GSK-3484862 mediates global demethylation in murine embryonic stem cells. *Epigenetics and Chromatin*, *14*(1). <https://doi.org/10.1186/s13072-021-00429-0>
- 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., Cleary, B., Lagiakos, H. R., Mieruszynski, S., Pacini, G., Vanyai, H. K., Bergamasco, M. I., May, R. E., Davey, B. K., Morgan, K. J., ... Thomas, T. (2018). Inhibitors of histone acetyltransferases KAT6A/B induce senescence and arrest tumour growth. *Nature*, *560*(7717), 253–257. <https://doi.org/10.1038/s41586-018-0387-5>
- Becht, D. C., Klein, B. J., Kanai, A., Jang, S. M., Cox, K. L., Zhou, B. R., Phanor, S. K., Zhang, Y., Chen, R. W., Ebmeier, C. C., Lachance, C., Galloy, M., Fradet-Turcotte, A., Bulyk, M. L., Bai, Y., Poirier, M. G., Côté, J., Yokoyama, A., & Kutateladze, T. G. (2023). MORF and MOZ acetyltransferases target unmethylated CpG islands through the winged helix domain. *Nature Communications*, *14*(1), 697. <https://doi.org/10.1038/s41467-023-36368-5>
- Behrends, M., & Engmann, O. (2020). Linker histone H1.5 is an underestimated factor in differentiation and carcinogenesis. *Environmental Epigenetics*, *6*(1). <https://doi.org/10.1093/eep/dvaa013>
- Bell, J. S. K., & Vertino, P. M. (2017). Orphan CpG islands define a novel class of highly active enhancers. *Epigenetics*, *12*(6), 449–464. <https://doi.org/10.1080/15592294.2017.1297910>
- Bhootra, S., Jill, N., Shanmugam, G., Rakshit, S., & Sarkar, K. (2023). DNA methylation and cancer: transcriptional regulation, prognostic, and therapeutic perspective. In *Medical Oncology* (Vol. 40, Issue 2). Springer. <https://doi.org/10.1007/s12032-022-01943-1>
- Bird, A. P. (1986). *CpG-rich islands and the function of DNA methylation*. <https://doi.org/10.1038/321209a0>
- Bogdanović, O., & Veenstra, G. J. C. (2009). DNA methylation and methyl-CpG binding proteins: Developmental requirements and function. In *Chromosoma* (Vol. 118, Issue 5, pp. 549–565). <https://doi.org/10.1007/s00412-009-0221-9>
- Borrow, J., Stanton, V. P., Andresen, J. M., Becher, R., Behm3, F. G., Chaganti4, R. S. K., Civin, C. I., Distcheche, C., Dube7, I., Frischaufs, A. M., Horsman, D., Mitelman, F., Volinia, S., Watmore, A. E., & Housman, D. E. (1996). *The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB-binding protein*. <http://www.nature.com/naturegenetics>

- Brennan, R. (1993). The Winged-Helix DNA-Binding Motif: Another Helix-Turn-Helix Takeoff. In *Cell* (Vol. 74).
- Bryan, E., Warburton, M., Webb, K. M., McLaughlin, K. A., Spanos, C., Ambrosi, C., Major, V., Baubec, T., Rappsilber, J., & Voigt, P. (2021). Nucleosomal Asymmetry Shapes Histone Mark Binding and Promotes Poising at Bivalent Domains. *BioRxiv*, 2021.02.08.430127.  
<https://www.biorxiv.org/content/10.1101/2021.02.08.430127v1%0Ahttps://www.biorxiv.org/content/10.1101/2021.02.08.430127v1.abstract>
- Caiafa, P., & Zampieri, M. (2005). DNA methylation and chromatin structure: The puzzling CpG islands. In *Journal of Cellular Biochemistry* (Vol. 94, Issue 2, pp. 257–265). <https://doi.org/10.1002/jcb.20325>
- Camós, M., Esteve, J., Jares, P., Colomer, D., Rozman, M., Villamor, N., Costa, D., Carrió, A., Nomdedéu, J., Montserrat, E., & Campo, E. (2006). Gene expression profiling of acute myeloid leukemia with translocation t(8;16)(p11;p13) and MYST3-CREBBP rearrangement reveals a distinctive signature with a specific pattern of HOX gene expression. *Cancer Research*, 66(14), 6947–6954.  
<https://doi.org/10.1158/0008-5472.CAN-05-4601>
- Chaffanet, M., Gressin, L., Preudhomme, C., Soenen-Cornu, V., Birnbaum, D., & Pébusque, M. J. (2000). MOZ is fused to p300 in an acute monocytic leukemia with t(8;22). *Genes Chromosomes and Cancer*, 28(2), 138–144. [https://doi.org/10.1002/\(SICI\)1098-2264\(200006\)28:2<138::AID-GCC2>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1098-2264(200006)28:2<138::AID-GCC2>3.0.CO;2-2)
- Champagne, N., Bertos, N. R., Pelletier, N., Wang, A. H., Vezmar, M., Yang, Y., Heng, H. H., & Yang, X. J. (1999). Identification of a human histone acetyltransferase related to monocytic leukemia zinc finger protein. *Journal of Biological Chemistry*, 274(40), 28528–28536. <https://doi.org/10.1074/jbc.274.40.28528>
- Champagne, N., Pelletier, N., & Yang, X. J. (2001). The monocytic leukemia zinc finger protein MOZ is a histone acetyltransferase. *Oncogene*, 20(3), 404–409. <https://doi.org/10.1038/sj.onc.1204114>
- Crump, J. G., Swartz, M. E., Eberhart, J. K., & Kimmel, C. B. (2006). Moz-dependent Hox expression controls segment-specific fate maps of skeletal precursors in the face. *Development*, 133(14), 2661–2669.  
<https://doi.org/10.1242/dev.02435>
- Davey, C., Pennings, S., & Allan, J. (1997). CpG Methylation Remodels Chromatin Structure in vitro. <https://doi.org/10.1006/jmbi.1997.0899>
- Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes and Development*, 25(10), 1010–1022. <https://doi.org/10.1101/gad.2037511>
- Defossez, P. A., & Stancheva, I. (2011). Biological functions of methyl-CpG-binding proteins. In *Progress in Molecular Biology and Translational Science* (Vol. 101, pp. 377–398). Elsevier B.V.  
<https://doi.org/10.1016/B978-0-12-387685-0.00012-3>
- Deguchi, K., Ayton, P. M., Carapeti, M., Kutok, J. L., Snyder, C. S., Williams, I. R., Cross, N. C. P., Glass, C. K., Cleary, M. L., & Gilliland, D. G. (2003). MOZ-TIF2-induced acute myeloid leukemia requires the MOZ nucleosome binding motif and TIF2-mediated recruitment of CBP. *Cancer Cell*, 3(3), 259–271.  
[https://doi.org/10.1016/S1535-6108\(03\)00051-5](https://doi.org/10.1016/S1535-6108(03)00051-5)
- Dreveny, I., Deeves, S. E., Fulton, J., Yue, B., Messmer, M., Bhattacharya, A., Collins, H. M., & 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 Research*, 42(2), 822–835. <https://doi.org/10.1093/nar/gkt931>
- 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., Koseki, H., Brockdorff, N., Ponting, C. P., Kessler, B. M., & Klose, R. J. (2012). KDM2B links the polycomb repressive complex 1 (PRC1) to recognition of CpG islands. *ELife*, 2012(1). <https://doi.org/10.7554/eLife.00205>
- Fischer, S., Weber, L. M., & Liefke, R. (2022). Evolutionary adaptation of the Polycomb repressive complex 2. In *Epigenetics and Chromatin* (Vol. 15, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s13072-022-00439-6>
- Gajiwala, K. S., & Burley, S. K. (2000). Winged helix proteins. *Curr Opin Struct Biol.*, 10(1), 110–116.
- Harami, G. M., Gyimesi, M., & Kovács, M. (2013). From keys to bulldozers: Expanding roles for winged helix domains in nucleic-acid-binding proteins. In *Trends in Biochemical Sciences* (Vol. 38, Issue 7, pp. 364–371). <https://doi.org/10.1016/j.tibs.2013.04.006>
- Hergeth, S. P., & Schneider, R. (2015). The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO Reports*, 16(11), 1439–1453. <https://doi.org/10.15252/embr.201540749>
- Holbert, M. A., Sikorski, T., Carten, J., Snowflack, D., Hodawadekar, S., & Marmorstein, R. (2007). The human monocytic leukemia zinc finger histone acetyltransferase domain contains DNA-binding activity implicated in chromatin targeting. *Journal of Biological Chemistry*, 282(50), 36603–36613. <https://doi.org/10.1074/jbc.M705812200>
- Hu, S., Wan, J., Su, Y., Song, Q., Zeng, Y., Nguyen, H. N., Shin, J., Cox, E., Rho, H. S., Woodard, C., Xia, S., Liu, S., Lyu, H., Ming, G. L., Wade, H., Song, H., Qian, J., & Zhu, H. (2013). DNA methylation presents distinct binding sites for human transcription factors. *ELife*, 2013(2). <https://doi.org/10.7554/eLife.00726>
- Hung, T., Binda, O., Champagne, K. S., Kuo, A. J., Johnson, K., Chang, H. Y., Simon, M. D., Kutateladze, T. G., & Gozani, O. (2009). ING4 Mediates Crosstalk between Histone H3 K4 Trimethylation and H3 Acetylation to Attenuate Cellular Transformation. *Molecular Cell*, 33(2), 248–256. <https://doi.org/10.1016/j.molcel.2008.12.016>
- Jeziorska, D. M., Murray, R. J. S., De Gobbi, M., Gaentzsch, R., Garrick, D., Ayyub, H., Chen, T., Li, E., Telenius, J., Lynch, M., Graham, B., Smith, A. J. H., Lund, J. N., Hughes, J. R., Higgs, D. R., & Tufarelli, C. (2017). DNA methylation of intragenic CpG islands depends on their transcriptional activity during differentiation and disease. *Proceedings of the National Academy of Sciences of the United States of America*, 114(36), E7526–E7535. <https://doi.org/10.1073/pnas.1703087114>
- Kang, H., Jung, Y. L., McElroy, K. A., Zee, B. M., Wallace, H. A., Woolnough, J. L., Park, P. J., & Kuroda, M. I. (2017). Bivalent complexes of PRC1 with orthologs of BRD4 and MOZ/MORF target developmental genes in *Drosophila*. *Genes and Development*, 31(19), 1988–2002. <https://doi.org/10.1101/gad.305987.117>
- KAT6 Foundation. (2023, May 20). *What is KAT6A Syndrome?* <https://kat6a.org/about-kat6a/>
- Katsumoto, T., Aikawa, Y., Iwama, A., Ueda, S., Ichikawa, H., Ochiya, T., & Kitabayashi, I. (2006). MOZ is essential for maintenance of hematopoietic stem cells. *Genes and Development*, 20(10), 1321–1330. <https://doi.org/10.1101/gad.1393106>

- Katsumoto, T., Yoshida, N., & Kitabayashi, I. (2008). Roles of the histone acetyltransferase monocytic leukemia zinc finger protein in normal and malignant hematopoiesis. In *Cancer Science* (Vol. 99, Issue 8, pp. 1523–1527). <https://doi.org/10.1111/j.1349-7006.2008.00865.x>
- Kennedy, J., Goudie, D., Blair, E., Chandler, K., Joss, S., McKay, V., Green, A., Armstrong, R., Lees, M., Kamien, B., Hopper, B., Tan, T. Y., Yap, P., Stark, Z., Okamoto, N., Miyake, N., Matsumoto, N., Macnamara, E., Murphy, J. L., ... Newbury-Ecob, R. (2019). KAT6A Syndrome: genotype–phenotype correlation in 76 patients with pathogenic KAT6A variants. *Genetics in Medicine*, *21*(4), 850–860. <https://doi.org/10.1038/s41436-018-0259-2>
- Kitabayashi, I., Aikawa, Y., Nguyen, L. A., Yokoyama, A., & Ohki, M. (2001). Activation of AML1-mediated transcription by MOZ and inhibition by the MOZ-CBP fusion protein. *EMBO Journal*, *20*(24), 7184–7196. <https://doi.org/10.1093/emboj/20.24.7184>
- Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., Presser, A., Nusbaum, C., Xie, X., Chi, A. S., Adli, M., Kasif, S., Ptaszek, L. M., Cowan, C. A., Lander, E. S., Koseki, H., & Bernstein, B. E. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genetics*, *4*(10). <https://doi.org/10.1371/journal.pgen.1000242>
- Laue, K., Daujat, S., Crump, J. G., Plaster, N., Roehl, H. H., van Bebber, F., Busch-Nentwich, E., Dahm, R., Frohnhöfer, H. G., Geiger, H., Gilmour, D., Holley, S., Hooge, J., Jülich, D., Knaut, H., Maderspacher, F., Maischein, H. M., Neumann, C., Nicolson, T., ... Hammerschmidt, M. (2008). The multidomain protein Brpf1 binds histones and is required for hox gene expression and segmental identity. *Development*, *135*(11), 1935–1946. <https://doi.org/10.1242/dev.017160>
- Lee, J. H., Voo, K. S., & Skalnik, D. G. (2001). Identification and Characterization of the DNA Binding Domain of CpG-binding Protein. *Journal of Biological Chemistry*, *276*(48), 44669–44676. <https://doi.org/10.1074/jbc.M107179200>
- Li, H., Liefke, R., Jiang, J., Kurland, J. V., Tian, W., Deng, P., Zhang, W., He, Q., Patel, D. J., Bulyk, M. L., Shi, Y., & Wang, Z. (2017). Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature*, *549*(7671), 287–291. <https://doi.org/10.1038/nature23881>
- Li, J., Jiang, L., Liang, X., Qu, L., Wu, D., Chen, X., Guo, M., Chen, Z., Chen, L., & Chen, Y. (2017). DNA-binding properties of FOXP3 transcription factor. *Acta Biochimica et Biophysica Sinica*, *49*(9), 792–799. <https://doi.org/10.1093/abbs/gmx079>
- Lv, D., Jia, F., Hou, Y., Sang, Y., Alvarez, A. A., Zhang, W., Gao, W. Q., Hu, B., Cheng, S. Y., Ge, J., Li, Y., & Feng, H. (2017). Histone acetyltransferase KAT6A upregulates PI3K/AKT signaling through TRIM24 binding. *Cancer Research*, *77*(22), 6190–6201. <https://doi.org/10.1158/0008-5472.CAN-17-1388>
- Marx, A., Kahan, T., & Simon, I. (2013). Integrative analysis of methylome and transcriptome reveals the importance of unmethylated CpGs in non-CpG island gene activation. *BioMed Research International*, *2013*. <https://doi.org/10.1155/2013/785731>
- Millan, F., Cho, M. T., Retterer, K., Monaghan, K. G., Bai, R., Vitazka, P., Everman, D. B., Smith, B., Angle, B., Roberts, V., Immken, L. D., Nagakura, H., DiFazio, M., Sherr, E., Haverfield, E., Friedman, B., Telegrafi, A., Juusola, J., Chung, W. K., & Bale, S. (2016). Whole exome sequencing reveals de novo pathogenic variants

- in KAT6A as a cause of a neurodevelopmental disorder. *American Journal of Medical Genetics, Part A*, 170(7), 1791–1798. <https://doi.org/10.1002/ajmg.a.37670>
- Miller, C. T., Maves, L., & Kimmel, C. B. (2004). *moz* regulates Hox expression and pharyngeal segmental identity in zebrafish. *Development*, 131(10), 2443–2461. <https://doi.org/10.1242/dev.01134>
- Miyamoto, R., Okuda, H., Kanai, A., Takahashi, S., Kawamura, T., Matsui, H., Kitamura, T., Kitabayashi, I., Inaba, T., & Yokoyama, A. (2020). Activation of CpG-Rich Promoters Mediated by MLL Drives MOZ-Rearranged Leukemia. *Cell Reports*, 32(13). <https://doi.org/10.1016/j.celrep.2020.108200>
- Miyamoto, R., & Yokoyama, A. (2021). Protocol for fractionation-assisted native ChIP (fanChIP) to capture protein-protein/DNA interactions on chromatin. *STAR Protocols*, 2(2). <https://doi.org/10.1016/j.xpro.2021.100404>
- Murati, A., Gervais, C., Carbuccia, N., Finetti, P., Cervera, N., Adélaïde, J., Struski, S., Lippert, E., Mugneret, F., Tigaud, I., Penther, D., Bastard, C., Poppe, B., Speleman, F., Baranger, L., Luquet, I., Cornillet-Lefebvre, P., Nadal, N., Nguyen-Khac, F., ... Birnbaum, D. (2009). Genome profiling of acute myelomonocytic leukemia: Alteration of the MYB locus in MYST3-linked cases. *Leukemia*, 23(1), 85–94. <https://doi.org/10.1038/leu.2008.257>
- Owen, B. M., & Davidovich, C. (2022). DNA binding by polycomb-group proteins: searching for the link to CpG islands. *Nucleic Acids Research*, 50(9), 4813–4839. <https://doi.org/10.1093/nar/gkac290>
- Pachano, T., Sánchez-Gaya, V., Ealo, T., Mariner-Faulí, M., Bleckwehl, T., Asenjo, H. G., Respuela, P., Cruz-Molina, S., Muñoz-San Martín, M., Haro, E., van IJcken, W. F. J., Landeira, D., & Rada-Iglesias, A. (2021). Orphan CpG islands amplify poised enhancer regulatory activity and determine target gene responsiveness. *Nature Genetics*, 53(7), 1036–1049. <https://doi.org/10.1038/s41588-021-00888-x>
- Pasini, D., Bracken, A. P., Hansen, J. B., Capillo, M., & Helin, K. (2007). The Polycomb Group Protein Suz12 Is Required for Embryonic Stem Cell Differentiation. *Molecular and Cellular Biology*, 27(10), 3769–3779. <https://doi.org/10.1128/mcb.01432-06>
- Pelletier, N., Champagne, N., Stifani, S., & Yang, X.-J. (2002). MOZ and MORF histone acetyltransferases interact with the Runt-domain transcription factor Runx2. *Oncogene*, 21, 2729–2740. <https://doi.org/10.1038/sj/onc/1205367>
- 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., Rendleman, E. J., Savas, J. N., Singer, B. D., Bulun, S. E., & Shilatifard, A. (2019). CATAcomb: An endogenous inducible gene that antagonizes H3K27 methylation activity of Polycomb repressive complex 2 via an H3K27M-like mechanism. In *Sci. Adv* (Vol. 5). <https://www.science.org>
- Qiu, Y., Liu, L., Zhao, C., Han, C., Li, F., Zhang, J., Wang, Y., Li, G., Mei, Y., Wu, M., Wu, J., & Shi, Y. (2012). Combinatorial readout of unmodified H3R2 and acetylated H3K14 by the tandem PHD finger of MOZ reveals a regulatory mechanism for HOXA9 transcription. *Genes and Development*, 26(12), 1376–1391. <https://doi.org/10.1101/gad.188359.112>
- Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, 8(11), 2281–2308. <https://doi.org/10.1038/nprot.2013.143>

- Rokudai, S., Laptenko, O., Arnal, S. M., Taya, Y., Kitabayashi, I., & Prives, C. (2013). MOZ increases p53 acetylation and premature senescence through its complex formation with PML. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(10), 3895–3900. <https://doi.org/10.1073/pnas.1300490110>
- Saha, A., & Dalal, Y. (2021). A glitch in the snitch: The role of linker histone H1 in shaping the epigenome in normal and diseased cells. *Open Biology*, *11*(8). <https://doi.org/10.1098/rsob.210124>
- Sheikh, B. N., Downer, N. L., Phipson, B., Vanyai, H. K., Kueh, A. J., McCarthy, D. J., Smyth, G. K., Thomas, T., & 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. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(17), 5437–5442. <https://doi.org/10.1073/pnas.1422872112>
- Sheikh, B. N., Yang, Y., Schreuder, J., Nilsson, S. K., Bilardi, R., Carotta, S., McRae, H. M., Metcalf, D., Voss, A. K., & Thomas, T. (2016). MOZ (KAT6A) is essential for the maintenance of classically defined adult hematopoietic stem cells. *Blood*, *128*(19), 2307–2318. <https://doi.org/10.1182/blood-2015-10-676072>
- Shen, L., Kondo, Y., Guo, Y., Zhang, J., Zhang, L., Ahmed, S., Shu, J., Chen, X., Waterland, R. A., & Issa, J. P. J. (2007). Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS Genetics*, *3*(10), 2023–2026. <https://doi.org/10.1371/journal.pgen.0030181>
- Shin Voo, K., Carlone, D. L., Jacobsen, B. M., Flodin, A., & Skalnik, D. G. (2000). Cloning of a Mammalian Transcriptional Activator That Binds Unmethylated CpG Motifs and Shares a CXXC Domain with DNA Methyltransferase, Human Trithorax, and Methyl-CpG Binding Domain Protein 1. In *MOLECULAR AND CELLULAR BIOLOGY* (Vol. 20, Issue 6). <https://journals.asm.org/journal/mcb>
- Stielow, B., Finkernagel, F., Stiewe, T., Nist, A., & Suske, G. (2018). MGA, L3MBTL2 and E2F6 determine genomic binding of the non-canonical Polycomb repressive complex PRC1.6. *PLoS Genetics*, *14*(1), 1–29. <https://doi.org/10.1371/journal.pgen.1007193>
- Stielow, B., Zhou, Y., Cao, Y., Simon, C., Pogoda, H. M., Jiang, J., Ren, Y., Phanor, S. K., Rohner, I., Nist, A., Stiewe, T., Hammerschmidt, M., Shi, Y., Bulyk, M. L., Wang, Z., & Liefke, R. (2021). The SAM domain-containing protein 1 (SAMD1) acts as a repressive chromatin regulator at unmethylated CpG islands. *Science Advances*, *7*(20), 1–18. <https://doi.org/10.1126/sciadv.abf2229>
- Tham, E., Lindstrand, A., Santani, A., Malmgren, H., Nesbitt, A., Dubbs, H. A., Zackai, E. H., Parker, M. J., Millan, F., Rosenbaum, K., Wilson, G. N., & Nordgren, A. (2015). Dominant mutations in KAT6A cause intellectual disability with recognizable syndromic features. *American Journal of Human Genetics*, *96*(3), 507–513. <https://doi.org/10.1016/j.ajhg.2015.01.016>
- Thomas, T., Corcoran, L. M., Gugasyan, R., Dixon, M. P., Brodnicki, T., Nutt, S. L., Metcalf, D., & Voss, A. K. (2006). Monocytic leukemia zinc finger protein is essential for the development of long-term reconstituting hematopoietic stem cells. *Genes and Development*, *20*(9), 1175–1186. <https://doi.org/10.1101/gad.1382606>
- Thomson, J. P., Skene, P. J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A. R. W., Deaton, A., Andrews, R., James, K. D., Turner, D. J., Illingworth, R., & Bird, A. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature*, *464*(7291), 1082–1086. <https://doi.org/10.1038/nature08924>

- Trinh, J., Hüning, I., Yüksel, Z., Baalman, N., Imhoff, S., Klein, C., Rolfs, A., Gillissen-Kaesbach, G., & Lohmann, K. (2018). A KAT6A variant in a family with autosomal dominantly inherited microcephaly and developmental delay. *Journal of Human Genetics*, *63*(9), 997–1001. <https://doi.org/10.1038/s10038-018-0469-0>
- Ullah, M., Pelletier, N., Xiao, L., Zhao, S. P., Wang, K., Degerny, C., Tahmasebi, S., Cayrou, C., Doyon, Y., Goh, S.-L., Champagne, N., Côté, J., & Yang, X.-J. (2008). Molecular Architecture of Quartet MOZ/MORF Histone Acetyltransferase Complexes. *Molecular and Cellular Biology*, *28*(22), 6828–6843. <https://doi.org/10.1128/mcb.01297-08>
- Vanyai, H. K., Garnham, A., May, R. E., McRae, H. M., Collin, C., Wilcox, S., Smyth, G. K., Thomas, T., & Voss, A. K. (2019). MOZ directs the distal-less homeobox gene expression program during craniofacial development. *Development (Cambridge)*, *146*(14). <https://doi.org/10.1242/dev.175042>
- Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Zidek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., ... Velankar, S. (2022). AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, *50*(D1), D439–D444. <https://doi.org/10.1093/nar/gkab1061>
- Varriale, A., & Bernardi, G. (2010). Distribution of DNA methylation, CpGs, and CpG islands in human isochores. *Genomics*, *95*(1), 25–28. <https://doi.org/10.1016/j.ygeno.2009.09.006>
- Völkel, S., Stielow, B., Finkernagel, F., Stiewe, T., Nist, A., & 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 Genetics*, *11*(3). <https://doi.org/10.1371/journal.pgen.1005102>
- Voss, A. K., Collin, C., Dixon, M. P., & Thomas, T. (2009). Moz and Retinoic Acid Coordinately Regulate H3K9 Acetylation, Hox Gene Expression, and Segment Identity. *Developmental Cell*, *17*(5), 674–686. <https://doi.org/10.1016/j.devcel.2009.10.006>
- Voss, A. K., Vanyai, H. K., Collin, C., Dixon, M. P., McLennan, T. J., Sheikh, B. N., Scambler, P., & Thomas, T. (2012). MOZ Regulates the Tbx1 Locus, and Moz Mutation Partially Phenocopies DiGeorge Syndrome. *Developmental Cell*, *23*(3), 652–663. <https://doi.org/10.1016/j.devcel.2012.07.010>
- Weber, L. M., Jia, Y., Stielow, B., Gisselbrecht, S. S., Cao, Y., Ren, Y., Rohner, I., King, J., Rothman, E., Fischer, S., Simon, C., Forné, I., Nist, A., Stiewe, T., Bulyk, M. L., Wang, Z., & Liefke, R. (2023). The histone acetyltransferase KAT6A is recruited to unmethylated CpG islands via a DNA binding winged helix domain. *Nucleic Acids Research*, *51*(2), 574–594. <https://doi.org/10.1093/nar/gkac1188>
- Wen, H., Li, Y., Li, H., & Shi, X. (2014). ZMYND11: An H3.3-specific reader of H3K36me3. In *Cell Cycle* (Vol. 13, Issue 14, pp. 2153–2154). Taylor and Francis Inc. <https://doi.org/10.4161/cc.29732>
- Widom, J. (1998). Chromatin structure: Linking structure to function with histone H1. *Current Biology*, *8*, 788–791. <http://biomednet.com/elecref/09609822008R0788>
- Wiesel-Motiuk, N., & Assaraf, Y. G. (2020). The key roles of the lysine acetyltransferases KAT6A and KAT6B in physiology and pathology. *Drug Resistance Updates*, *53*, 100729. <https://doi.org/https://doi.org/10.1016/j.drug.2020.100729>

- Xiong, X., Panchenko, T., Yang, S., Zhao, S., Yan, P., Zhang, W., Xie, W., Li, Y., Zhao, Y., Allis, C. D., & Li, H. (2016). Selective recognition of histone crotonylation by double PHD fingers of MOZ and DPF2. *Nature Chemical Biology*, *12*(12), 1111–1118. <https://doi.org/10.1038/nchembio.2218>
- Xiong, X., Tu, S., Wang, J., Luo, S., & Yan, X. (2019). CXXC5: A novel regulator and coordinator of TGF- $\beta$ , BMP and Wnt signaling. In *Journal of Cellular and Molecular Medicine* (Vol. 23, Issue 2, pp. 740–749). Blackwell Publishing Inc. <https://doi.org/10.1111/jcmm.14046>
- Yan, F., Li, J., Milosevic, J., Petroni, R., Liu, S., Shi, Z., Yuan, S., Reynaga, J. M., Qi, Y., Rico, J., Yu, S., Liu, Y., Rokudai, S., Palmisiano, N., Meyer, S. E., Sung, P. J., Wan, L., Lan, F., Garcia, B. A., ... Andrés Blanco, M. (2022). KAT6A and ENL Form an Epigenetic Transcriptional Control Module to Drive Critical Leukemogenic Gene-Expression Programs. *Cancer Discovery*, *12*(3), 792–811. <https://doi.org/10.1158/2159-8290.CD-20-1459>
- 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., Potocki, L., Zhou, D., Cadieux-Dion, M., Bellus, G. A., Wagner, M. D., Hale, R. J., Esber, N., Riley, A. F., Solomon, B. D., ... Yang, X. J. (2020). Deficient histone H3 propionylation by BRPF1-KAT6 complexes in neurodevelopmental disorders and cancer. *Science Advances*, *6*(4). <https://doi.org/10.1126/sciadv.aax0021>
- Yang, X. J. (2015). MOZ and MORF acetyltransferases: Molecular interaction, animal development and human disease. In *Biochimica et Biophysica Acta - Molecular Cell Research* (Vol. 1853, Issue 8, pp. 1818–1826). Elsevier. <https://doi.org/10.1016/j.bbamcr.2015.04.014>
- Yang, X. J., & Ullah, M. (2007). MOZ and MORF, two large MYSTic HATs in normal and cancer stem cells. In *Oncogene* (Vol. 26, Issue 37, pp. 5408–5419). <https://doi.org/10.1038/sj.onc.1210609>
- Yu, B., Luo, F., Sun, B., Liu, W., Shi, Q., Cheng, S. Y., Chen, C., Chen, G., Li, Y., & Feng, H. (2021). KAT6A Acetylation of SMAD3 Regulates Myeloid-Derived Suppressor Cell Recruitment, Metastasis, and Immunotherapy in Triple-Negative Breast Cancer. *Advanced Science*, *8*(20). <https://doi.org/10.1002/advs.202100014>
- Yu, L., Liang, Y., Cao, X., Wang, X., Gao, H., Lin, S. Y., Schiff, R., Wang, X. S., & Li, K. (2017). Identification of MYST3 as a novel epigenetic activator of ER $\alpha$  frequently amplified in breast cancer. *Oncogene*, *36*(20), 2910–2918. <https://doi.org/10.1038/onc.2016.433>
- Zhou, C., Liu, W., & Duan, Y. (2020). MOZ/KAT6A: a promising target for acute myeloid leukemia therapy. In *Future Medicinal Chemistry* (Vol. 12, Issue 9, pp. 759–761). Future Medicine Ltd. <https://doi.org/10.4155/fmc-2020-0047>
- Zu, G., Liu, Y., Cao, J., Zhao, B., Zhang, H., & You, L. (2022). BRPF1-KAT6A/KAT6B Complex: Molecular Structure, Biological Function and Human Disease. In *Cancers* (Vol. 14, Issue 17). MDPI. <https://doi.org/10.3390/cancers14174068>
- Zwaveling-Soonawala, N., Maas, S. M., Alders, M., Majoie, C. B., Fliers, E., van Trotsenburg, A. S. P., & Hennekam, R. C. M. (2017). Variants in KAT6A and pituitary anomalies. In *American Journal of Medical Genetics, Part A* (Vol. 173, Issue 9, pp. 2562–2565). Wiley-Liss Inc. <https://doi.org/10.1002/ajmg.a.38330>



## APPENDIX

### I. Publications

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

*“Investigation of SAMD1 ablation in mice”*

### II. Curriculum vitae

### III. PhD Portfolio

### IV. List of academic teachers

### V. Danksagung

### VI. Ehrenwörtliche Erklärung

## *I. Publications*

The following publications, which are discussed in this thesis, are attached:

**Weber, L. M.**, Jia, Y., Stielow, B., Gisselbrecht, S. S., Cao, Y., Ren, Y., Rohner, I., King, J., Rothman, E., Fischer, S., Simon, C., Forné, I., Nist, A., Stiewe, T., Bulyk, M. L., Wang, Z., & Liefke, R. (2023). The histone acetyltransferase KAT6A is recruited to unmethylated CpG islands via a DNA binding winged helix domain. *Nucleic Acids Research*, 51(2), 574–594. <https://doi.org/10.1093/nar/gkac1188>

Supplementary Table S3 (PBM data) and Supplementary Table S4 (MS data) are excluded from this document due to their size. Find them under the following link:

<https://academic.oup.com/nar/article/51/2/574/6946999#supplementary-data>

Campbell, B., **Weber, L. M.**, Engle, S. J., Ozolinš, T. R. S., Bourassa, P., Aiello, R., & Liefke, R. (2023). Investigation of SAMD1 ablation in mice. *Scientific Reports*, 13(1), 3000. <https://doi.org/10.1038/s41598-023-29779-3>

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

Lisa Marie Weber<sup>1,†</sup>, Yulin Jia<sup>2,†</sup>, Bastian Stielow<sup>1</sup>, Stephen S. Gisselbrecht<sup>3</sup>, Yinghua Cao<sup>2</sup>, Yanpeng Ren<sup>2</sup>, Iris Rohner<sup>1</sup>, Jessica King<sup>3</sup>, Elisabeth Rothman<sup>3</sup>, Sabrina Fischer<sup>1</sup>, Clara Simon<sup>1</sup>, Ignasi Forné<sup>4</sup>, Andrea Nist<sup>5</sup>, Thorsten Stiewe<sup>5</sup>, Martha L. Bulyk<sup>3,6</sup>, Zhanxin Wang<sup>2,\*</sup> and Robert Liefke<sup>1,7,\*</sup>

<sup>1</sup>Institute of Molecular Biology and Tumor Research (IMT), Philipps University of Marburg, Marburg 35043, Germany, <sup>2</sup>Key Laboratory of Cell Proliferation and Regulation Biology of Ministry of Education, College of Life Sciences, Beijing Normal University, Beijing 100875, China, <sup>3</sup>Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA, <sup>4</sup>Protein Analysis Unit, Biomedical Center (BMC), Faculty of Medicine, Ludwig-Maximilians-University (LMU) Munich, Martinsried 82152, Germany, <sup>5</sup>Genomics Core Facility, Institute of Molecular Oncology, Member of the German Center for Lung Research (DZL), Philipps University of Marburg, Marburg 35043, Germany, <sup>6</sup>Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA and <sup>7</sup>Department 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 $\gamma$  (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 OD<sub>600</sub> 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 homemade 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 MgCl<sub>2</sub>, 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 MgCl<sub>2</sub>, 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  $1 \times 10^8$  cells with NP-40 lysis buffer (20 mM Tris-HCl, pH 7.6, 300 mM KCl, 12.5 mM MgCl<sub>2</sub>, 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 Laemml 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 NH<sub>4</sub>HCO<sub>3</sub> 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  $3 \times 10^6$ ). The 10 most intense peptide ions with charge states between 2 and 5 were sequentially isolated to a target value of  $1 \times 10^5$ , 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  $1 \times 10^5$  cells per well. The cell viability was determined 1, 3 and 6 days after seeding using MTT assay by adding 90  $\mu$ 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  $\mu$ 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 ( $1 \times 10^3$  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 dH<sub>2</sub>O 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  $\mu$ M and 10  $\mu$ 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 NaHCO<sub>3</sub>, 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

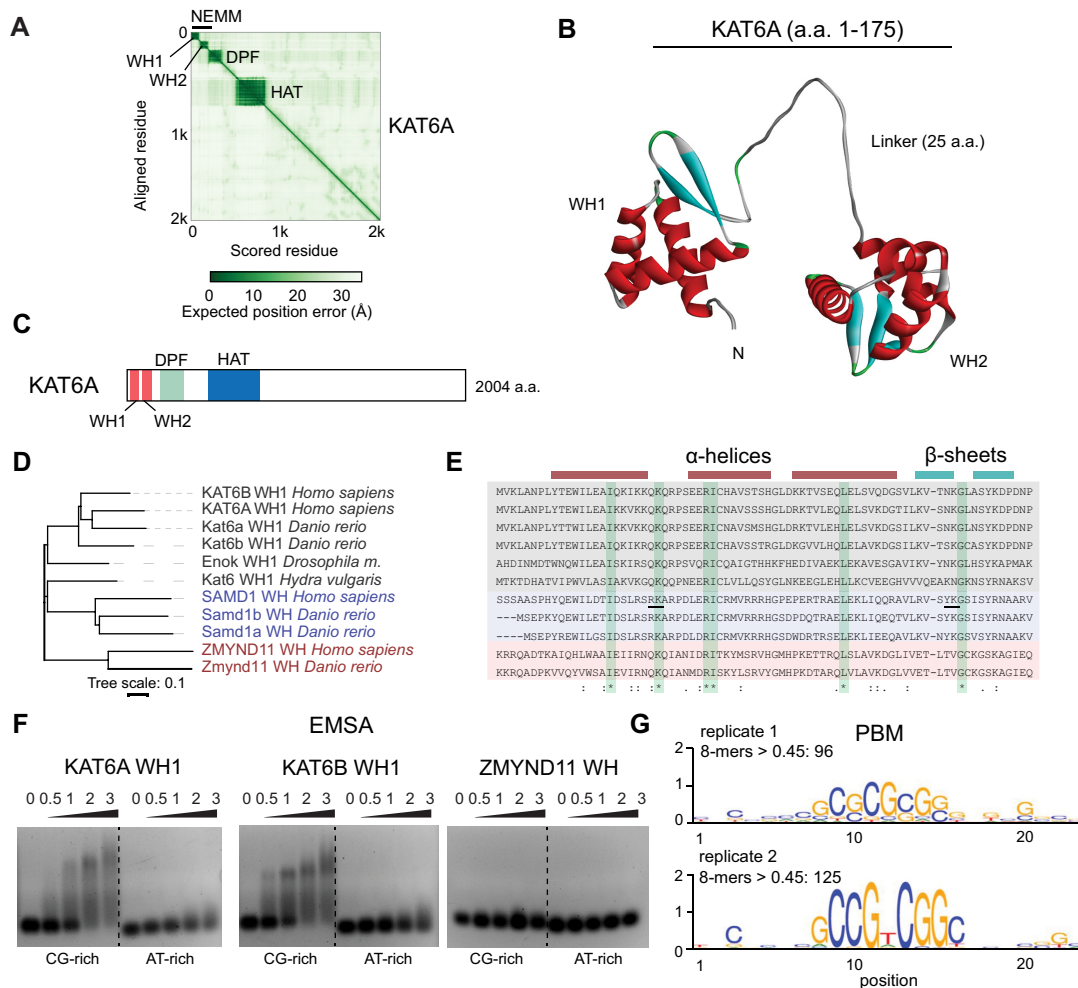
### KAT6A possesses a conserved N-terminal winged helix domain

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-





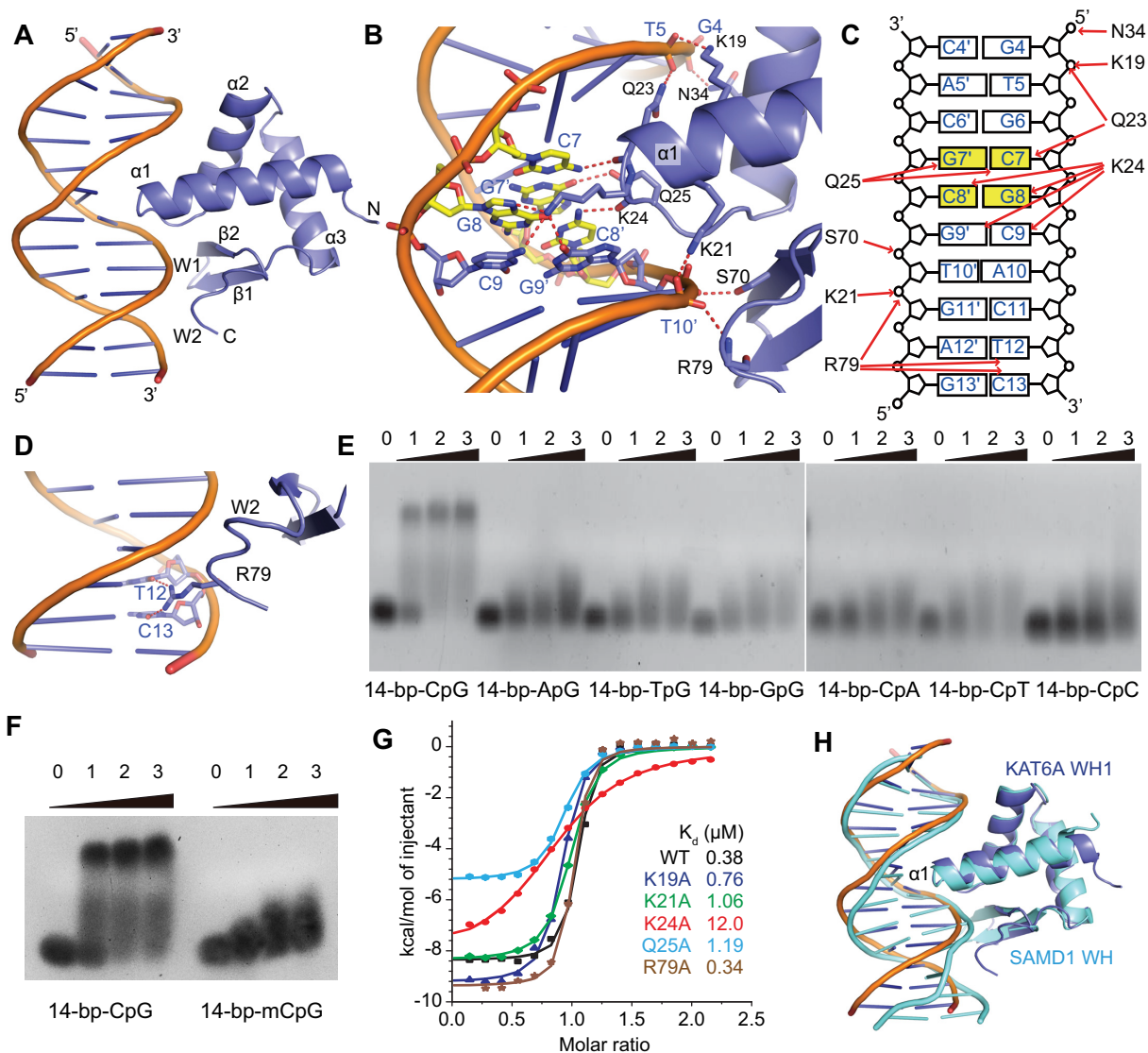
**Figure 1.** (A) AlphaFold predicted alignment error for KAT6A (65), showing two distinguishable WH domains at the N-terminus, reflecting the NEMM, DPF = double PHD finger. See also Supplementary Figure S1. (B) AlphaFold predicted structure of WH1 and WH2 of KAT6A. (C) Domain structure of KAT6A based 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  $\alpha$  helices at the N-terminal half and two  $\beta$  strands at the C-terminal half (Figure 2A). Wing-like loops between strands  $\beta_1$  and  $\beta_2$  (W1) and after  $\beta_2$  (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  $\alpha_1$  helix. The C-terminal end of the  $\alpha_1$  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 crystalized 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

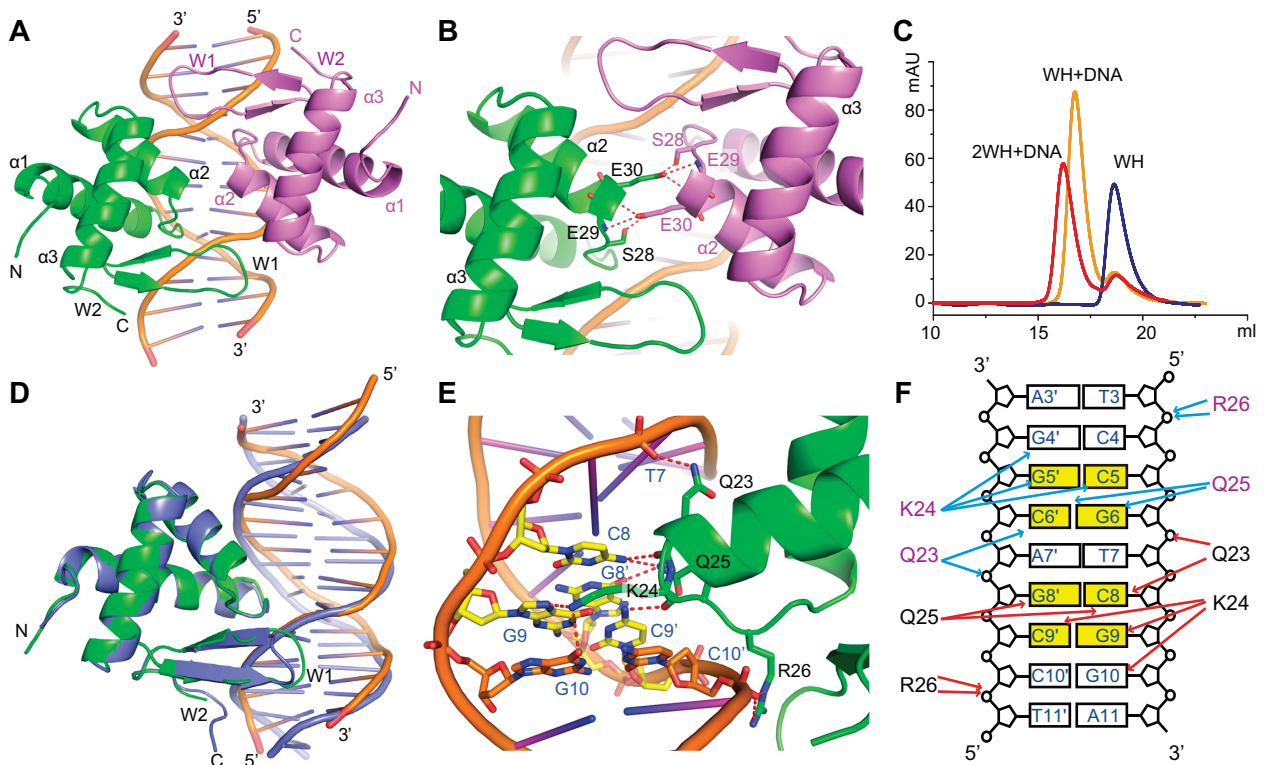
verifying that the CGG/CGC-motifs are the most critical sequences required for the recognition of KAT6A WH1.

### 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*, *Skid1* 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.

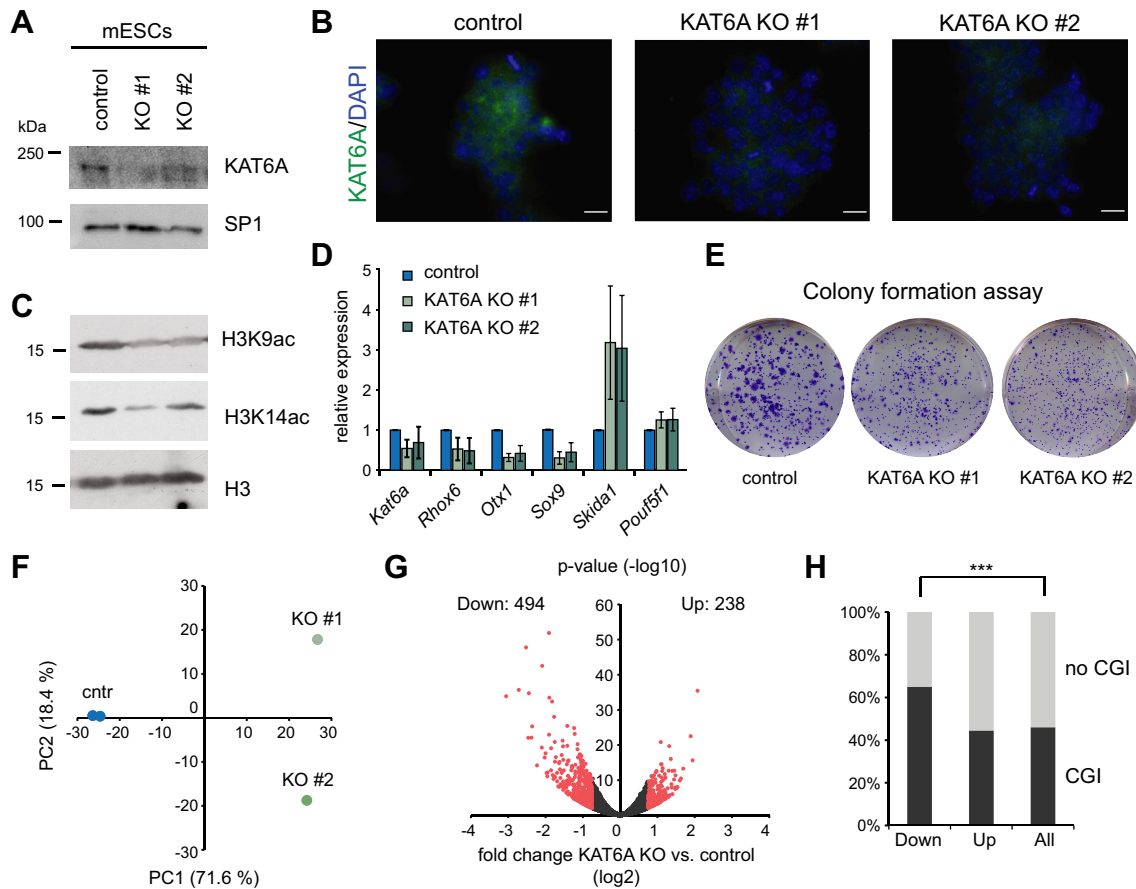
nificantly enriched for CGIs at their promoters (Figure 4H). This result supports that KAT6A may bind to CGI-containing promoters in mouse ES cells for gene activation, and that the WH1 domain could play a role in the chromatin binding of KAT6A in these cells. However, in mouse ES cells endogenous KAT6A 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  $\mu$ m. (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. \*\*\*  $P < 0.001$ .

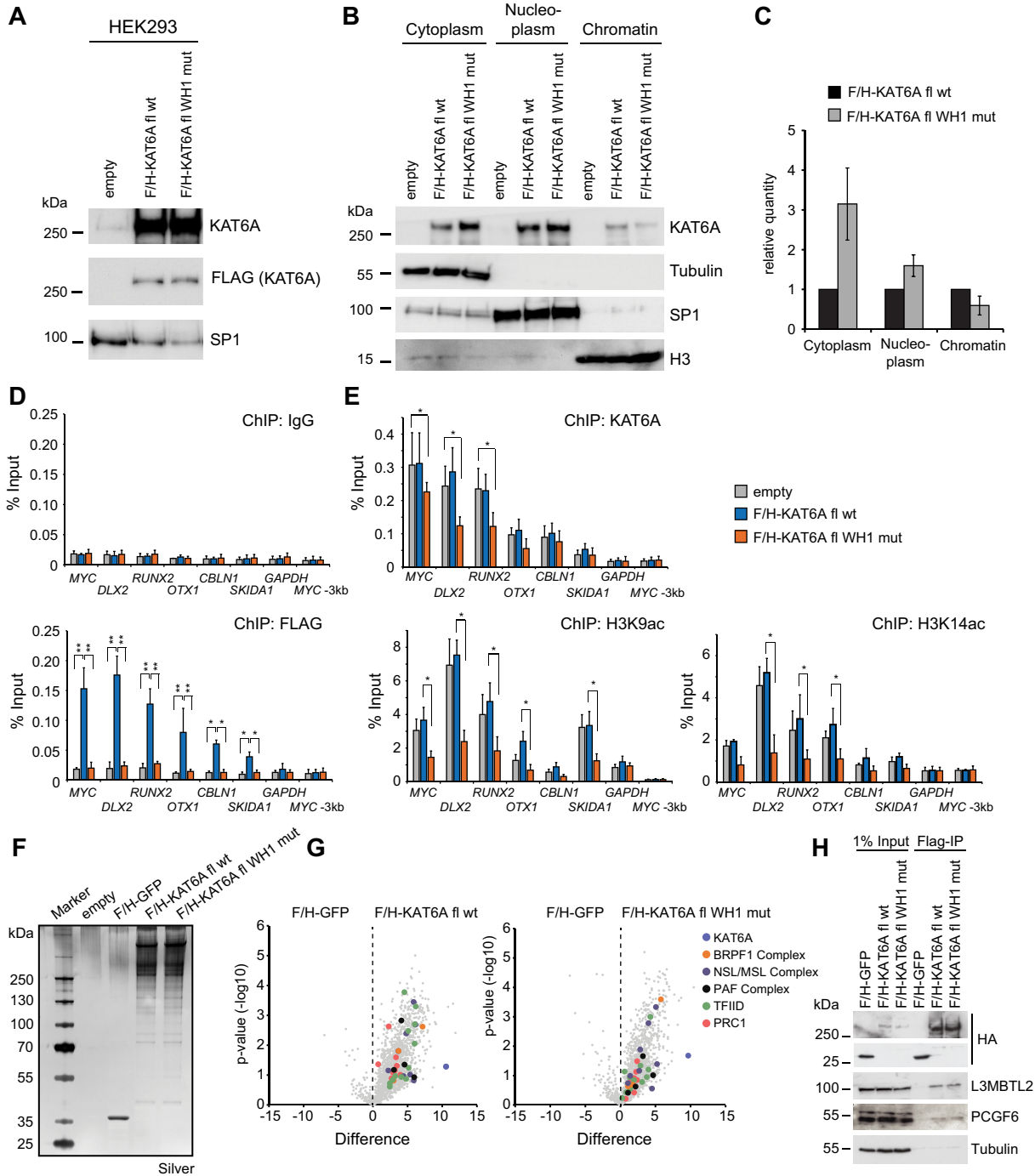
line with the relatively low DNA binding affinity of the WH1 domain (Figure 2G) and is similar to the WH domain of SAMD1, which is also not able to bind to chromatin in isolation (36). To address, which additional regions of KAT6A are required for chromatin binding, we investigated KAT6A proteins that 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.

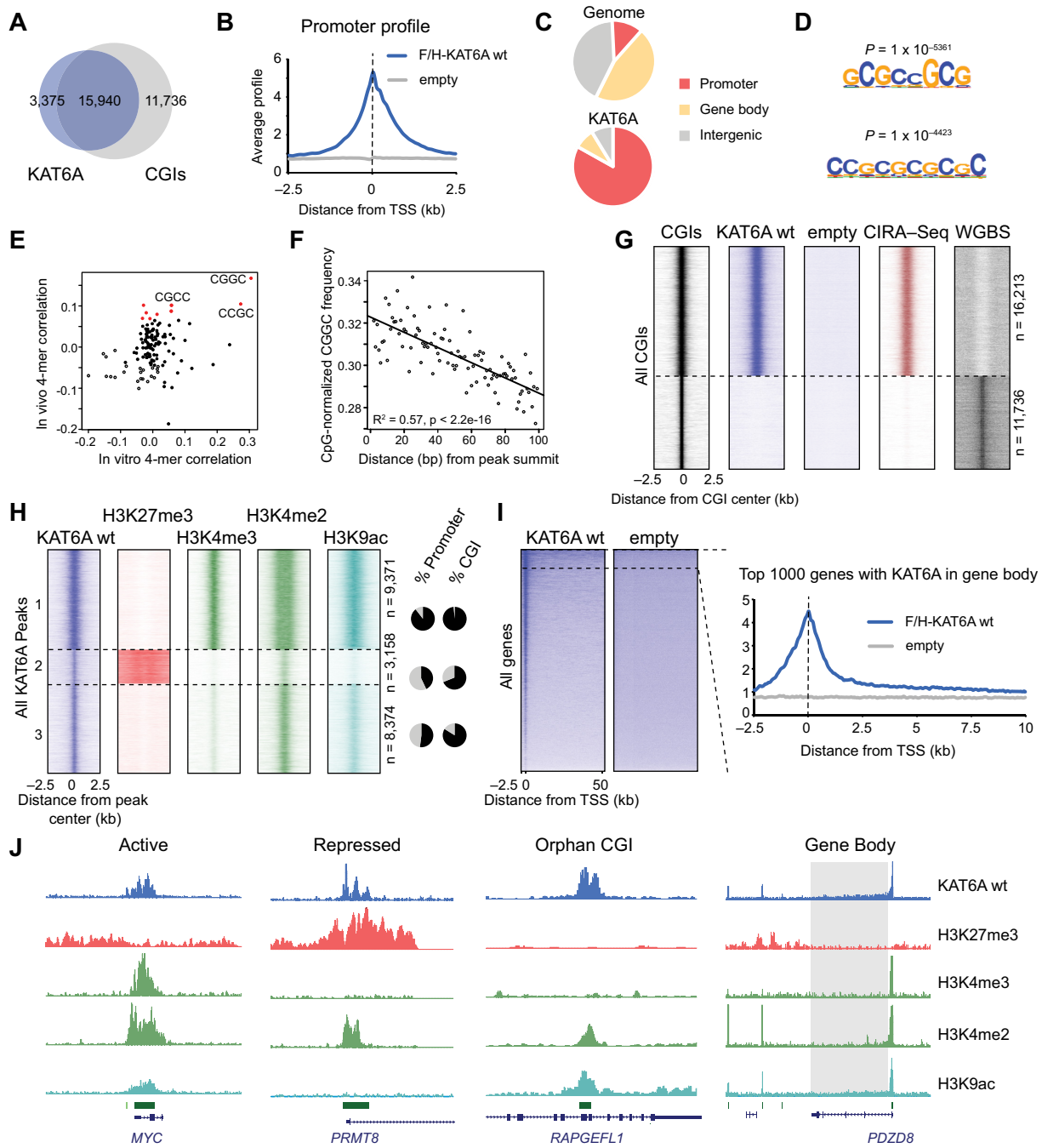
### KAT6A localizes genome-wide to unmethylated CpG islands and gene bodies

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

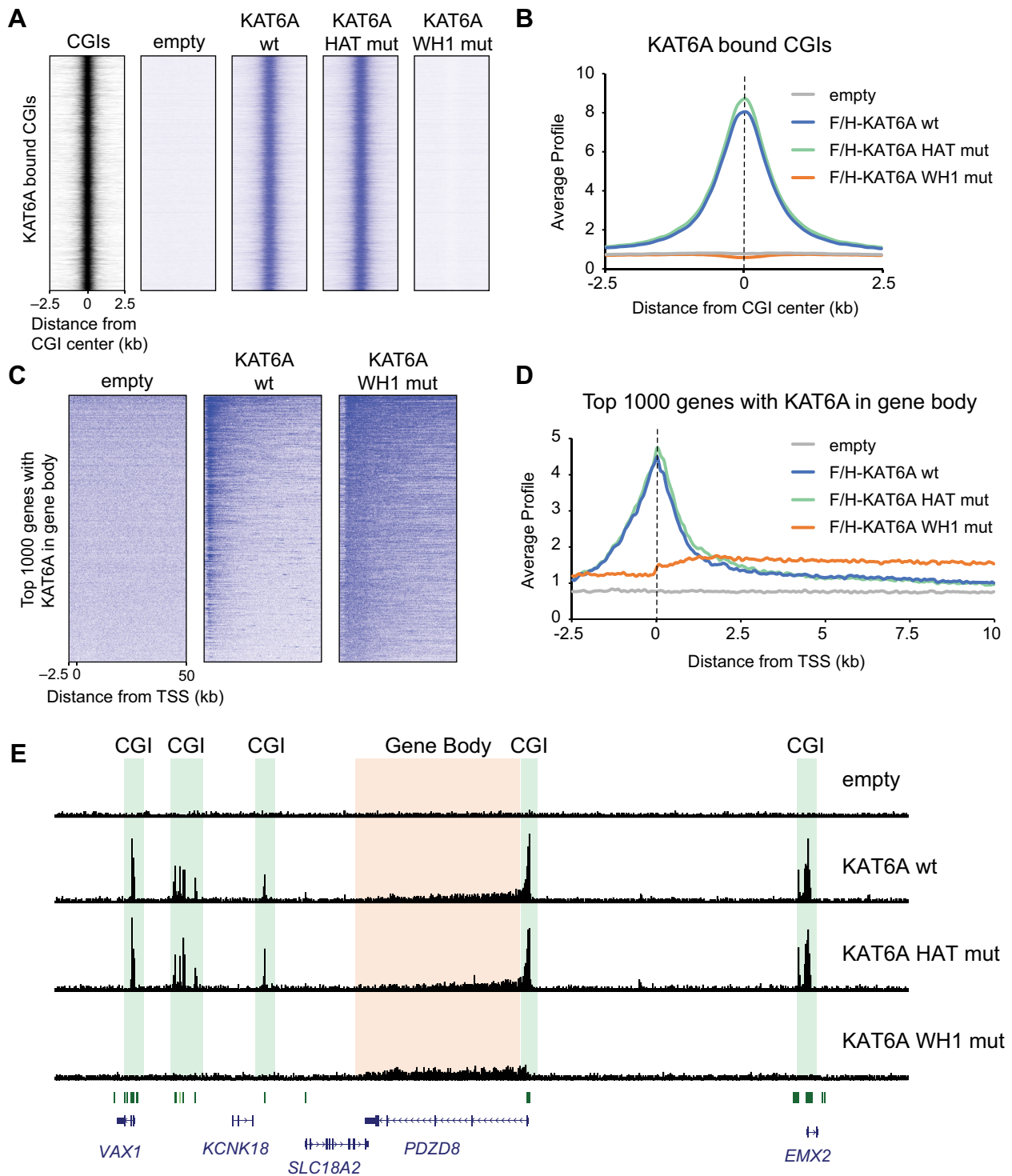
Based on our ChIP-qPCR experiments mutating the WH1 domain of KAT6A almost completely abrogated chromatin binding at the investigated genes (Figure 5D). On the contrary, cellular fractionation experiments suggested that the mutated KAT6A still bound to some extent to chromatin (Figure 5B). To investigate 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.

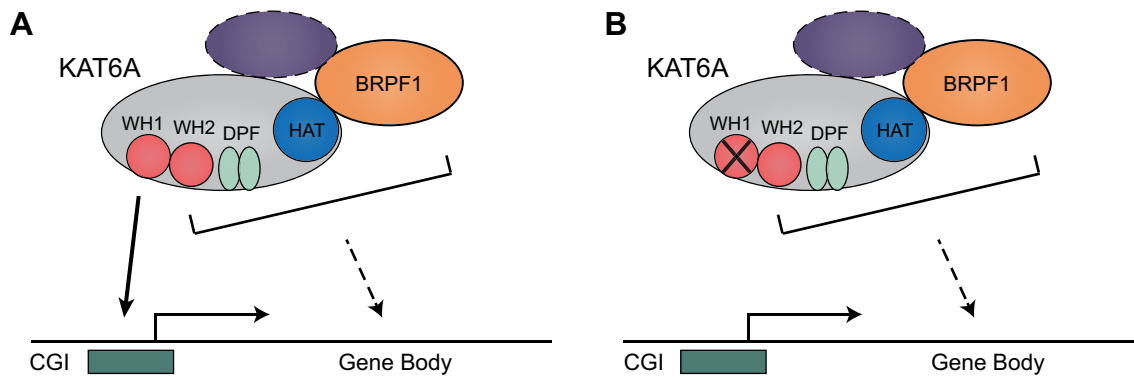
Here, we revealed that the NEMM of KAT6A consists of two winged-helix domains, and we demonstrated that the more N-terminal WH1 domain of KAT6A is required for the recruitment of KAT6A to unmethylated CpG islands (Figure 8A). The structure and the DNA binding mode of the WH1 domain of KAT6A are very 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 CFP1 (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: <https://deeptools.readthedocs.io/>  
 Galaxy Europe: <https://usegalaxy.eu/>  
 GENCODE: <https://www.genecodegenes.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

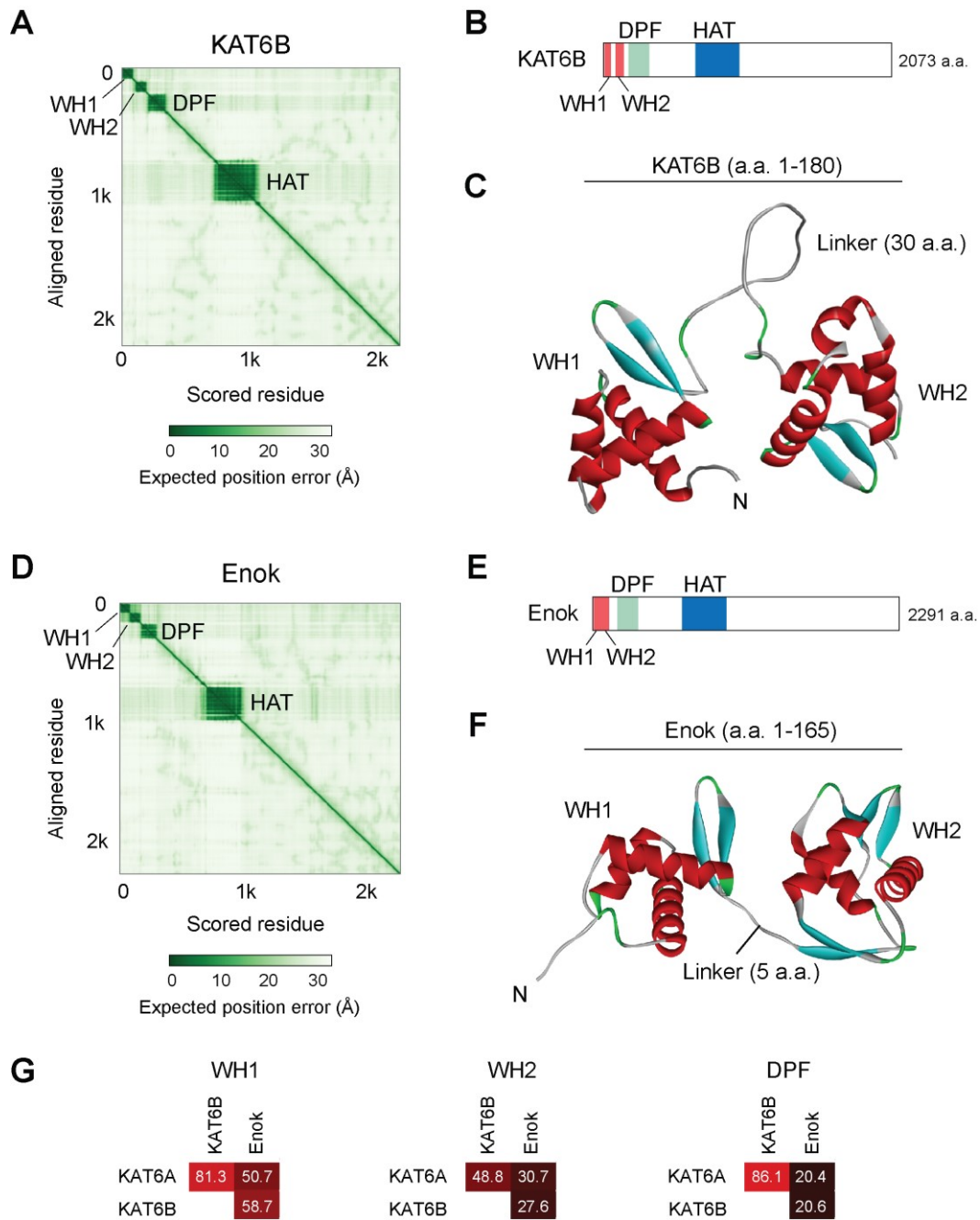
R.L. acknowledges support from the German Research Foundation (DFG) [109546710, 416910386]; Fritz Thyssen Foundation [10.20.1.005MN]; Z.W. was supported by funds from the National Natural Science Foundation of China (NSFC) [32071204, 31870725, 32125008]; M.L.B acknowledges funding from the National Institutes 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., Distchele, 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, I., 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. Sheikth,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 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 hep2 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., Eeckhoutte,J., Johnson,D.S., Bernstein,B.E., Nussbaum,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,E.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 DP2. *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., Hodawadkar, 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 CFPI 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., Rolf, A., Gillissen-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., Gillissen-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.



**Supplementary Figure S1: Domain structures of human KAT6B and Drosophila Enok.**

A) AlphaFold predicted alignment error for KAT6B.

B) Domain structure of KAT6B, based on AlphaFold prediction.

C) AlphaFold predicted structure of WH1 and WH2 of KAT6B.

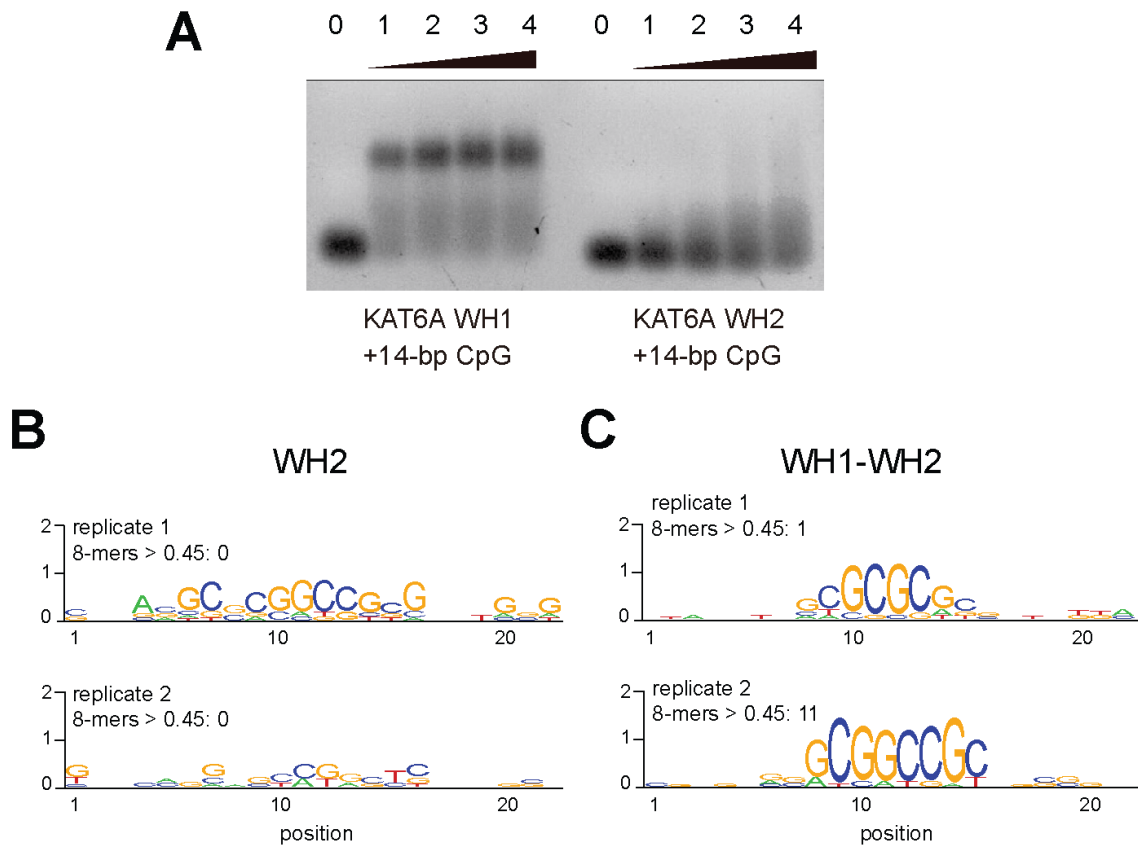
D) AlphaFold predicted alignment error for Enok.

E) Domain structure of Enok, based on AlphaFold prediction.

F) AlphaFold predicted structure of WH1 and WH2 of Enok.

G) Aminoacid identity in percent comparing the N-terminal domains between KAT6A, KAT6B and Enok



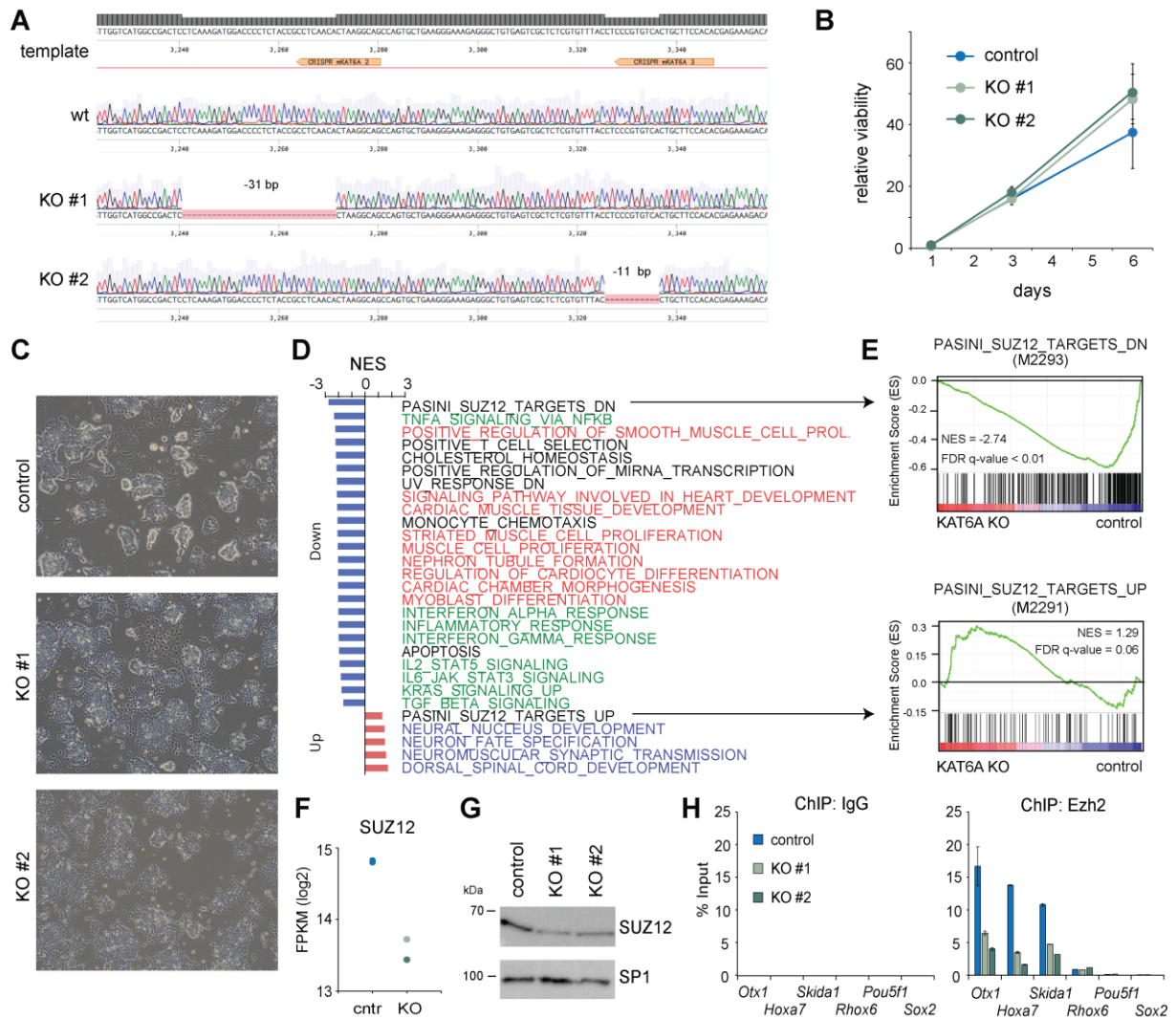


**Supplementary Figure S2: KAT6A WH2 does not bind to DNA**

A) EMSA of the WH1 and WH2 domains with CG-rich DNA. Numbers indicate molar ratio.

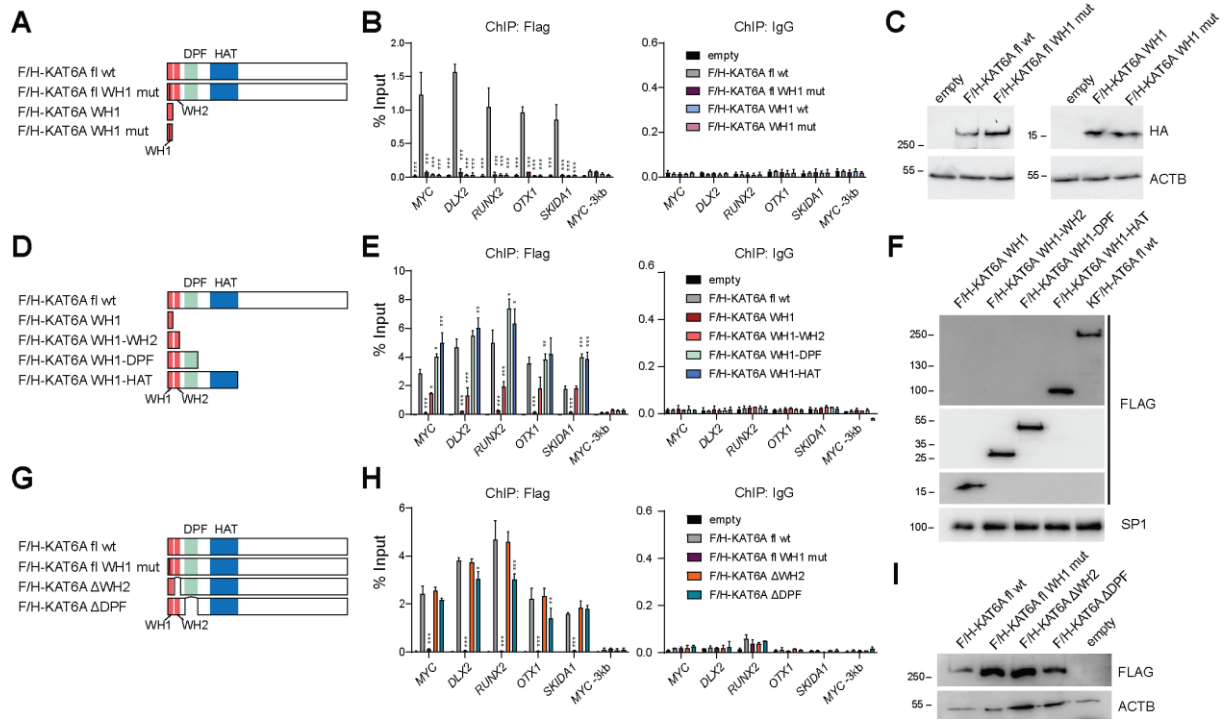
B) PBM results of KAT6A WH2 domain alone. No 8-mer had a score higher than 0.45, implicating no DNA binding ability of this domain.

C) PBM results of KAT6A WH1-WH2 domains in combination.



**Supplementary Figure S3: KAT6A influences developmental pathways in mouse ES cells.**

- A) Validation of CRISPR targeting via Sanger sequencing.  
 B) Proliferation assay of control and KAT6A KO mouse ES cells.  
 C) Bright-field microscopy of control and KAT6A KO mouse ES cells.  
 D) Gene set enrichment analysis (GSEA) of RNA-Seq data demonstrating the dysregulation of pathways involved in development (red), signaling (green) and neuronal pathways (blue).  
 E) GSEA of gene sets dysregulated upon SUZ12 deletion.  
 F) Gene expression of SUZ12 in control and KAT6A KO cells based on the RNA-Seq data.  
 G) Western blot of SUZ12 in control and KAT6A KO cells. (Same samples as in Fig. 3A).  
 H) ChIP-qPCR of EZH2 in control and KAT6A KO cells at PRC2 target genes.



**Supplementary Figure S4: The KAT6A WH1 domain alone is not sufficient for chromatin binding.**

A) Schematic overview of used constructs in B) and C).

B) ChIP-qPCR upon ectopic overexpression in HEK293 cells of constructs shown in A) using a FLAG antibody (IgG as a control).

C) Western blot of ectopic expression in HEK293 cells of constructs shown in A).

D) Schematic overview of used constructs in E) and F)

E) ChIP-qPCR upon ectopic overexpression in HEK293 cells of constructs shown in D) using a FLAG antibody (IgG as a control).

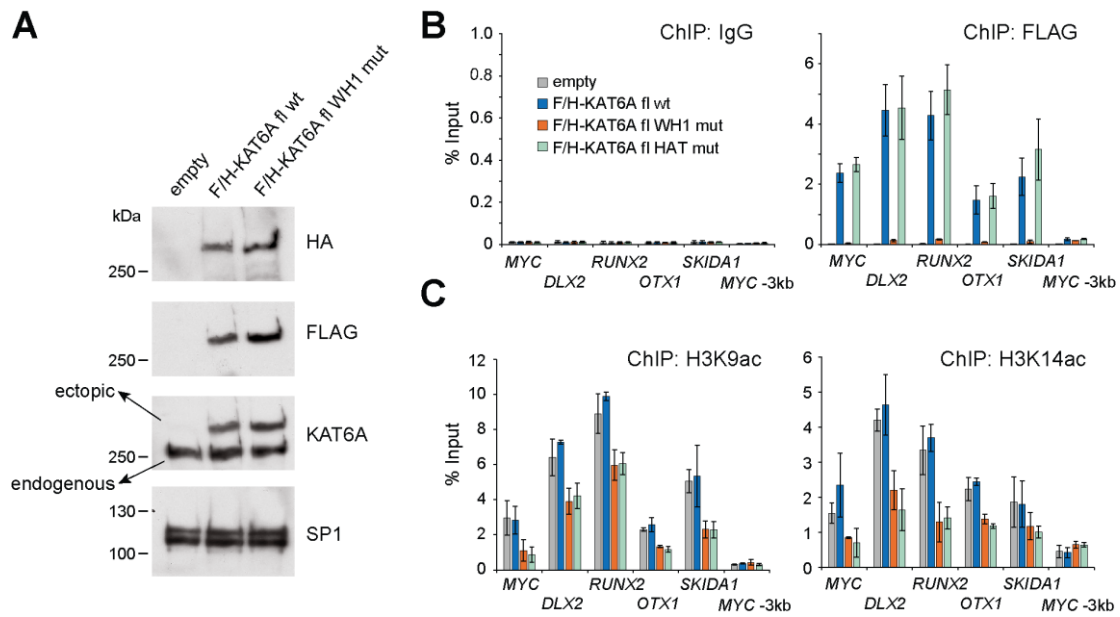
F) Western blot of ectopic expression in HEK293 cells of constructs shown in E).

G) Schematic overview of used constructs in H) and I).

H) ChIP-qPCR upon ectopic overexpression in HEK293 cells of constructs shown in G) using a FLAG antibody (IgG as a control).

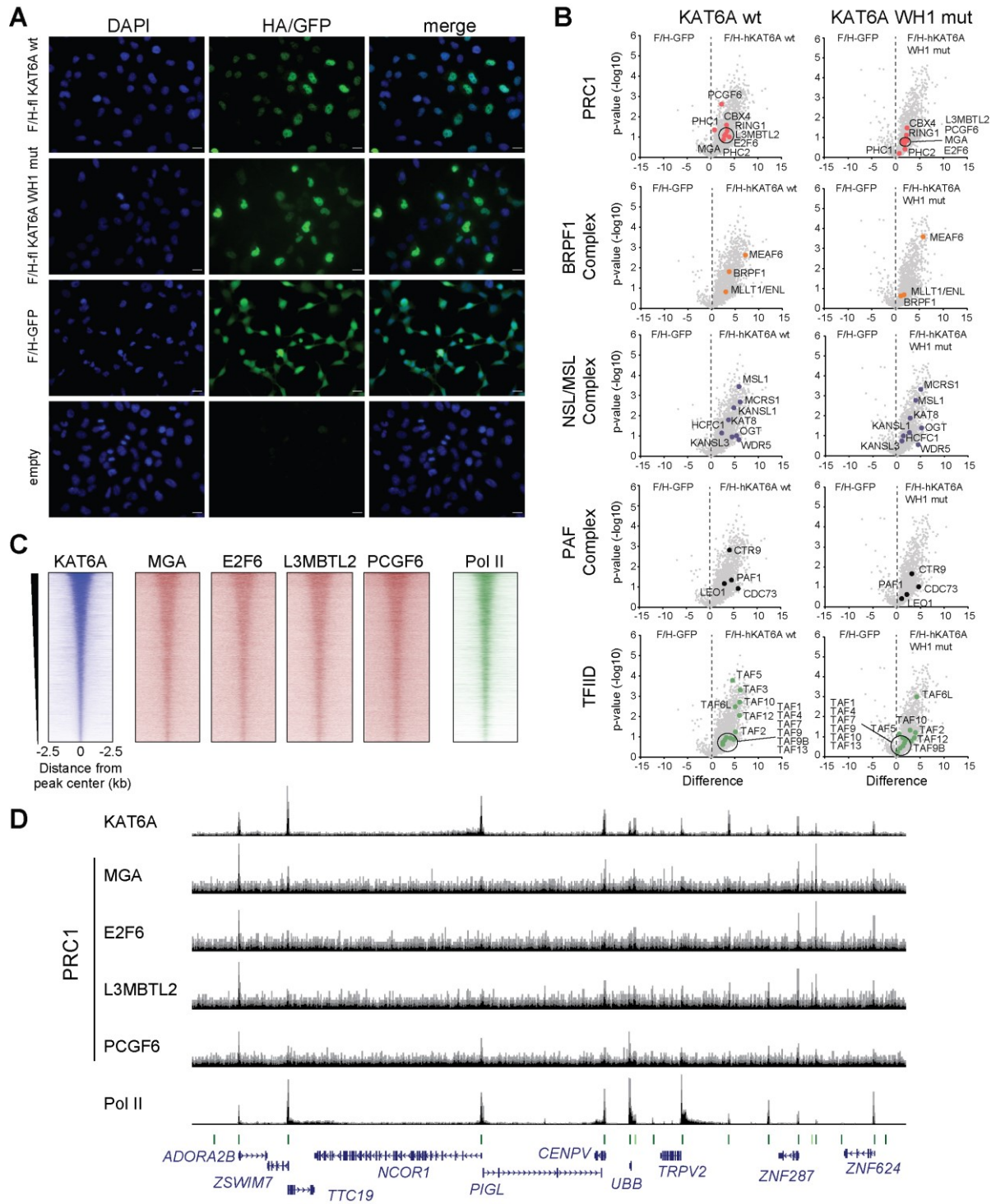
I) Western blot of ectopic expression in HEK293 cells of constructs shown in G).

In B), E), H), data represent 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 wildtype KAT6A. \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$



**Supplementary Figure S5: Overexpression of KAT6A WH1 mutant has a dominant negative effect on histone acetylation**

- A) Western blot of endogenous KAT6A upon ectopically overexpressed KAT6A wt and KAT6A WH1 mut.  
 B) ChIP-qPCR using antibodies for Flag (IgG as negative control) in HEK293 cells with ectopically expressed full-length KAT6A, KAT6A WH1 mutant and KAT6A HAT mutant.  
 C) As in B) with antibodies against H3K9ac (left) and H3K14ac (right)



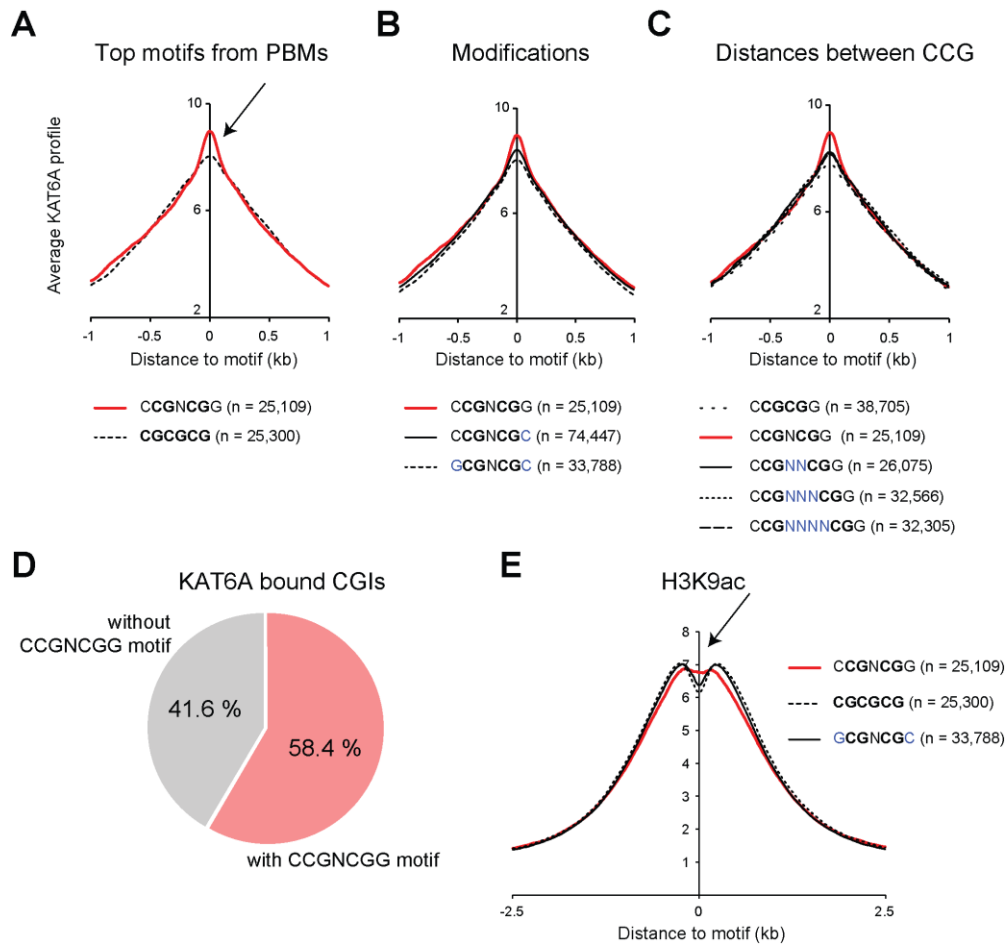
**Supplementary Figure S6: Detailed analysis of mass spectrometry results**

A) Immunofluorescence of ectopically expressed KAT6A in HEK293 cells.

B) Detailed volcano plot of mass spectrometry results from 3G).

C) Heatmap of the overlap of KAT6A with members of PRC1.

D) Genome browser view of the overlap of KAT6A with members of PRC1.



**Supplementary Figure S7: Palindromic CCGNCGG motif leads to local KAT6A enrichment**

A) KAT6A profile around the two motifs, found most enriched in PBM experiments (Figure 1G).

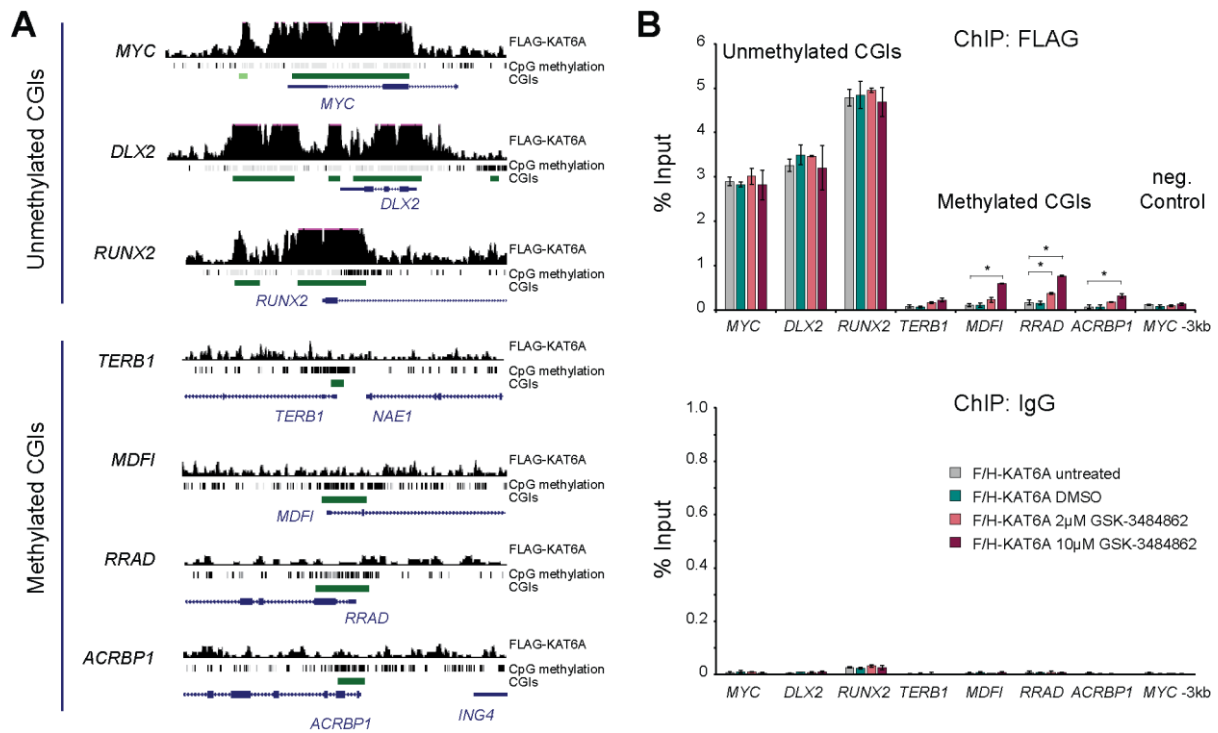
B) KAT6A profile around the palindromic CCGNCGG motif and two modified motifs.

C) KAT6A profile around the palindromic CCGN<sub>x</sub>CGG with various spacers.

D) Pie diagram showing the percentage of CGIs with or without a CCGNCGG motif.

E) H3K9ac profile at CCGNCGG, CGCGCG and a GCGNCGC motifs.

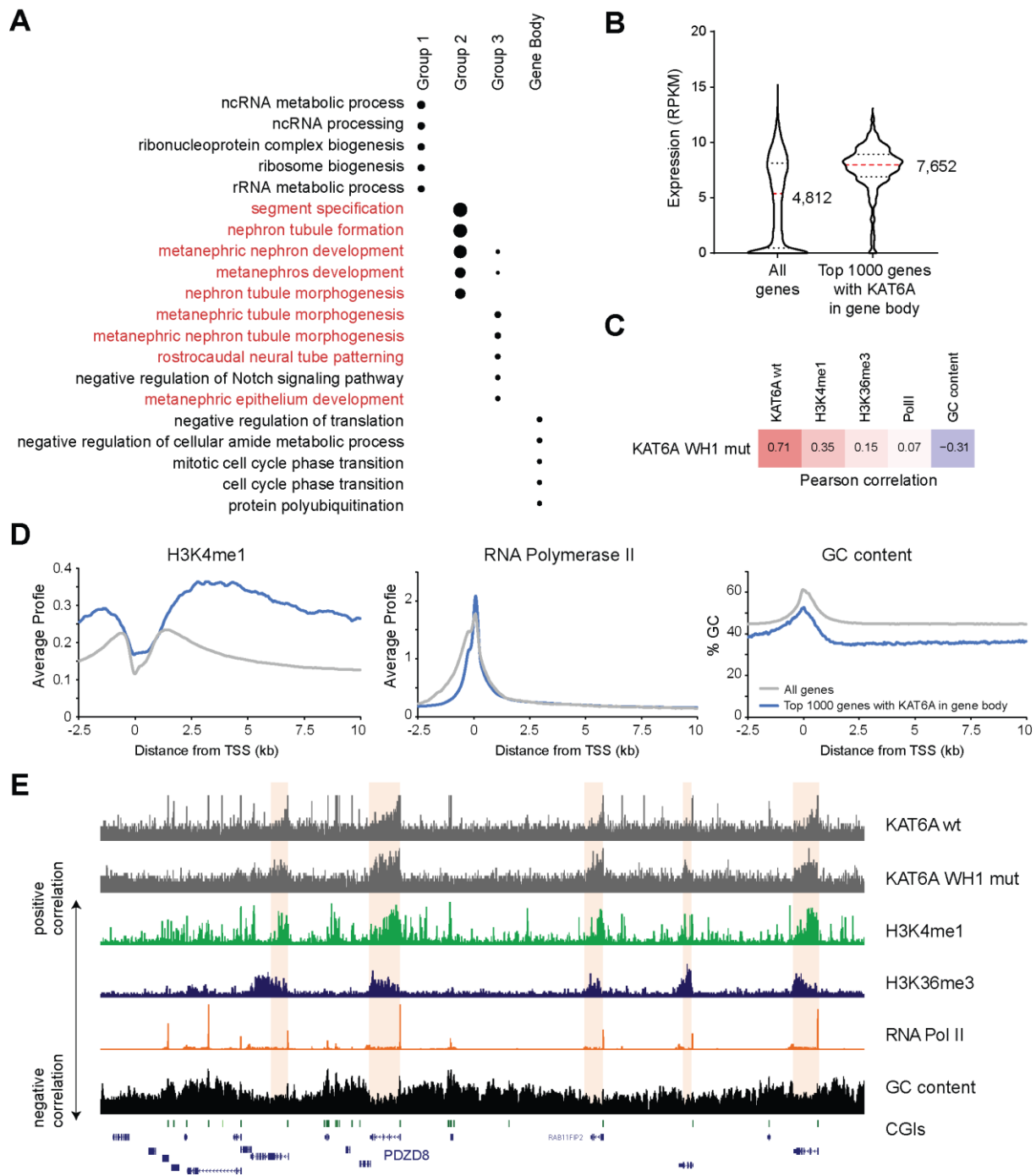
For A, B, C and E only locations within CGIs were analyzed.



**Supplementary Figure S8: Inhibition of DNMT1 leads to de novo recruitment of KAT6A to methylated CGIs.**

A) Example CGIs with or without DNA methylation.

B) ChIP-qPCR of ectopically expressed FLAG-HA-tagged KAT6A wt at CGIs upon DNMT1 inhibitor treatment. Data represent 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. \*  $P < 0.05$ .



**Supplementary Figure S9: Gene ontology of KAT6A-bound clusters and analysis of genes with KAT6A in the gene body.**

A) Gene ontology analysis of the clusters from Figure 5H, using GREAT (46). Gene sets in red are related to kidney development.

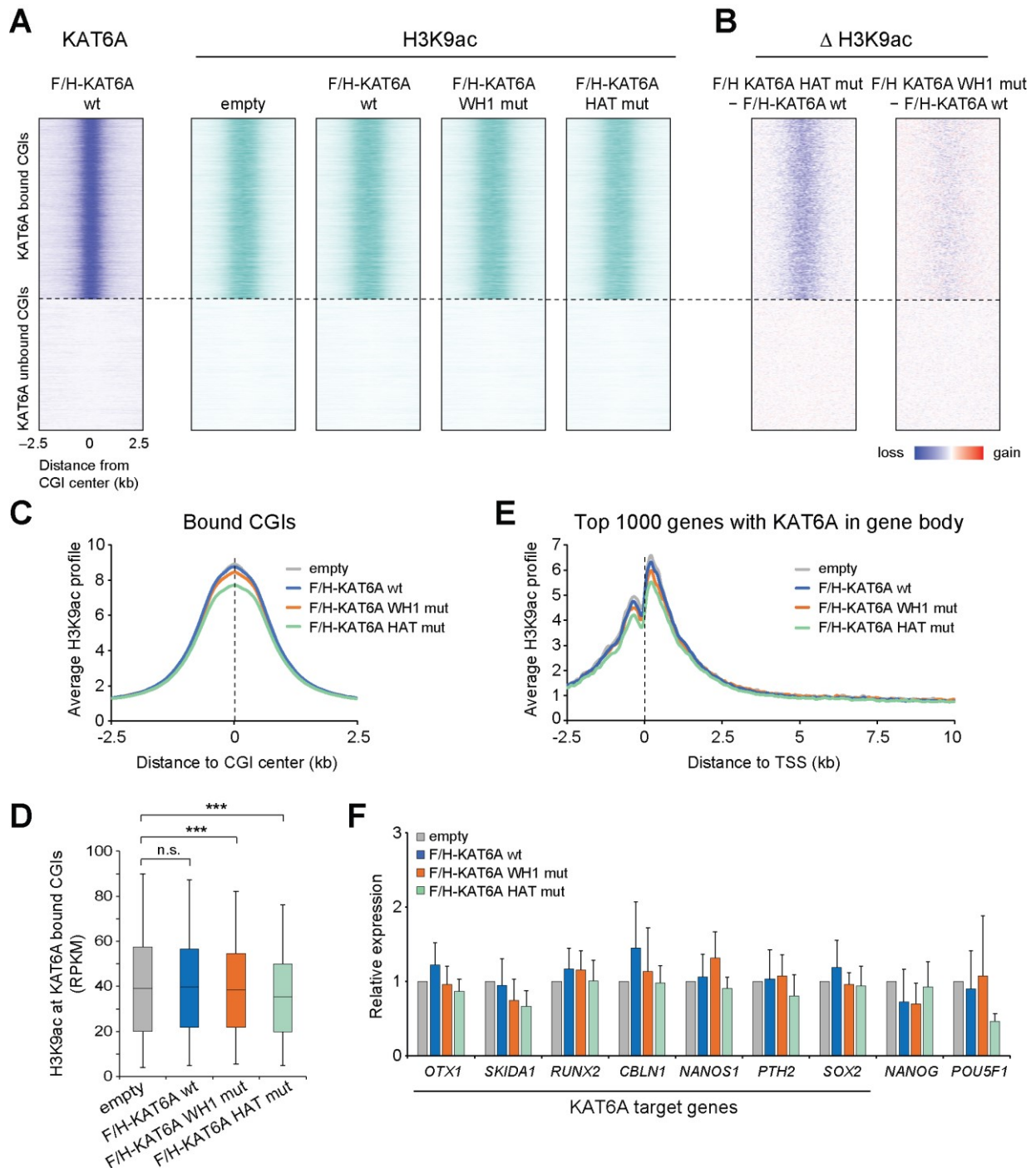
B) Gene expression of the top 1,000 genes with KAT6A in the gene body compared to all genes, using expression data from (9).

C) Correlation of KAT6A levels in gene body (using the data from KAT6A WH1 as reference), with other features of the gene bodies. KAT6A in gene body shows a positive correlation with H3K4me1 and a negative correlation with CG-content.

D) Profile of H3K4me1 (positive correlation), RNA Polymerase II (no strong correlation) and % GC content (negative correlation) of the top 1,000 genes with KAT6A in the gene body compared to all genes.

E) UCSC browser view showing positive correlation of chromatin features with presence of KAT6A in gene bodies.





**Supplementary Figure S10: Overexpression of the KAT6A WH1 mutant has mild dominant negative effects on H3K9ac, similar to a KAT6A HAT mutant.**

A) Heatmap of H3K9ac levels in HEK293 cells transfected with empty vector, KAT6A wt, KAT6A WH1 mut and KAT6A HAT mut. The heatmap shows the H3K9ac levels at CGIs, separated as in Figure 4E.

B) Heatmap of differences in H3K9ac levels upon ectopic expression of the KAT6A HAT mutant and KAT6A WH1 mutant in comparison to ectopic expression of the wild-type protein.

C) Profile of H3K9ac at CGIs.

D) Levels of H3K9ac at KAT6A bound CGIs. \*\*\* =  $P < 0.0001$ .

E) Profile of H3K9ac at the top 1000 genes with KAT6A in the gene body.

F) RT-qPCR measuring the gene expression of representative KAT6A target genes upon ectopic overexpression of KAT6A wild-type, WH1 mutant and KAT6A HAT mutant.

**Supplementary Table S1: Data collection and refinement statistics of KAT6A-WH1/DNA crystals**

Data collection and refinement statistics		
Crystal	KAT6A-WH/DNA 1:1 complex	KAT6A-WH/DNA 2:1 complex
Beam line	SSRF-BL18U1	SSRF-BL19U1
Wavelength	0.97930	0.97853
Space group	<i>P</i> 4 <sub>3</sub> 2 <sub>1</sub> 2	<i>P</i> 1
Unit cell		
a, b, c (Å)	46.5, 46.5, 153.5	40.0, 40.1, 81.3
$\alpha$ , $\beta$ , $\gamma$ (°)	90.0, 90.0, 90.0	90.1, 90.0, 113.9
Resolution (Å)	50.0-1.50 (1.53-1.50) <sup>a</sup>	50.0-1.92 (1.96-1.93)
R <sub>sym</sub>	0.097 (1.290)	0.046 (0.399)
I/ $\sigma$ (I)	41.7 (2.0)	21.0 (2.0)
Completeness (%)	100 (100)	97.6 (96.1)
Redundancy	24.7 (22.5)	3.5 (3.4)
Number of unique reflections	28018	33965
R <sub>work</sub> /R <sub>free</sub> (%)	23.4/24.6	19.5/23.5
Number of none-H atoms		
Protein	626	2439
DNA	568	1054
Water	147	257
Average B factors (Å <sup>2</sup> )		
Protein	24.9	34.8
DNA	46.9	39.5
Water	34.7	37.7
R.m.s. deviations		
Bond lengths (Å)	0.007	0.008
Bond angles (°)	1.006	0.956

<sup>a</sup> Highest resolution shell (in Å) shown in parentheses.

## Supplementary Table S2 – qPCR Primers

### RT-qPCR

#### *Housekeeping genes*

RT mActb fw	GTA-CCC-AGG-CAT-TGC-TGA-CA
RT mActb rv	AGG-GTG-TAA-AAC-GCA-GCT-CAG
RT mGapdh fw	AGA-CGG-CCG-CAT-CTT-CTT-GT
RT mGapdh rv	GCC-TTG-ACT-GTG-CCG-TTG-AA
RT hGAPDH fw	AGC-CAC-ATC-GCT-CAG-ACA-C
RT hGAPDH rv	GCC-CAA-TAC-GAC-CAA-ATC-C
RT hHPRT fw	TGA-CCT-TGA-TTT-ATT-TTG-CAT-ACC
RT hHPRT rv	CGA-GCA-AGA-CGT-TCA-GTC-CT

#### *Target genes*

RT mKat6A fw	CGT-GTC-AGT-TTG-GGG-CAT-CT
RT mKat6A rv	AAG-CCA-GTC-TAG-GGG-GTC-AA
RT mRhox6 fw	CTG-GCT-CAA-CTG-CGG-TAC-AG
RT mRhox6 rv	ACC-AAT-TCT-GCA-CAT-CAC-ATT-CA
RT mOtx1 fw	AAC-CGA-GCA-AGA-CAA-GCC-AC
RT mOtx1 rv	GTT-CAT-GCC-GTA-TGG-GGG-TT
RT mSox9 fw	CTC-AGC-AAG-ACT-CTG-GGC-AA
RT mSox9 rv	TGC-TCA-GTT-CAC-CGA-TGT-CC
RT mSkida1 fw	CAG-CAT-CTG-GCG-GAG-GAT-TA
RT mSkida1 rv	GGT-CGT-CCT-CGG-GAT-GTT-TT
RT mPou5f1 fw	TGG-AAA-GGT-GTT-CAG-CCA-GA
RT mPou5f1 rv	CCT-CAC-ACG-GTT-CTC-AAT-GC
RT hOTX1 fw	CAA-GCC-ACT-CCG-ACA-AGG-TT
RT hOTX1 rv	ATG-CCG-TAT-GGG-GGT-TGT-TT
RT hRUNX2 fw	CGC-CTC-ACA-AAC-AAC-CAC-AG
RT hRUNX2 rv	TCA-CTG-TGC-TGA-AGA-GGC-TG
RT hSKIDA1 fw	TCG-CCT-AGA-GTT-TGG-CAC-TT
RT hSKIDA1 rv	GGT-CGT-CCT-CGG-GAT-GTT-TT
RT hCBLN1 fw	GAA-CGC-AGC-ACT-TTC-ATC-GC
RT hCBLN1 rv	ATC-AAG-TTT-CCC-CGC-TCC-AG
RT hSOX2 fw	CAT-GGG-TTC-GGT-GGT-CAA-GT
RT hSOX2 rv	AGC-TGT-CAT-TTG-CTG-TGG-GT
RT hPTH2 fw	AGG-TGT-CCA-GGA-GCC-CTC
RT hPTH2 rv	ATC-CAA-CAC-CAG-CAG-CTT-GT
RT hNANOS1 fw	CGG-AAC-AAC-AAG-GAG-GCG-AT
RT hNANOS1 rv	AGA-GCG-GGC-AGT-ACT-TGA-TG
RT hNANOG fw	AAT-GGT-GTG-ACG-CAG-GGA-TG
RT hNANOG rv	AAG-GAT-TCA-GCC-AGT-GTC-CAG
RT hPOU5F1 fw	GGC-TTC-GGA-TTT-CGC-CTT-CT
RT hPOU5F1 rv	CCC-CCA-CAG-AAC-TCA-TAC-GG

### ChIP -qPCR

ChIP mOtx fw	AGT-AGG-CGT-GCT-CAG-AGA-GG
ChIP mOtx rv	GGC-CGG-TCA-AGA-AGA-AGT-C
ChIP mHoxa7 fw	GAG-AGG-TGG-GCA-AAG-AGT-GG
ChIP mHoxa7 rv	CCG-ACA-ACC-TCA-TAC-CTA-TTC-CTG

ChIP mSkida1 fw	CAC-ATC-CTG-CAA-AAC-CGA-GC
ChIP mSkida1 rv	CGA-ACG-ATC-TCC-GGG-TAG-TG
ChIP mRhox6 fw	TAA-TTT-GGG-TGA-GGC-GCC-AG
ChIP mRhox6 rv	CTC-TTC-TTT-CCA-CCC-ACC-GTT
ChIP mPou5f1	GTT-TGT-GAG-GTG-TCC-GGT-GA
ChIP mPou5f1	GCT-CAC-CTA-GGG-ACG-GTT-TC
ChIP mSox2 fw	CCA-TCC-ACC-CTT-ATG-TAT-CCA-AG
ChIP mSox2 rv	CGA-AGG-AAG-TGG-GTA-AAC-AGC-AC
ChIP hMYC CGI fw1	AGC-TGG-CAA-AAG-GAG-TGT-TG
ChIP hMYC CGI rv1	AAA-GTT-TTG-CGC-CAC-CTG-AA
ChIP hDLX2 fw	AAT-CGT-AAG-AAC-AGC-GCA-ACC
ChIP hDLX2 rv	CCT-TGC-CAC-CAG-CTT-CTT-TC
ChIP hRUNX2 fw	CAA-CTT-CCT-GTG-CTC-GGT-GC
ChIP hRUNX2 rv	TTT-CTG-GGG-TTA-GAG-CCG-CC
ChIP hOTX1 fw	AAT-TTC-GTT-GGT-CTG-GCG-GC
ChIP hOTX1 rv	GGA-GCC-CGC-TGA-GTG-AAA-AG
ChIP hCBLN1 fw	GCG-CCG-GCT-GCA-TCA-ATA-AT
ChIP hCBLN1 rv	TCG-CAC-TCC-GGG-ACT-AGC-GT
ChIP hSKIDA1 fw	TGC-GTT-ATT-GCA-AGC-TCA-GG
ChIP hSKIDA1 rv	CCA-GCA-TGC-TCC-CCT-ATT-GA
ChIP hGAPDH fw	GAG-CCT-CCT-TCC-TCT-CCA-G
ChIP hGAPDH rv	GTA-GTG-ACA-CCG-GAC-TGC-T
ChIP hMYC -3kb fw	AAC-CTC-CAC-TGC-CAG-AAG-TCC-TTA
ChIP hMYC -3kb rv	GAA-ATT-TAC-CTG-GCA-CGT-GTC-CCT
ChIP hTERB1 fw	CAT-TGC-CCT-TGG-CAA-CCT-AG
ChIP hTERB1 rv	CCC-TTT-CTT-CAT-CAC-CGC-TC
ChIP hMDFI fw	AAA-CGG-AGA-GAG-ACG-GTG-AG
ChIP hMDFI rv	TCT-GGA-GAG-CTG-AGC-CAA-CT
ChIP hRRAD fw	GGC-GGG-TGG-AGG-CTT-AAA-TA
ChIP hRRAD rv	ACA-CAG-ACA-CGC-TCA-GGA-CT
ChIP hACRBP1 fw	ATG-GGC-TCT-GGC-CAC-TTT-AG
ChIP hACRBP1 rv	TCA-GGA-GTG-AGG-GAA-GGA-AG

Sequencing primers

Seq mKAT6A fw1	ATG-GCC-GAC-TCC-TCA-AAG-AT
Seq mKAT6A rv1	ACA-CTG-GTG-CTA-ACG-TTC-CT



# OPEN Investigation of SAMD1 ablation in mice

Bruce Campbell<sup>1,8</sup>✉, Lisa M. Weber<sup>2,8</sup>, Sandra J. Engle<sup>4,7</sup>, Terence R. S. Ozolinš<sup>5,7</sup>, Patricia Bourassa<sup>6,7</sup>, Robert Aiello<sup>6,7</sup> & Robert Liefke<sup>2,3</sup>✉

SAM domain-containing protein 1 (SAMD1) has been implicated in atherosclerosis, as well as in chromatin and transcriptional regulation, suggesting a versatile and complex biological function. However, its role at an organismal level is currently unknown. Here, we generated SAMD1<sup>-/-</sup> and SAMD1<sup>+/-</sup> mice to explore the role of SAMD1 during mouse embryogenesis. Homozygous loss of SAMD1 was embryonic lethal, with no living animals seen after embryonic day 18.5. At embryonic day 14.5, organs were degrading and/or incompletely developed, and no functional blood vessels were observed, suggesting failed blood vessel maturation. Sparse red blood cells were scattered and pooled, primarily near the embryo surface. Some embryos had malformed heads and brains at embryonic day 15.5. In vitro, SAMD1 absence impaired neuronal differentiation processes. Heterozygous SAMD1 knockout mice underwent normal embryogenesis and were born alive. Postnatal genotyping showed a reduced ability of these mice to thrive, possibly due to altered steroidogenesis. In summary, the characterization of SAMD1 knockout mice suggests a critical role of SAMD1 during developmental processes in multiple organs and tissues.

Human SAMD1 (SAM domain-containing protein 1) is a 538 amino acid protein that has orthologs in most vertebrates, including zebrafish. Originally, SAMD1 was identified as a protein involved in atherosclerotic low-density lipoprotein (LDL) binding in humans<sup>1</sup> and mice<sup>2</sup>. SAMD1 may participate in LDL retention and uptake, and its knockdown suppresses vascular smooth muscle cell (VSMC) differentiation and proliferation, which affects foam cell development<sup>3</sup>. However, SAMD1 is widely expressed in most organs and tissues, both in mice and humans (see ProteinAtlas, GTEx, etc.). The extensiveness of expression suggests roles in adult homeostasis, but the molecular function of SAMD1 has been minimally studied so far. SAMD1 was found to directly associate with CpG islands at the chromatin, acting as a transcriptional regulator, and be required for proper embryonic stem (ES) cell differentiation<sup>4</sup>. Earlier reports further support epigenetic functions for SAMD1. SAMD1 was enriched at nucleosomes and H3K4me3-possessing chromatin regions, as revealed by mass spectrometry<sup>5,6</sup>. It was also among the highest 5% of proteins that associate with both H3K4me3-modified and bivalently H3K4me3/H3K27me3-modified chromatin<sup>7</sup>.

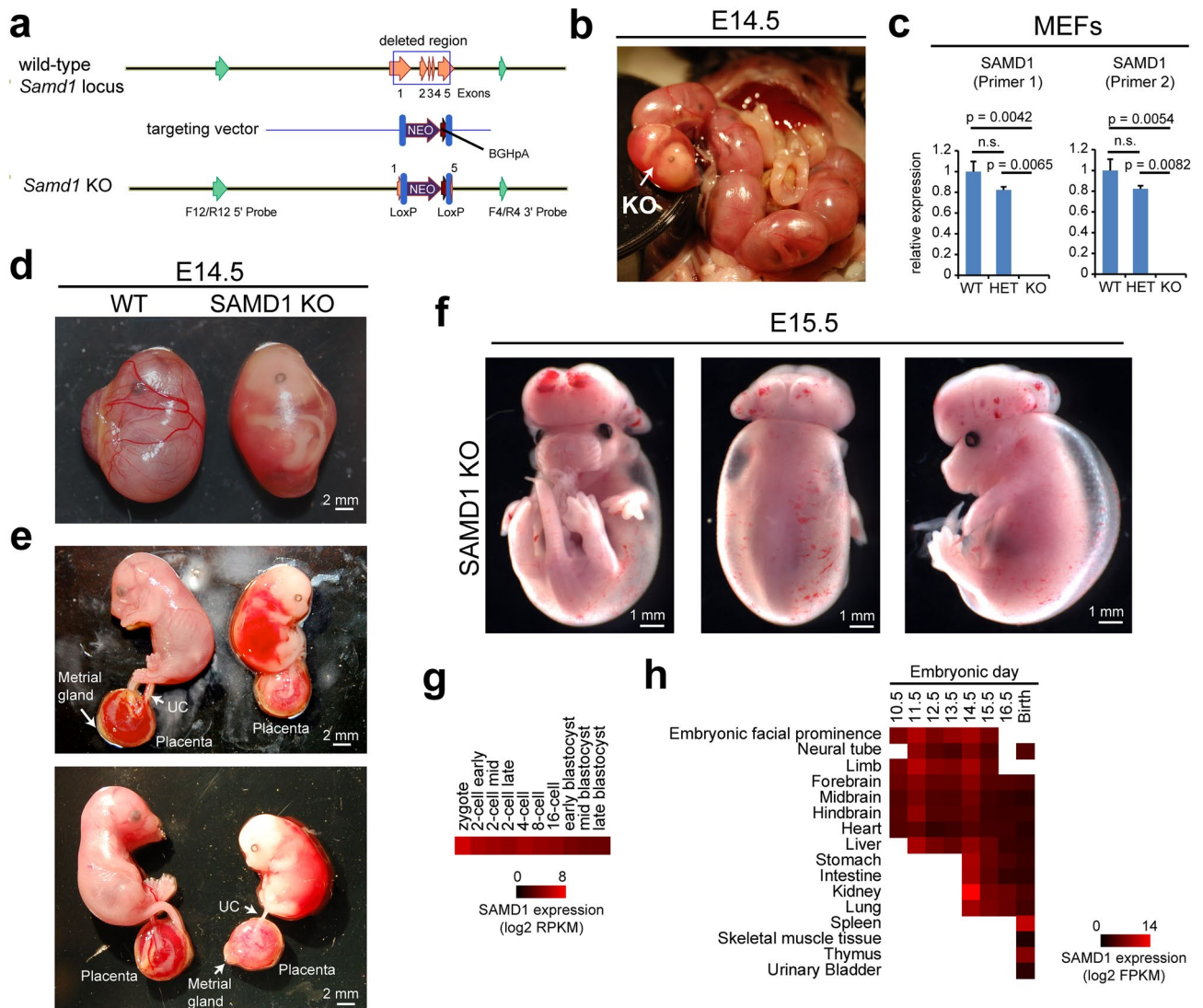
The SAMD1 protein has two globular domains. At the C-terminus, it possesses a SAM domain that can interact with other SAM domain-containing proteins, such as L3MBTL3, and allows multimerization<sup>4</sup>. The N-terminal domain belongs to a group of related winged helix domains that bind to unmethylated CpG motifs<sup>4,8</sup>, explaining the enrichment of SAMD1 at CpG islands on the genome<sup>4,9</sup>. SAMD1 associates with repressive chromatin regulatory complexes that contain the histone demethylase KDM1A and several other SAM domain proteins, including the Polycomb group-related proteins L3MBTL3 and SFMBT1<sup>4,10–12</sup>. Consequently, the absence of SAMD1 typically leads to the derepression of its target genes<sup>4,9</sup>. SAMD1 is often upregulated in cancer tissues and correlates with worse prognosis in some cancer types, such as adenoid cystic carcinoma (ACC) and liver cancer<sup>9,13</sup>. A CRISPR screen in K562 cells suggested that SAMD1 is required for the efficient proliferation of these cells<sup>14</sup>. Further work demonstrated that SAMD1 deletion in HepG2 hepatocellular carcinoma cells leads to a global readjustment of the active H3K4me2 chromatin mark and a more favorable gene signature, supporting an oncogenic role in this context<sup>9</sup>. SAMD1 has also been described to play a role in muscle adaptation<sup>15</sup>. Taken together, these findings suggest that SAMD1 likely has a function in multiple biological processes and diseases, but the biological role of SAMD1 at the organismal level has yet to be investigated.

<sup>1</sup>Atherex Inc., Lincoln, MA 01773, USA. <sup>2</sup>Institute of Molecular Biology and Tumor Research (IMT), Philipps University of Marburg, 35043 Marburg, Germany. <sup>3</sup>Department of Hematology, Oncology, and Immunology, University Hospital Giessen and Marburg, 35043 Marburg, Germany. <sup>4</sup>Biogen, Cambridge, MA 02142, USA. <sup>5</sup>Department of Biomedical and Molecular Sciences, Queen's University, Kingston, ON K7L 3N6, Canada. <sup>6</sup>Cybrexa Therapeutics, Groton, CT 06340, USA. <sup>7</sup>Pfizer Inc., Groton, CT 06340, USA. <sup>8</sup>These authors contributed equally: Bruce Campbell and Lisa M. Weber. ✉email: bcampbell9516@gmail.com; robert.liefke@imt.uni-marburg.de

Here, we report the first knockout of SAMD1 in mice. Fewer homozygous ( $SAMD1^{-/-}$ ) embryos were observed at embryonic day (E) 14.5, and none were observed after E18.5. Development appeared grossly normal at E12.5, but at E14.5 the investigated embryos exhibited numerous defects, including degrading endothelial cell (EC) tubes, degrading inner organs and an almost complete absence of red blood cells (RBCs). The heterozygous ( $SAMD1^{+/-}$ ) mice had a less severe phenotype and approximately 70% of the mice survived past week 3 postnatally but they exhibited reduced weight gain and hormonal changes.

## Results

**SAMD1 knockout mice are embryonic lethal.** The *SAMD1* gene was deleted from C57BL/6J mice by recombineering (Fig. 1a), which led to the removal of 2396 bp of the *SAMD1* gene. Successfully obtained heterozygous mice were mated. The ratios between the wild-type (WT, aka  $SAMD1^{+/+}$ ), heterozygous (HET, aka  $SAMD1^{+/-}$ ), and knockout (KO, aka  $SAMD1^{-/-}$ ) mice were close to the expected Mendelian ratios through E12.5. All embryos removed at E12.5 had heartbeats. The first dead *SAMD1* KO embryos and abnormal phenotypes were observed at E14.5 (Fig. 1b). At this time only 16 of the Mendelian expected 25 KO embryos were present, meaning that approximately 9 embryos had been resorbed since E12.5 ( $p=0.13$ ). Additional reductions



**Figure 1.** *SAMD1* KO mice are embryonic lethal and lack intact blood vessels. **(a)** Gene targeting strategy to obtain *SAMD1* KO mice. Southern probes are marked in green. **(b)** A *SAMD1* KO embryo is easily distinguished from the WT embryos as soon as the litter is partially removed from the uterus. **(c)** RT-qPCR of *SAMD1* in mouse embryonic fibroblasts (MEFs) obtained from wild-type and *SAMD1* homo- and heterozygous embryos. Data are presented as the mean  $\pm$  SD of two biological replicates. P-values via two-way ANOVA with Tukey's multiple comparisons test. n.s. = not significant. **(d)** Pictures of embryos at E14.5 within the yolk sac. **(e)** Pictures of embryos at E14.5, without yolk sac, but still within the amnion. The placenta is shown underneath. **(f)** Picture of a *SAMD1* KO mouse at approx. E15.5. **(g)** Gene expression of *SAMD1* during early embryogenesis<sup>23</sup>. **(h)** *SAMD1* gene expression in various embryonic tissues based on data from ENCODE<sup>24</sup>.

in KO embryos manifested at later time points (Supplementary Table S1) ( $p=0.049$ ). Dying pups were not seen, so the few KO mice that had not been resorbed by E18.5 died just prior to birth (Supplementary Table S1). This suggests there was variable expression and penetrance, and thus multiple embryonic lethal SAMD1 KO phenotypes. Crossing SAMD1<sup>+/-</sup> mice with SAMD1<sup>+/-</sup> mice and phenotyping 3 weeks after birth (P21) also showed no KO ( $p=5.1 \times 10^{-13}$ ) (Supplementary Table S2), confirming the lethality of the SAMD1 KO.

In contrast, the E14.5 heterozygous (HET) embryos were not noticeably different from that of the WT embryos, and HET mice were born alive (Supplementary Table S1). This observation suggests that heterozygous deletion of SAMD1 had no or only a minor impact on embryogenesis. To understand the underlying reason, embryonic fibroblasts (MEFs) raised from WT, HET, and KO mice were investigated for SAMD1 expression via RT-qPCR. SAMD1 mRNA was slightly but not significantly reduced in the heterozygous cells and absent in the homozygous KO cells (Fig. 1c). The mild reduction in SAMD1 expression in the heterozygous cells raises the possibility that a feedback mechanism at the wildtype allele compensates for the loss of SAMD1 expression from the knockout allele<sup>16</sup>. This mechanism could be the underlying reason for the mostly normal development of the heterozygous mice. However, genotyping of the surviving mice at P21 showed that the number of observed HETs was significantly lower than the expected number, implying that the heterozygous mice also have some impairments. Reduced numbers of HET mice at P21 were observed both when SAMD1<sup>+/-</sup> mice were crossed with SAMD1<sup>+/-</sup> mice ( $p=0.003$ ) (Supplementary Table S2), and when SAMD1<sup>+/-</sup> mice were crossed with SAMD1 $\pm$  mice ( $p=0.003$ ) (Supplementary Table S3).

**Defects of the SAMD1 KO embryos are visually obvious.** Given the considerably severer impairment of KO embryogenesis compared to HET mice, we first analyzed the KO embryos in more detail. We focused on E14.5 embryos for further investigations, because the E12.5 KO mice appeared normal, and E14.5 has been established as an optimal time point for mouse developmental disorder phenotype analyses<sup>17,18</sup>. Macrolevel photography of the E14.5 WT embryos within the yolk sac showed normal development, noting branching blood vessels in the yolk sac (Fig. 1b,d). In contrast, the yolk sac of the KO embryos lacked obvious blood vessels, it had only a few broken thin red lines, and the amniotic fluid was pink as if blood had leaked into it (Fig. 1b,d). In the KO mice the yolk sac did not seem to be normally attached to the placental disk, suggesting initiation of the abortion process.

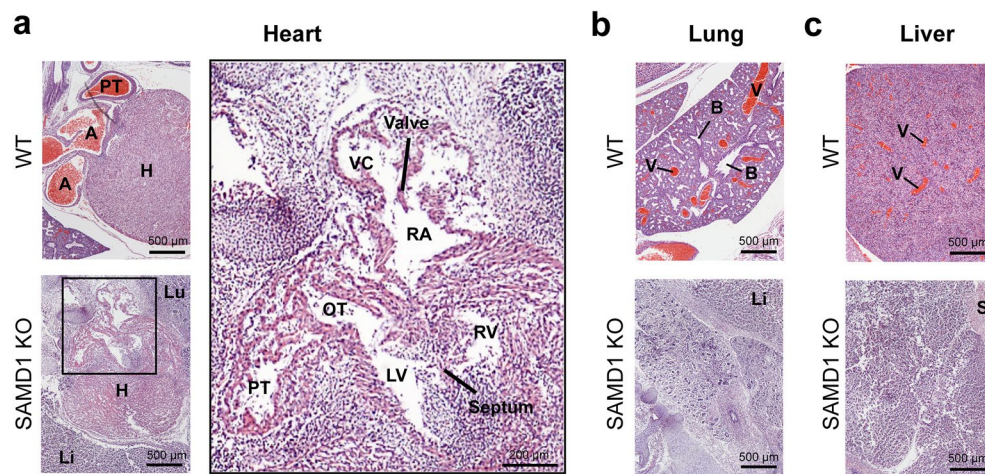
When the yolk sac was removed, the KO embryo within the amnion appeared developmentally delayed and visually smaller than the WT embryo (Fig. 1e). Bloody fluid was also pooled around the midsection within the amnion, but the source of leakage was not obvious (Fig. 1e). The embryo surface was very pale, and across the back, clear edema could be seen separating the skin from the tissue beneath. Red blood cells were not obvious in the skin, other than on the head and back, where a few small hemorrhages and a few broken red lines that were probably failed blood vessels were recognizable. This observation may suggest abnormal vessel regression. Regression, also called “pruning”, is a normal part of embryonic maturation of blood vessels<sup>19,20</sup>. It can include pruning of individual EC (endothelial cell) tubes, as well as complete regional network regression, leaving behind “sleeves” of collagen and apoptotic ECs<sup>21</sup>. During maturation, progenitor/pericyte/smooth muscle cells connect to EC tubes, providing stability needed to control regression<sup>20</sup>. The presence of RBCs in the KO skin implies that a functional circulatory system existed prior to an apparent abnormal regression of the EC tube vascular plexus in the KO embryo.

Similar to the embryo, the KO placenta was pale compared to the WT (Fig. 1e), probably because the labyrinth contained fewer RBCs and the vasculature seemed to be pink instead of red. The metrial gland lacked obvious blood vessels and the decidua appeared to be thinner. Maternal blood vessels could be seen in a few locations immediately beneath the decidua. The umbilical cord was avascular, showing cessation of maternal-embryo blood flow (Fig. 1e).

At E15.5 (Fig. 1f), a SAMD1 KO embryo lacked a skull vault, had obvious exencephaly, and had a hypertrophic brain. Clear edema fluid separated the skin from most of the embryo surface. This kind of edema appears in lymph endothelial cell progenitor knockouts<sup>22</sup>. The E15.5 embryo was less pale than the E14.5 embryo, had longer and larger red lines, and had a few apparently intact blood vessels. Some areas of the embryo surface in and beneath the skin, and above the brain, had small hemorrhages and lines of RBCs that may have once been contained in vessels, suggesting a failing circulatory system. Since the E14.5 KO had a skull vault and was paler, the investigated E15.5 KO embryo is likely an example of a different SAMD1 KO phenotype, but the surface RBC patterns appear similar to the E14.5 KO. Failed skull vault development may be caused by failure of neural tube closure. Clear fluid leakage due to failing lymphatic and blood vessels likely produced edema (Fig. 1f). Both vessel types are constructed from CD31+/VEGFR2+ ECs, suggesting a common point of failure.

Given the strong defects observed in SAMD1 KO embryos, we used public gene transcription data to assess embryonic SAMD1 expression<sup>23,24</sup>. SAMD1 mRNA was detected in very early developmental stages (Fig. 1g) as well as in all investigated organs during all embryonic stages (Fig. 1h). Similar results were also obtained using RT-qPCR experiments from isolated embryonic organs (Supplementary Fig. S1). This observation suggests that SAMD1 has a biological function during all embryonic stages and in most embryonic tissues.

**SAMD1 KO embryos have multiple organ defects.** To further assess the consequence of the SAMD1 KO on specific organs, we first used hematoxylin and eosin (H&E) staining on sagittal slices of an E14.5 embryo (Fig. 2). Hematoxylin reveals not only nuclei, but also glycosaminoglycans, for example in cell walls, and thus allows visualization of fragmenting cells, while eosin stains cytoplasm and most connective tissue pink, orange, and/or red<sup>25</sup>. A comparison of embryonic tissues and organs in WT, HET and KO mice revealed several abnormalities in the KO mice, while the HET embryo organs were indistinguishable from the WT organs (Fig. 2, Supplementary Fig. S2). Although cardiac muscle cells in the KO embryo stained red and had associated nuclei, the



**Figure 2.** SAMD1 KO embryos exhibit organ degradation. (a) H&E staining of the heart at E14.5. *H* heart muscle, *A* aorta, *PT* pulmonary trunk, *Lu* lung, *Li* liver, *RA* right atrium, *OT* outflow tract of left ventricle, *VC* vena cava, *LV* left ventricle, *RV* right ventricle. (b) H&E staining of the lungs at E14.5. *B* bronchiole, *V* blood vessel, *S* skeletal muscle. (c) H&E staining of the liver at E14.5. *V* blood vessel, *S* skeletal muscle.

heart appeared to be breaking up, and the atria, ventricles, and pulmonary trunk had partially collapsed (Fig. 2a, right panel). The endocardium was separating from the chamber walls. Red-stained RBCs were absent, and the ventricles appeared to contain some lymphocytes/immune cells instead of RBCs (Fig. 2a, right panel). This suggests RBCs were trapped due to ejection failure and became necrotic.

The KO lung also appeared to be fragmenting, and RBCs were not apparent, in contrast to the WT lung where RBCs filled large vessels (Fig. 2b). Bronchioles had formed but were smaller than those in the WT. They appeared to be degrading and may not have formed correctly. Faint pink staining of connective tissue in the lung and diaphragm may indicate dying cells.

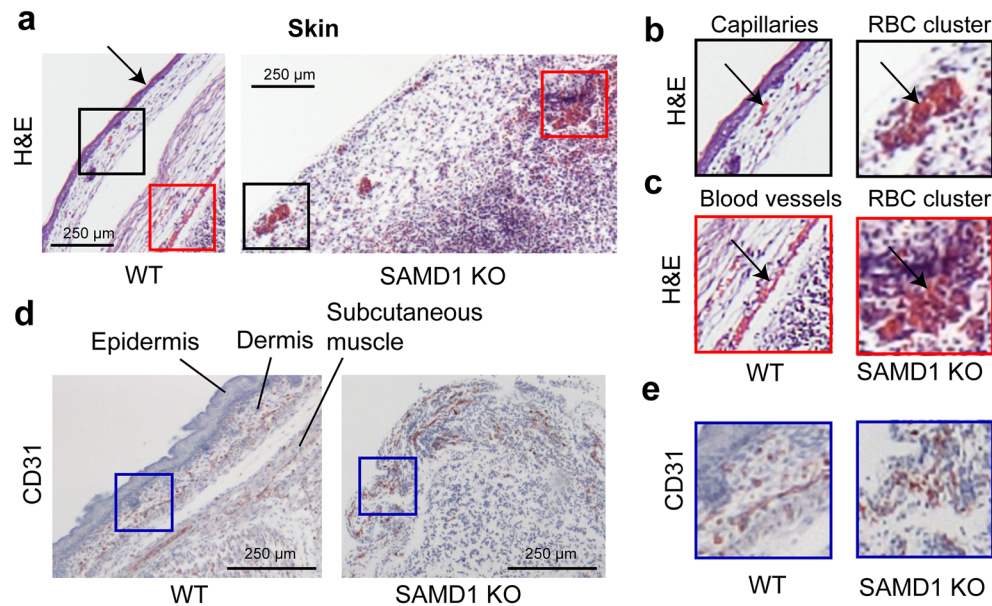
The KO liver had lobes and was large enough to have developed at least until E12.5 (Fig. 2c). Similar to the heart, the liver had no RBCs, again in contrast to the WT, where RBCs filled large vessels. The tissue was broken up in patterns that suggested degradation rather than a malformation. A possible portal triad is visible, but other features were not readily identifiable. The fact that organs were substantially developed means that a functional circulatory system necessarily existed prior to E14.5.

**SAMD1 deletion led to disorganization and reduced levels of RBCs in the skin.** The most obvious difference between the appearances of WT and KO embryos was the paleness of the latter, suggesting impaired blood circulation. Thus, we tested an E14.5 embryo for the presence of endothelial cells (ECs), which line all blood and lymph vessels, from primitive endothelial tubes to fully mature vessels. We used H&E, CD31 or VEGFR2 staining (with hematoxylin counterstain) to mark ECs and provide morphological identification of other cells. Cluster of differentiation 31 (CD31, also known as PECAM1) is widely used as a marker for ECs, staining the cell surface. Vascular endothelial growth factor 2 (VEGFR2, also known as FLK1) stains the cytoplasm of ECs. Using these different markers, we investigated the organs and tissues of the SAMD1 KO embryos.

Looking first at the embryo surface, H&E staining of the WT embryo showed groups of RBCs in the skin and RBCs in a larger blood vessel beneath the subcutaneous tissue. Here, CD31 stained ECs of capillaries and larger vessels (Fig. 3a–c), as expected. In contrast, KO skin, except around the limbs and skull, appeared to have been lost during fixation (Fig. 3a), likely due to edema separating the skin from the tissue beneath. In locations beneath the lost skin, the KO embryonic surface appeared to consist of degrading and fragmenting skeletal muscle (Fig. 3a,d). This tissue included individual RBCs and a few scattered clusters of RBCs that, lacking typical organizational patterns, did not appear to be contained in vessels (Fig. 3a–c). Coagulation in the clusters could not be determined, but gradations in red staining implied the recent onset of RBC degradation. These extravascular RBCs could be the result of fresh microhemorrhages or RBCs recently stranded as vessel development failure halted circulation. Consistent with this second hypothesis, intact vessels were not obvious, but RBCs could not have arrived at their present locations unless at least functional endothelial tubes had once been present. CD31 stained numerous misshapen subcutaneous ECs in broken brown lines (Fig. 3e). These patterns suggested that the ECs were part of capillaries or larger vessels that had degraded. Cytoplasmic instead of cell surface CD31 brown staining patterns suggest necrosis, and the association of CD31 staining with rounded nuclei suggests possible phagocytosis of ECs. Several broken small circles of stained ECs also suggested failed vessels (Fig. 3d,e).

Edema did not separate the skin from the KO forepaws, which appeared to have developed normally except for being smaller, pale, and lacking intact blood vessels (Supplementary Fig. S3). H&E staining of the forepaw showed a small group of faintly stained RBCs, where faint blurred hematoxylin stain suggested impending necrosis. CD31 staining showed a few nuclei partially surrounded by brown-stained EC material, again suggesting ingestion. The patterns of broken lines suggest the presence of failed small vessels in the forepaw (Supplementary Fig. S3,





**Figure 3.** E14.5 SAMD1 KO mice have clusters of RBCs and disorganized and damaged endothelial cells (ECs) in the skin. **(a)** H&E staining of embryo skin. **(b)** Higher magnification showing capillaries. Arrows indicate capillary (WT) or RBC clusters (KO). **(c)** Higher magnification of blood vessels (WT) or RBC clusters (KO) are indicated by arrows. **(d)** CD31 (brown) staining of embryo skin **(e)** Higher magnification of CD31 ECs in capillaries in WT skin and disordered structures in KO skin.

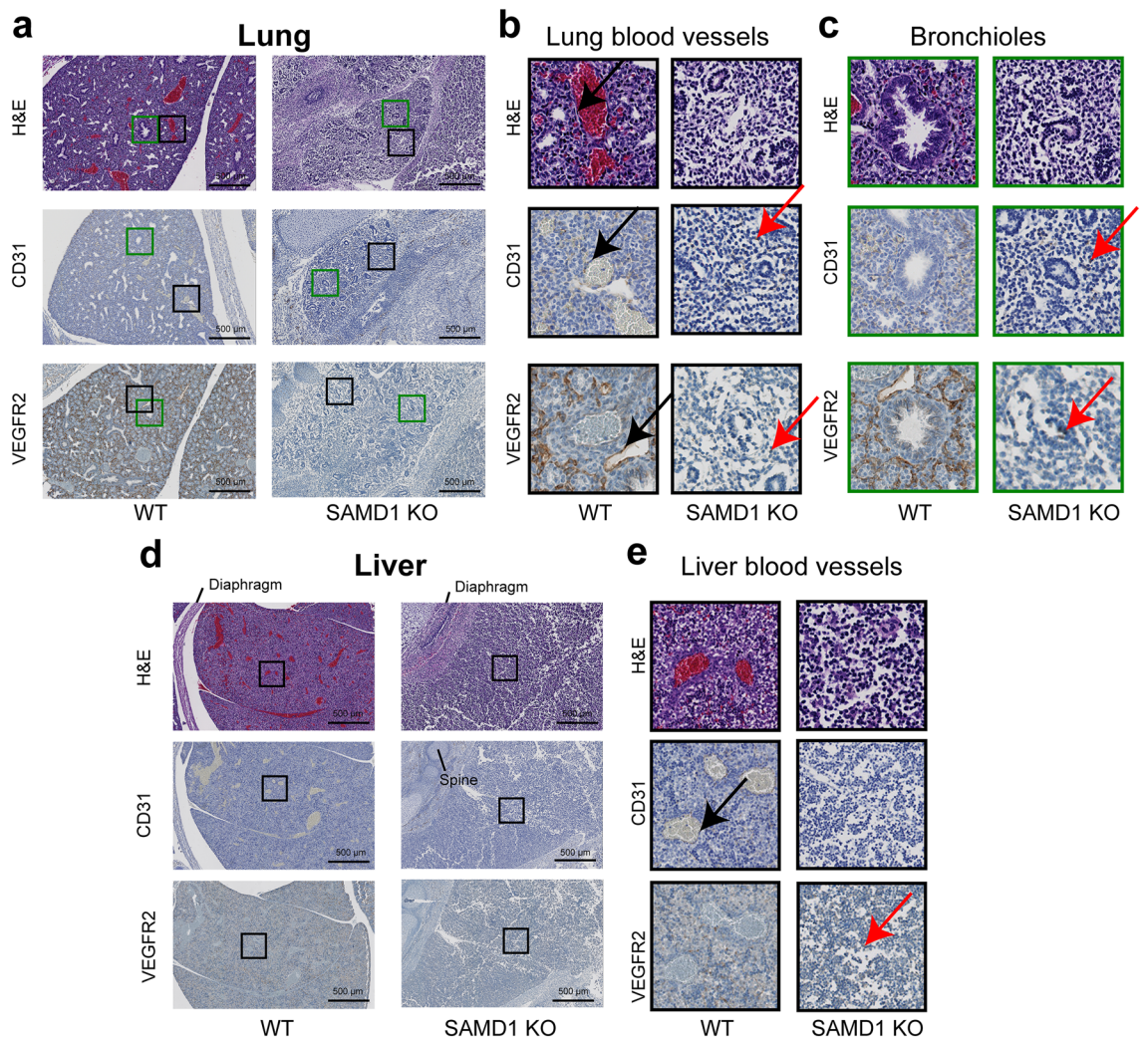
arrows). Interestingly, red-stained RBCs were noted primarily in skeletal muscle near the embryo surface but not in the heart or other organs. Surface perfusion may have been sufficient to delay RBC necrosis compared to locations deeper in the embryo.

**SAMD1 KO embryos have degraded internal organs and blood vessels.** Our initial analysis (Fig. 2) suggests that internal organs are fragmented and that they lack proper blood vessels. To further assess the impact of SAMD1 deletion on blood vessels in these internal organs, we stained lung, liver (Fig. 4), and heart samples (Fig. 5) from E14.5 WT and KO mice with H&E, CD31, or VEGFR2. In all WT samples, H&E staining showed normal organ development and large blood vessels filled with RBCs (Figs. 4, 5). Larger lung and liver vessels, identified by sizeable areas of RBCs, were encircled by stained ECs (Fig. 4b,d). Capillaries were also made evident by the presence of EC markers and RBCs. Pulmonary vascular ECs normally line the surfaces of the lung vasculature, which stained strongly for CD31 and VEGFR2, and epithelial cells in bronchioles stained faintly for VEGFR2 (Fig. 4c). Liver sinusoidal ECs also stained for CD31 and VEGFR2 (Fig. 4e).

In contrast, in the lungs and livers of the SAMD1 KO mice, CD31 and VEGFR2 staining was almost absent. In KO organs, no RBCs or intact ECs were visible (Fig. 4a,b,d,e). CD31 and VEGFR2 stained vessels that should have surrounded alveoli were absent (Fig. 4c). The limited CD31 and VEGFR2 staining in the lung and liver appeared to be cell fragments of necrotic vessel wall ECs, and the thin lines of associated faint hematoxylin stain likely mark dermatan sulfate and heparan sulfate from these ECs. Examples of probable phagocytosis in the liver were indicated by uneven brown staining around strongly blue-stained round nuclei (Fig. 4b,c,e, red arrow, Supplementary Fig. S4a). This implies that ECs had been present, became necrotic and were being removed. Bronchioles had formed in the KO lungs but were much smaller than those in the WT lungs (Fig. 4c). Based on the available data, it is not possible to judge whether the smaller bronchioles in the KO were due to incomplete formation during development or whether they had degraded (Fig. 4c). Malformed and/or underdeveloped liver lobules and tubes appear to be degrading (Fig. 4d).

In the WT hearts, normal CD31 and VEGFR2 staining patterns were noted in ECs of capillaries and larger vessels (Fig. 5a–c). H&E staining of the WT pulmonary trunk showed numerous RBCs contained in the vessel, demonstrating normal heart function. VEGFR staining also identified ECs at the lumen and several layers of circumferentially oriented spindle-shaped vascular smooth muscle cells (VSMCs) (Fig. 5d). As expected, epithelial cells in the endocardium stained for CD31 and VEGFR2 (Fig. 5e).

In the KO heart, H&E staining uncovered signs of epithelial, cardiac muscle cell, and EC degradation (Figs. 2a, 5b). Hematoxylin counterstain in the CD31 and VEGFR2 slides revealed fragmenting cardiac muscle cells (Fig. 5b). In a few places, round hematoxylin-blue nuclei were loosely associated with faint CD31 or VEGFR2-brown patterns demonstrating the presence of necrotic ECs and epithelial cells (Fig. 5c). Phagocytosis in the heart was indicated by uneven and foamy brown staining around strongly blue-stained round nuclei (Supplementary Fig. S4b). CD31 and VEGFR2 stained broken circles and lines of misshapen ECs and EC fragments. These patterns suggested incomplete and/or failed vessels (Fig. 5b,c). This was seen deep in the heart muscle where capillaries should have been (Fig. 5b), and more frequently at the heart's surface, where larger stained circles suggested

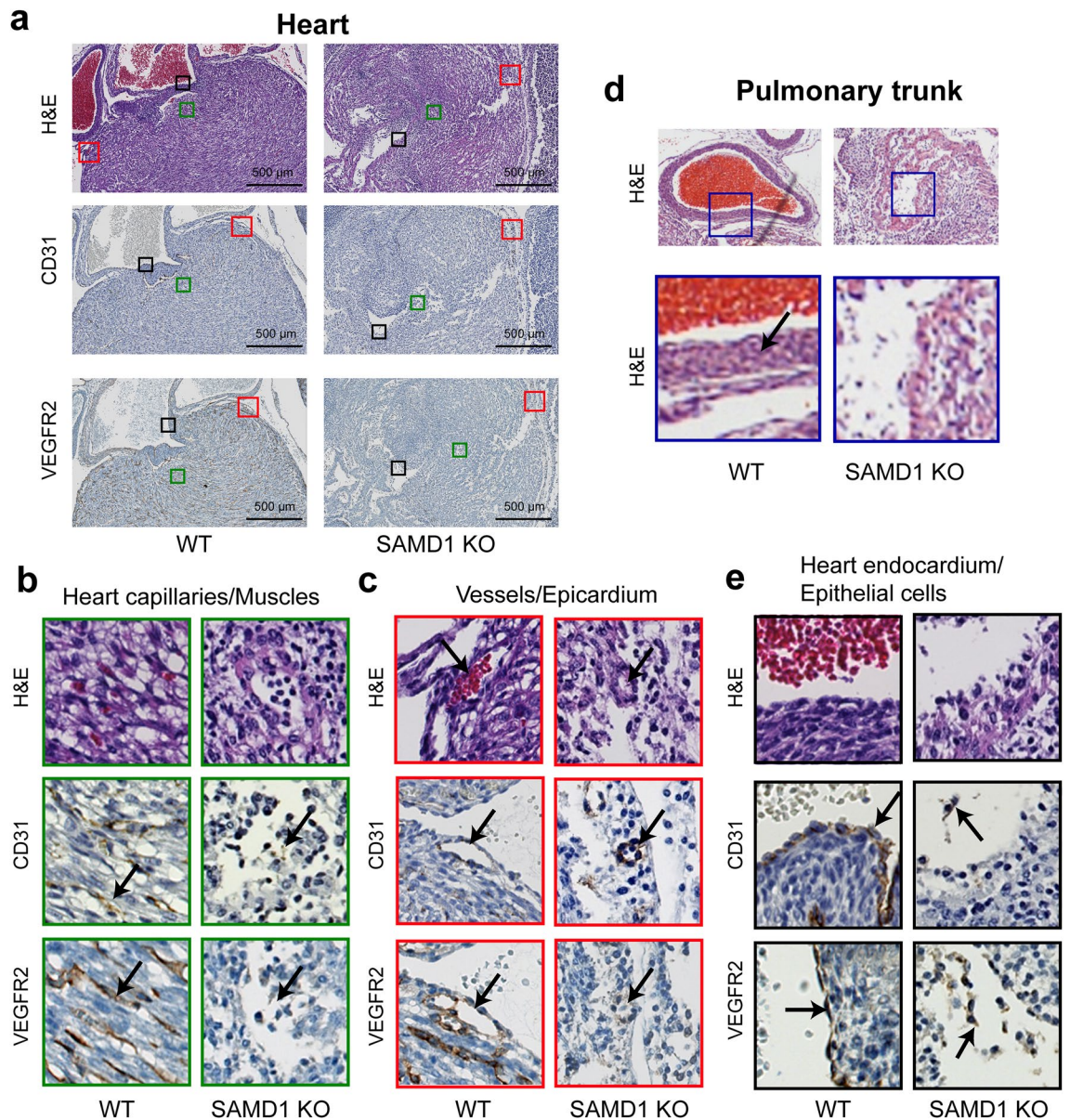


**Figure 4.** Absence of blood vessels in the lung and liver of SAMD1 KO embryos. (a) Macroscale microscopy of the lungs from wild-type and SAMD1 KO mice at E14.5 using H&E, CD31 (brown), and VEGFR2 (brown) staining. (b) Higher magnification of lung wild-type blood vessels; the black arrows point to spindle-shaped vascular smooth muscle cells (VSMCs) adjacent to ECs, that are not present in the KO lung. (c) Higher magnification of bronchioles, which are smaller and/or degraded in the KO mice. The lung appears to be fragmenting along the paths of failed vessels. (d) Macroscale microscopy of the lungs from wild-type and SAMD1 KO mice at E14.5 using H&E, CD31, and VEGFR2 staining. (e) Higher magnification of blood vessels in the wild-type liver, and their absence in the KO. The KO liver also appears to be fragmenting. Red arrows in b, c, and e indicate possible fragments of phagocytized cells.

failed coronary vessels (Fig. 5c). Spindle-shaped VSMCs (blue) that were noted next to coronary artery ECs in the WT were absent from the KO (Fig. 5e). The KO pulmonary trunk wall consisted only of the myocardium, with a few disconnected ECs above the lumen, and lacked the VSMCs that were seen in the WT (Fig. 5a,d). The lumen of the KO pulmonary trunk contained no RBCs and only a few rounded probable immune cells (Fig. 5d). The heart chambers and large vessels were partially collapsed, possibly due to serum fluid being insufficient to maintain shape (Figs. 2a, 5a–c). The ventricles contained dark blue stained rounded probable immune cells and cell fragments that may have been RBCs, but trabeculation was not observed. Epithelial and endothelial cells are closely linked. During embryonic development, the epithelial layer is the source of endothelial cells for the heart's blood vessels<sup>26</sup>. CD31- and VEGFR2-stained epithelial cell fragments were detectable in the KO endocardium, including detached cells and cells undergoing probable necrosis and phagocytosis (Fig. 5e). These above findings suggest that in the investigated SAMD1 KO mouse the development of the heart may have stopped before VSMC differentiation.

Together our analysis of internal organs using markers for ECs suggests that the blood vessel system was degrading and non-functional at E14.5, leading to organ degradation.

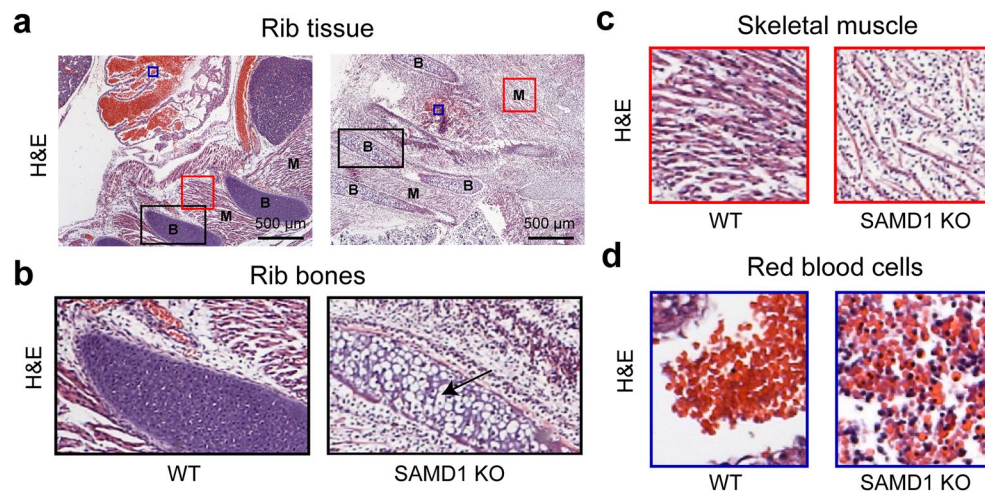
**SAMD1 KO ribs showed premature ossification.** Next, we investigated the ribs and associated skeletal muscles. Structural organization in the E14.5 WT and KO mice were roughly similar (Fig. 6a), but H&E staining



**Figure 5.** SAMD1 KO mouse hearts are fragmenting, capillaries are missing, and larger vessels and the epithelial layer are failing. **(a)** Microscopy of the heart using H&E staining as well as staining for CD31 and VEGFR2 at E14.5. **(b)** Higher magnification comparing WT with myocardium capillaries to KO missing myocardium capillaries. The arrows indicate intact and degraded ECs; note degraded myocardium. **(c)** Higher magnification of an intact WT compared to a degrading KO coronary vessel beneath the epicardium. The arrows indicate intact and degraded coronary vessel ECs. **(d)** H&E staining of the pulmonary trunk from WT and SAMD1 KO mice at E14.5 (derived from same embryo as in Fig. 2a). Instead of spindle-shaped cells (blue), marked with an arrow in the WT, a compact myocardium lines the SAMD1 KO pulmonary trunk. **(e)** Higher magnification of heart endocardium and epithelial cells. The arrows indicate intact (WT) and degraded (KO) epithelial cells (brown).

highlighted substantial differences. The WT rib consisted of cartilage primordium, while in the KO mouse, the rib was ossifying, as seen by the bubbly appearance of hypertrophic chondrocytes, some of which lacked nuclei (Fig. 6b, arrow). Bone development is tightly linked to hypoxia<sup>27</sup>, which is required for the transition from proliferating cartilage to endochondral ossification. Hypoxia interior to the developing bone causes chondrocytes to become hypertrophic and delays chondrocyte apoptosis/necrosis. Subsequent apoptosis/necrosis is required for ossification<sup>27</sup>. Chondrocytes in a vertebrae stain for VEGFR2 (Supplementary Fig. S4c), a receptor for VEGF, which is a chondrocyte survival factor necessary for bone formation<sup>28</sup>. Thus, we hypothesize that in the SAMD1 KO mice, the hypertrophic chondrocytes appeared to have become hypoxic before E14.5, instead of the normal condition at approximately E18, possibly due to the lack of RBC delivery.

Skeletal muscle cells in the WT mouse stained strongly with H&E and had associated blue nuclei (Fig. 6c). Strands of cells are separated by lines of unstained connective tissue (Fig. 6c). In the KO, there are several signs



**Figure 6.** The absence of SAMD1 led to early bone ossification, muscle degradation, and the enduring presence of nucleated RBCs. **(a)** Macroscale microscopy of embryonic rib tissue using H&E staining at E14.5. *B* bone, *M* muscle. **(b)** Higher magnification of the rib bones, showing cartilage primordium (purple) in the wild-type mice and ossification (bubbly, indicated by the arrow) in the KO mice. **(c)** Higher magnification of the skeletal muscles shows separated and fainter muscle fibers (pink) in the KO and more unstained tissue. **(d)** Higher magnification of H&E staining of red blood cell pools from WT and SAMD1 KO embryos at E14.5 shows nucleated RBCs and immune cells in the KO.

of degradation. The KO has fewer eosin-stained red strands, which are less strongly stained red and are more widely separated from each other than in the WT (Fig. 6c). The strands are also interspersed with many nuclei. Some of these nuclei may be immune cells that are attracted to necrotizing muscle cells. Hypoxia from a lack of delivery of RBCs is the likely cause of muscle cell necrosis.

Most RBCs (bright orange) in a blood vessel above a WT rib do not have nuclei, and few if any lymphocytes are seen (Fig. 6d). RBCs are absent from the KO, other than a few scattered pools near the embryo surface and near ribs (Fig. 6a). A small blood pool in the skeletal muscle above a KO rib contains some RBCs with eccentric nuclei and many lymphocytes/immune cells (rounded, blue) (Fig. 6d). Skeletal muscle cells immediately adjacent to this pool appears to be less degraded. The initiation of heartbeat around E8.5 marks the onset of embryo-vitelline circulation<sup>29</sup>, as yolk sac-derived hematopoietic cells are spread through the developing embryo<sup>30</sup>. These primitive erythroid (EryP) cells have nuclei, and begin to mix with definitive RBCs from approximately E12.5, and enucleating EryPs at E12.5<sup>31</sup>. A mix of nucleated and definitive RBCs remains in circulation until gestation. Obviously, nucleated and enucleated RBCs could not have arrived at these locations unless a functional circulatory system existed later than 12.5. This observation suggests circulation failure trapped EryPs and definitive RBCs in a nonfunctioning vessel around E13.5, when approximately 50% of RBCs would not have nuclei<sup>31</sup>. Some KO RBCs appeared faded (Fig. 6d), implying a recent loss of O<sub>2</sub> from hemoglobin, and faint pink staining between RBCs and lymphocytes/immune cells suggests phagocytosis of RBCs.

Taken together, our work so far suggests that the absence of SAMD1 leads to improper development of all investigated organs and tissues in the embryo. We have summarized our findings in Table 1.

**SAMD1 is required for proper neuronal differentiation in vitro.** To better understand the molecular source of the many defects upon SAMD1 deletion, we investigated our previously published RNA-Seq data upon undirected ES cell differentiation in the absence of SAMD1<sup>4</sup>. Although SAMD1 functions mainly as a transcriptional repressor, we found both up and downregulated pathways upon SAMD1 deletion, suggesting direct and indirect consequences of SAMD1 deletion. Thus, it appears that SAMD1 plays a rather pleiotropic role during differentiation processes, consistent with the multifaceted mouse phenotype.

Our previous work showed that direct SAMD1 target genes are commonly associated with brain related pathways<sup>4</sup> and several genes related to brain development, such as *Cbln1*, *Ntrk2*, and *Plxna4*, *Eph4*, become derepressed upon SAMD1 deletion during differentiation (Supplementary Fig. S5a). Consequently, pathways that are linked to brain development, such as synapse assembly (Supplementary Fig. S5b), become predominantly upregulated in SAMD1 KO cells upon differentiation<sup>4</sup>, suggesting that SAMD1 absence may be involved in regulating neuronal differentiation processes.

In contrast, pathways related to angiogenesis were mostly downregulated (Supplementary Fig. S5c). Intriguingly, vascular endothelial growth factor A (*Vefga*), *Pecam1* (CD31), *Thy1* and *Tie1*, which are critical factors for angiogenesis<sup>32–34</sup>, were strongly downregulated in SAMD1 KO cells, both in differentiated and undifferentiated cells (Supplementary Fig. S5d). Furthermore, pathways related to cardiac chamber development were also significantly dysregulated in differentiated SAMD1 KO cells (Supplementary Fig. S5e). This includes key transcription factors such as *Gata4*, *Gata6*, *Zfmp2* (*Fog2*) and *Mef2c*<sup>35</sup> (Supplementary Fig. S5f). Thus, dysregulation of these genes and pathways during embryogenesis in the absence of SAMD1 could potentially contribute to the observed failure of angiogenesis and heart development. However, it is probable that angiogenesis would also fail upon

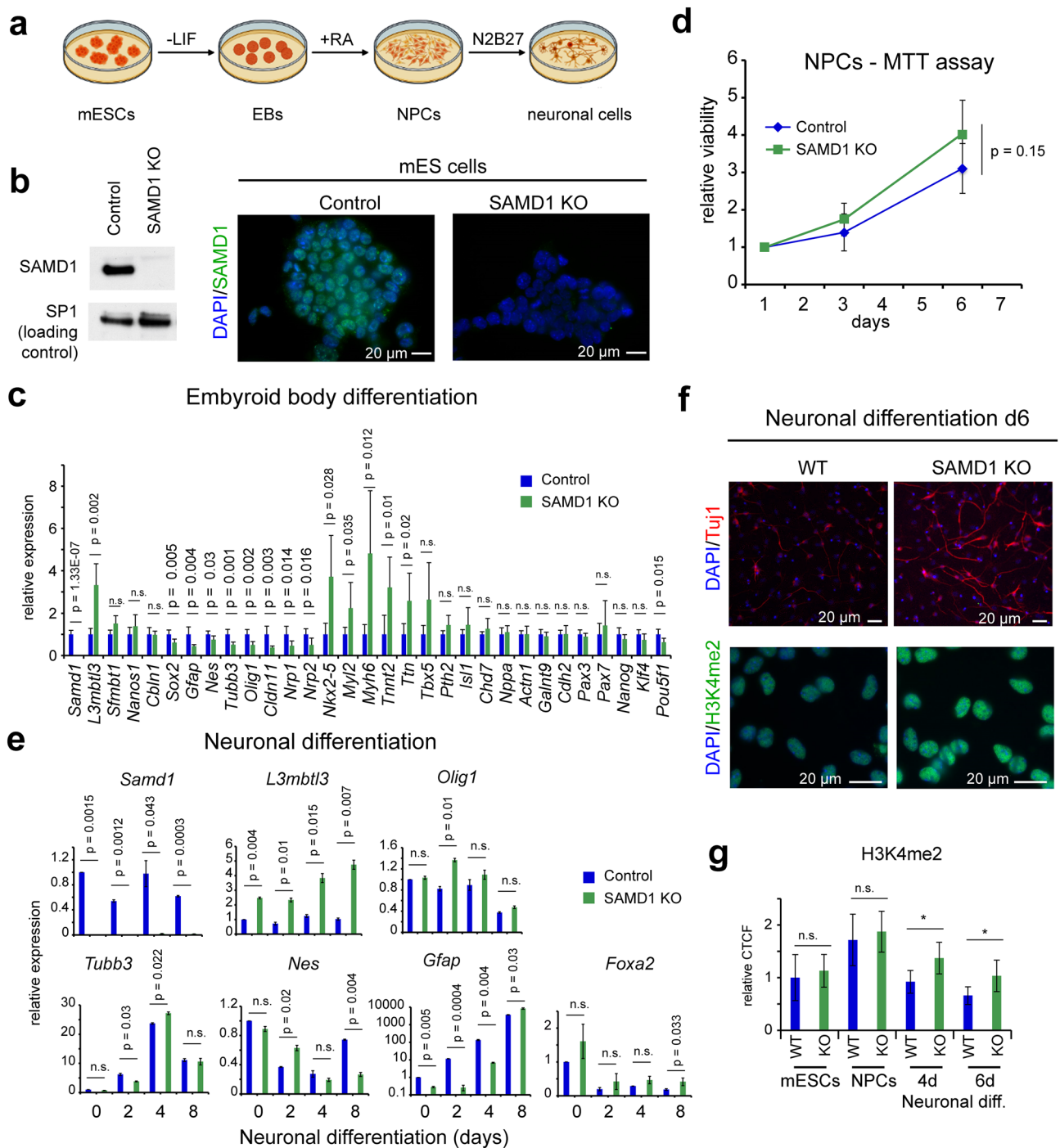
Organ/tissue	Observed phenotype
Whole embryo	
E14.5	Pale Smaller Intact blood vessels not obvious
E15.5	Pale Some embryos show exencephaly, with lack of skull vault Intact blood vessels not obvious Edema fluid beneath skin scattered small blood pools and RBC clusters
Extra-embryonic (E14.5)	
Placenta	Pale
Yolk sac	Lacks blood vessels
Amnion	Pink amniotic fluids
Organs (E14.5)	
Blood cells	Extra-vascular, primarily in subcutaneous muscle, and adjacent to bone Pooled and scattered Excess of nucleated RBCs
Heart	Heart chambers collapsing; devoid of RBCs Lymphocytes/immune cells in chambers Muscle cells, epithelial layer, capillary ECs, and larger vessel ECs are degrading Broken lines of CD31+ and VEGFR2+ cells and degrading cells, indicate failed vessels
Liver	Malformed and/or underdeveloped lobules and tubes are degrading RBCs and blood vessels are absent Scattered clusters of CD31+ cells, CD31+ degrading cells, and very faint VEGFR2, indicating degrading blood vessels
Lung	Malformed and/or underdeveloped bronchioles are degrading RBCs and blood vessels are absent Scattered clusters of CD31+ cells and very faint VEGFR2, indicating degrading blood vessels
Skeletal muscle	Sparse muscle fibers, clusters of RBCs Absence of intact capillaries and larger vessels, indicated by broken lines of CD31+ and VEGFR2+ cells and degrading cells
Bone	Premature ossification; VEGFR2+ chondrocytes/osteoblasts Scattered CD31 and VEGFR2 stained material adjacent to bone, indicating degrading blood vessels
Paw	Several scattered RBCs; broken lines of CD31+ cells, indicating failing capillaries

**Table 1.** Summary of defects observed in SAMD1 KO embryos.

the failure of arteriogenesis. Although significant research is ongoing, factors involved in arteriogenesis are still poorly understood compared to angiogenesis<sup>36</sup>.

Next, we wanted to gain more insights into whether ablation of SAMD1 can indeed influence specific differentiation processes. Direct SAMD1 targets are enriched at brain-related genes<sup>4</sup>, and the absence of SAMD1 leads to the upregulation of brain related pathways (Supplementary Fig. S5a,b). In contrast genes of other dysregulated pathways, such as angiogenesis, are not commonly targeted by SAMD1<sup>4</sup>, suggesting a more indirect influence of SAMD1 on these genes. Thus, we reasoned that neuronal pathways are more likely directly regulated by SAMD1, making them attractive for further investigation. Mouse ES cells can be differentiated into various lineages using specific protocols<sup>37</sup>, including neurons (Fig. 7a), offering us a starting point to investigate the role of SAMD1 in this process. Using our previously established SAMD1 KO mouse ES cells (Fig. 7b)<sup>4</sup>, we first created embryoid bodies, which already showed dysregulation of multiple genes (Fig. 7c). The embryoid bodies were differentiated into neuronal progenitor cells (NPCs) and subsequently into neuronal cells (Fig. 7a). We observed no significant differences in the cell growth of the NPCs, although SAMD1 KO NPCs tended to grow slightly faster than wild-type NPCs (Fig. 7d). Upon differentiation of NPCs into neuronal cells, we observed that several marker genes, such as glial fibrillary acidic protein (*Gfap*) and Nestin (*Nes*), were dysregulated upon SAMD1 KO (Fig. 7e). Immunofluorescence of the differentiated cells on day 6 showed an enhanced level of Tuj1 (Tubb3)-stained neurons in the KO cells (Fig. 7f), indicating that SAMD1 KO enhances the differentiation preferentially towards neurons. This observation is consistent with the increased expression of genes related to neuronal pathways during ES cell differentiation<sup>4</sup> (Supplementary Fig. S5a). Furthermore, we observed higher levels of H3K4me2 in the differentiated cells (Fig. 7f,g). This observation supports that SAMD1 deletion may impair the activity of the KDM1A histone demethylase complex during neuronal differentiation, in line with our previous observation in undifferentiated ES cells<sup>4</sup>. However, other indirect effects could also be the source of this observation. Combined, these data support the hypothesis that the absence of SAMD1 leads to aberrant neuronal differentiation and supports that SAMD1 is critical for proper differentiation processes.

**SAMD1 heterozygous mice fail to thrive.** Finally, we assessed the consequence of a heterozygous deletion of SAMD1. H&E staining of various organs of an E14.5 heterozygous (HET) embryo was not noticeably different from that in the WT embryo (Supplementary Fig. S2), and HET mice were born alive. This less severe phenotype of the heterozygous mice, compared to the knockout mice, can likely be explained by a still relatively high expression level of SAMD1 in heterozygous cells (Fig. 1c). However, analysis of the surviving mice showed



**Figure 7.** SAMD1 KO affects neuronal differentiation in vitro. (a) Schematic steps of the neuronal differentiation procedure. The graphic was created on BioRender.com. (b) Representative immunofluorescence and Western blots of wild-type (Control) and SAMD1 KO ES cells, which were established previously<sup>4</sup>. Full Western blots are shown in Supplementary Fig. S6. (c) RT-qPCR of differentiation-related genes after embryoid body formation. Data are presented as the mean  $\pm$  SD of two biological replicates. (d) MTT assay of neuronal progenitor cells. Data are presented as the mean  $\pm$  SD of three biological replicates. (e) RT-qPCR of marker genes upon neuronal differentiation, starting from NPCs. Data are presented as mean  $\pm$  SD of two biological replicates. (f) Representative immunofluorescence for Tuj1 (Tubb3) and H3K4me2 in differentiated neuronal cells. (g) Quantification of H3K4me2 levels at distinct stages during the neuronal differentiation procedure. CTCF = corrected total cell fluorescence. Data are presented as mean  $\pm$  SD of at least 15 cells per condition. Significance in (c, d, e, g) was evaluated via a two-tailed unpaired Student's t-test. n.s. not significant.

that approximately 30% of the HETs failed to survive past P21 (Supplementary Tables S2, S3), suggesting that the HET mice have distinct phenotypes.

To gain a clearer picture of why some of the postnatal HET mice died, we analyzed key body measures of the surviving mice. Longitudinal body weight analysis on regular chow revealed that male and female HET mice weighed less than WT mice (Fig. 8a), suggesting a reduced ability to thrive. A high-fat diet (HFD) is commonly used to induce metabolic changes in mice. HFD (60% fat by kcal, vegetable shortening) was started at 4 weeks and continued for 15 weeks. We found that after HFD feeding, HETs weighed less than control WT mice (Fig. 8b), and fat depots from HET mice weighed approximately half as much as those from the WT mice (Fig. 8c). Consistent with what appeared to be reduced adiposity in the SAMD1 HET mice on the HFD, glucose disposal was increased in the HET mice during the oral glucose tolerance test (OGTT) (Fig. 8d) and baseline (time 0) fasted glucose levels were significantly lower in the HET mice following the HFD challenge (Fig. 8e). Insulin measurements from the blood samples collected during the OGTT found that insulin levels in the HET mice did not increase following the HFD (Fig. 8f) and were significantly lower than the corresponding levels in the WT mice (Fig. 8g).

Serum VLDL (very low-density lipoprotein) and LDL (low-density lipoprotein) were higher in the HETs, while HDL (high-density lipoprotein) was lower in the HETs (Fig. 8h). Proper steroidogenesis is essential for correct lipid metabolism<sup>38</sup>. Indeed, steroid hormone levels appeared to be altered in adult HETs. Corticosterone, aldosterone, and angiotensin II levels were significantly increased, and testosterone levels trended higher (Fig. 8i). Steroidogenesis and transformation of cholesterol to oxysterols occurs in many cell types<sup>39</sup>. It is possible that the approximately 10–15% reduction in SAMD1 expression in SAMD1 HET mice (Fig. 1c) caused or required a compensatory change in steroid expression. The lower weights and adiposity may also be related to steroidal differences. Notably, we measured large differences in the above markers between individual adult HETs. It is likely that the HETs dying prior to postnatal week 3 expressed a more severe phenotype, resulting in a failure to thrive. We did not observe any, probably because they are typically quickly eaten by the dams.

Collectively, these data suggest that although the heterozygous deletion of SAMD1 has no substantial consequences on embryogenesis, it still has an impact on the ability of the mice to develop normally after birth.

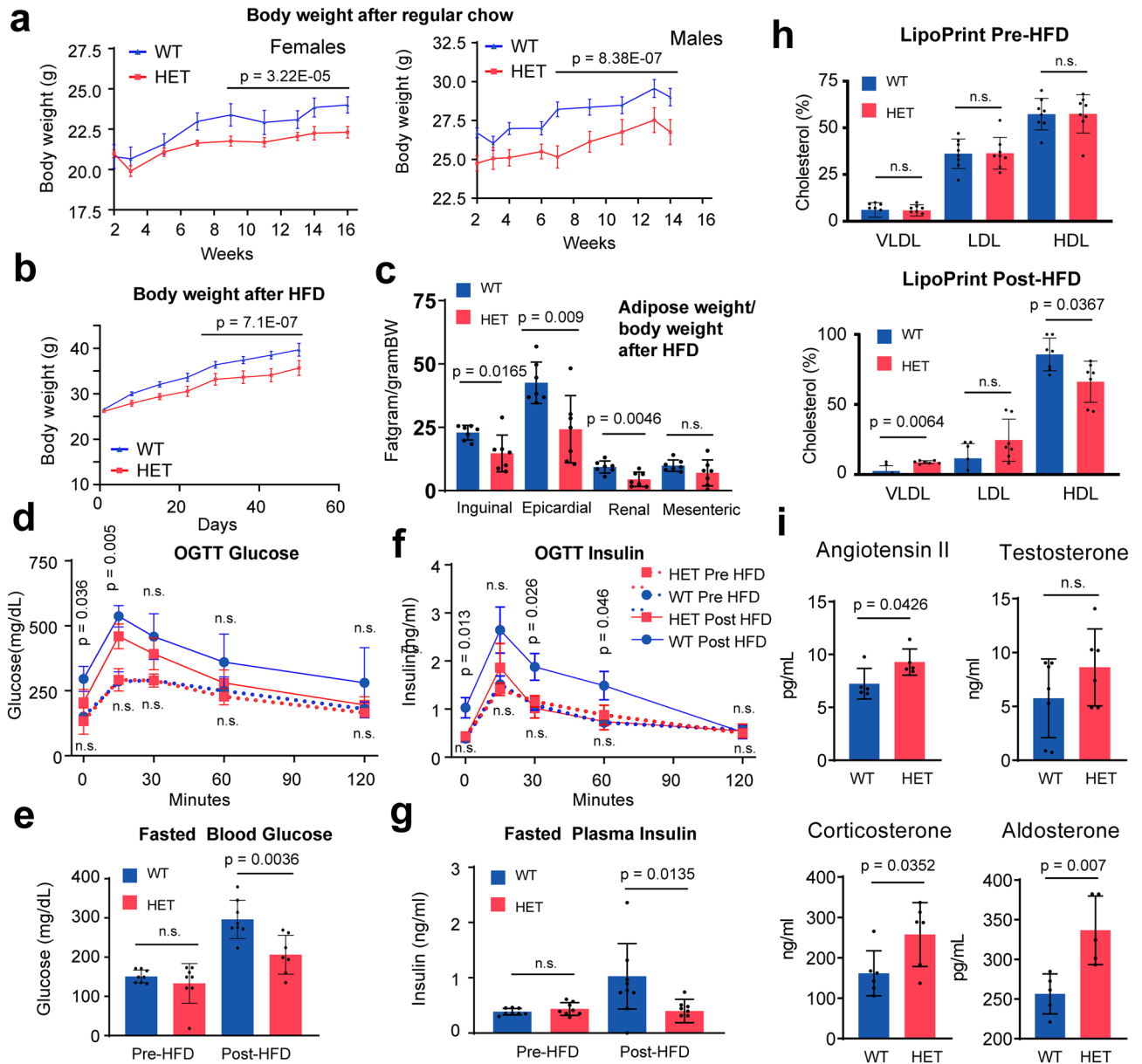
## Discussion

SAMD1 is an epigenetic regulator that plays a repressive gene regulatory role at unmethylated CpG islands<sup>4,9</sup>. Upon mouse ES cell differentiation, the absence of SAMD1 enhances the expression of genes involved in neuronal pathways, while factors important for metabolism and angiogenesis are downregulated (Supplementary Fig. S5)<sup>4</sup>, likely due to both direct and indirect effects. In HepG2 liver cancer cells, SAMD1 also strongly impacts the transcriptional network<sup>9</sup>. In addition to gene regulatory functions SAMD1 may also have a role in atherosclerosis<sup>1,3</sup> and muscle adaptation<sup>15</sup>. In this study, we investigated the effects of ablation of the SAMD1 gene in mice.

Looking at the whole organism, we noted that without SAMD1 there was an absence of functional blood vessels leading to a pale appearance of the embryo by E14.5 (Fig. 1e). Similarly, RBCs, and the lack thereof, in the E14.5 KO yolk sac, placenta, and umbilical cord seem to mark failed vessels, resembling those seen in the embryo. This should not be surprising, since vessel development is similar in embryonic and extraembryonic tissue<sup>40–42</sup>. Given that by approximately E12.5 the definitive placenta is mature<sup>43</sup> and the E12.5 embryos and extra-embryonic tissue appeared grossly normal, we conclude that embryonic lethality of the SAMD1 KO mice is unlikely to have placental defects as a primary cause. We hypothesize therefore that lack of SAMD1 causes roughly simultaneous failure of blood vessels in the embryo, the yolk sac and the placental vasculature.

Nonetheless, SAMD1 KO mice, including the extraembryonic tissues, necessarily had developed a vascular plexus, at least to the point of EC tubes. This is shown by the extent of organ development and the presence of apoptotic EC fragments in locations where vasculature would normally have been. Additionally, the observed RBCs and RBC clusters could not have arrived at the observed locations without a circulatory system to transport RBCs. The sparse mix of nucleated and enucleated RBCs in the KO demonstrated that early hematopoiesis, functional heartbeat, and vasculogenesis had occurred (Figs. 1f, 3a, 6)<sup>44,45</sup>. These findings suggest that in the SAMD1 KO mice at least partially functional ECs differentiated, proliferated, connected, and organized into a functional vascular plexus during the early stage of embryogenesis. However, the E14.5 KO no longer has functional vasculature. At this stage, staining for EC markers appeared to mark only necrotic cells (Figs. 3, 4, 5), and RBCs were seen only in small clusters and scattered broken lines suggesting that previously functional blood vessels had degraded. Failure of arteriogenesis and subsequent pruning<sup>20</sup> seems a possible explanation for the termination of circulation in the SAMD1 KO mice at this time point, since that such a failure would likely also lead to collapse of angiogenesis<sup>19,20</sup>. Although significant research is ongoing, factors involved in arteriogenesis are still poorly understood compared to angiogenesis<sup>36,46</sup>, and it remains to be determined by which mechanisms absence of SAMD1 may impair vessel maturation.

The absence of functional vasculature in the entire embryo at this time point likely caused severe hypoxia and thus cessation of heartbeat. Prior to initiation of maternal-embryo circulation, nutrients and sufficient oxygen to metabolize glucose for growth is available in the uterus. Normal heartbeat starts by E8.5, providing circulating nutrients and RBCs for oxygen<sup>47</sup>. Cessation of circulation at any time after approximately E9 would cause hypoxia such that the heart could no longer function, and resorption would begin. The SAMD1 KO heart, liver, and lung were the correct size at about E12.5–E13.5. Obviously, organs could not have developed to this size without a functional circulatory system to provide nutrients and oxygen. At E14.5, we found only fragmented vasculogenesis-level blood vessels. RBCs were absent from most locations, and pooled in a few places near the embryo surface, meaning that circulation had ceased, causing severe hypoxia and thus, cessation of the heartbeat. The observed tissue degradation means that resorption was occurring, which is very rapid after cessation of heartbeat<sup>48</sup> suggesting that the death in the E14.5 mouse occurred at approximately E13.5.



**Figure 8.** SAMD1 heterozygous (HET) mice gain less weight and have reduced adiposity, higher hormone levels, and different serum cholesterol levels. (a) Weight gain of WT and HET mice fed regular chow. Data are presented as the mean  $\pm$  SE of at least 7 mice per condition. Significance was evaluated by a two-tailed unpaired Student's t-test for the last 5 time points combined. (b) Weight gain of WT and HET mice fed regularly with high-fat diet (HFD). Data are presented as mean  $\pm$  SE of at least 7 mice per condition. Significance was evaluated by a two-tailed unpaired student's t-test. (c) Adipose weight upon HFD ( $n = 7$ ). Data are presented as the mean  $\pm$  SD of 7 mice per condition. P-value via two-tailed unpaired Student's t-test. (d) Blood glucose concentration during oral glucose tolerance test, pre- or post-HFD. Data are presented as mean  $\pm$  SE of at least 7 mice per condition. P-values via two-tailed unpaired student's t-test. (e) Fasted Blood Glucose level pre- or post-HFD. Data are presented as the mean  $\pm$  SD of at least 7 mice per condition. Significance was evaluated by a two-tailed unpaired Student's t-test. (f) Serum insulin during the oral glucose tolerance test, pre- or post-HFD. Data are presented as the mean  $\pm$  SE of at least 7 mice per condition. Significance was evaluated by a two-tailed unpaired Student's t-test. (g) Fasted plasma insulin pre- or post-HFD. Data are presented as the mean  $\pm$  SD of at least 7 mice per condition. Significance was evaluated by a two-tailed unpaired Student's t-test. (h) Serum cholesterol levels before and after an HFD. Data are presented as mean  $\pm$  SD of 7 mice per condition. P values via two-tailed Mann-Whitney U test. (i) Levels of metabolic hormones in WT and HET mice. Data are presented as mean  $\pm$  SD of at least 5 mice per condition. P-values via two-tailed unpaired Student's t-test. n.s. = not significant.



Taken together, we hypothesize that the death of the SAMD1 KO embryos is likely driven by several successive events. Possibly, arteriogenesis failed first, leading to failed vessel maturation<sup>19</sup>, vessel pruning<sup>20</sup>, and thus failing circulation. It is unlikely that this occurred simultaneously throughout the embryo, but at some point, it would prevent sufficient oxygen supply to the heart, leading to cessation of the heartbeat, an event that marks death and initiates resorption followed by the observed degradation of the embryos. Circulation failure caused insufficient oxygen and nutrient supply to internal organs, likely affecting their development. Secondary events, such as bone ossification are possibly driven by the hypoxic conditions during the failure of the circulation. Notably, the observed SAMD1 KO phenotype is markedly different from reported deletions of regulators of vasculogenesis, angiogenesis, and lymphatic development<sup>47</sup>. Thus, SAMD1 may modulate vessel formation in a unique fashion, which requires further clarification.

The presence of SAMD1 mRNA in the entire WT embryo and yolk sack (Fig. 1g,h, Supplementary Fig. S1) suggests that SAMD1 is relevant beyond vessel development. Thus, some SAMD1 KO phenotypes may not be related to failure of vascular development. The high expression of SAMD1 in the head, as well as the consequences of SAMD1 deletion on neuronal differentiation processes (Fig. 7), supports a role of SAMD1 during head and neuronal development (Fig. 1f). Given that undirected ES cell differentiation in the absence of SAMD1 influences multiple distinct pathways<sup>4</sup>, it is likely that SAMD1 is involved in additional cellular processes, which awaits further clarification. Thus, experiments that address the role of SAMD1 during other differentiation processes will be important to assess the relevance of SAMD1 for distinct lineages. Generating conditional SAMD1 knockout mice may be a suitable strategy to address the role of SAMD1 during the development of specific tissues and organs. Useful data might also be gathered from studies of SAMD1 KO mice that were embryonic lethal later in development.

Mutations that result in the absence of a gene often do not have uniform effects across a population. The mechanisms can be environmental, epigenetic, synergistic, and multigenic<sup>49</sup>. Variable expression and penetrance are likely explanations for the survival until as late as E18.5, of the few KOs that had heartbeats past E14.5 (Supplementary Table S1), and for the approximately 30% postnatal HET mortality by 3 weeks of age (Supplementary Tables S2 and S3). It is possible that compensating proteins, which typically have domain similarities with a knocked-out protein, allowed at least partial rescue of the developing embryo<sup>16</sup>. The E15.5 KO shown in Fig. 1f may be a somewhat different phenotype from the E14.5. The scarcity of blood vessels is not as extreme, but the absence of skull vault and the exencephaly in the E15.5 KO is different, and might have been caused by an “open” neural tube defect<sup>50</sup>.

Heterozygous SAMD1 KO mice were born alive, suggesting that here the embryogenesis was grossly normal. Likely, the relatively minor reduction of SAMD1 expression in the HET mice (Fig. 1c), possibly achieved due to compensatory effects<sup>16</sup>, is the underlying source for this weak phenotype. However, living HET mice had substantially increased aldosterone, corticosterone, angiotensin II, and testosterone levels (Fig. 8i), suggesting that correct SAMD1 levels are important for proper body function. Steroidogenesis and transformation of cholesterol to oxysterols occurs in many cell types<sup>39</sup>. Thus, it is currently unclear whether SAMD1 directly regulates hormone production pathways in cells or whether these changes are caused by compensatory effects, due to other alterations in body homeostasis. The HETs also showed reduced body weight and adiposity (Fig. 8a–c), which may be linked to steroidal changes.

This investigation aimed to present the first description of the effect of SAMD1 ablation on mouse embryonic development. Based on the E14.5 and E15.5 phenotypes, we focused on features related to vasculogenesis, arteriogenesis and neuronal differentiation. The study is therefore limited to a restricted time window of embryonic development and does not provide deeper insights into the role of SAMD1 during earlier or later time points. Given the use of a limited set of vascular markers and probes, only cells related to differentiated ECs were unambiguously identified. Therefore, it is not determined whether failure of the early vascular plexus is due to failure of migration, proliferation, differentiation, communication, or connection of necessary progenitor cells or of an unknown mechanism. Furthermore, we investigated only a restricted subset of animals and organs for the KO mice and only selected measures for the heterozygous mice. Thus, it is likely that other phenotypes or defects occur in the absence of SAMD1. Additionally, this study does not address the cellular and molecular mechanisms by which the absence of SAMD1 leads to the observed phenotypes. Specifically, it remains to be determined, which phenotypes are directly regulated by SAMD1, and which occur due to indirect effects. Given that SAMD1 is expressed in all tissues and organs (Fig. 1g,h), more research will be necessary to elucidate the role of SAMD1 in specific cellular contexts.

This is the first report on ablation of SAMD1 in mice. We observed embryonic lethality in the total knockout mice, which may be the consequence of the myriad effects stemming from the lack of appropriate vascular development. Given the wide expression of SAMD1 in most organs, numerous other processes may also be affected, potentially explaining the complex phenotype, both in the homozygous and the heterozygous mice.

## Methods

**Animal handling.** Mice were bred, studied, and maintained in the Longwood Medical Research Center facility in accordance with guidelines of the Committee on Animals of the Harvard Medical School and those prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council [DHEW publication No. (NIH) FS-23]. The Pfizer Institutional Animal Care and Use Committee (IACUC) approved all of the animal procedures and protocols according to the criteria stated by the National Academy of Sciences National Research Council (NRC) publication 86-23, 1985. All methods were performed in accordance with the relevant guidelines and regulations, as well as in accordance with the ARRIVE guidelines.

**SAMD1 gene inactivation.** SAMD1 knockout mice were genetically engineered by Pfizer (Groton, CT). To generate the SAMD1 targeting vector, recombineering<sup>51</sup> was used to replace 2396 bp of the mouse SAMD1 (Accession ID D3YXK1) gene encompassing the 3' 615 bp of exon 1, exons 2–4 and all of exon 5 except for the 3' 83 bp with a neomycin phosphotransferase cassette in a 9433 bp genomic subclone obtained from a C57BL/6J BAC (RP23-128H6; Invitrogen). The vector was introduced into Bruce4 (mixed C57Bl/6N:C57Bl/6J) mouse embryonic stem cells<sup>52</sup> using a standard homologous recombination technique. Southern blot analysis was used to identify correctly targeted mouse embryonic stem cell clones. Chimeric mice were generated by injection of the targeted ES cells into Balb/c blastocysts. Chimeric mice were bred with C57Bl/6J mice to produce F1 heterozygotes. Germline transmission was confirmed by PCR analysis (SAMD1g1: 5'-CCAAACCCCTCTTCA GTTCA-3'; SAMD1g2: 5'-GCCGTAGCTATTCTGCCTCA-3'; dNEO2: 5'-ACATAGCGTTGGCTACCCGTG ATA-3'). F1 heterozygous males and females were mated to produce F2 mice. The colony was maintained by intercrossing under specific pathogen-free conditions with unrestricted access to food and water.

**RNA and cDNA measurements for mouse experiments.** RNA and cDNA measurements were performed at the Brigham & Women's Hospital (Boston, MA). Synthesis of cDNA (ThermoScript RT-PCR System (Invitrogen Cat 11146-016)). RT-PCR was performed in a MyiQ single-color real-time PCR system (Bio-Rad Laboratories, Inc., Hercules, CA). Total ribonucleic acid (RNA) from 250,000 cells was reverse transcribed by Superscript II (Invitrogen) according to the manufacturer's instructions. Quantitative PCR was performed with SYBR green PCR mix (primers for SAMD1 from QIAGEN N.V.) and analysis performed with StepOne Software (ThermoFisher, Applied Biosystems). Levels of mRNA were normalized to glyceraldehyde 3-phosphate dehydrogenase (GAPDH) mRNA levels. cDNAs were synthesized using a High-Capacity RNA-to-cDNA™ Kit (Thermo Fisher, Catalog No. 4387406). Total 2 µg RNA/100 µl cDNA synthesis reaction and 1 µl cDNA per TaqMan reaction. GAPDH was used as an endogenous control (Applied Biosystem 4352339E, Catalogue No. 43-523-39E).

**Isolation of mouse embryonic fibroblasts.** Mouse embryonic fibroblasts (MEFs) were isolated as described<sup>53</sup>. In short, embryos from WT, SAMD1<sup>+/-</sup>, and SAMD1<sup>-/-</sup> mice were isolated between E12.5 and E18.5. Internal organs, heads, tails, and limbs were removed from embryos prior to treatment with trypsin to dissociate the cells. Cells were seeded into T-75 cell culture dishes in 15 ml of MEF media: KO DMEM (Life Technologies, 10829-018), 15% FBS (Life Technologies), nonessential amino acids (Life Technologies 111140-050), 2 mM L-glutamine (Life Technologies, 25030-081), 0.1 mM β-mercaptoethanol (Sigma M7522), and 0.25 mg/ml gentamicin (Life Technologies, 15710-072).

**Phenotypic studies.** All behavioral, physical, hormonal, and response phenotype comparisons of SAMD1<sup>+/+</sup> to SAMD1<sup>+/-</sup> were performed by Caliper Life Sciences (Xenogen Biosciences, Cranbury, NJ) based on the Phenotype Pfunder platform<sup>54</sup>.

**Immunohistochemistry.** Antibodies against CD31 and VEGFR2 were purchased from Abcam (Cambridge, UK). Immunohistochemical staining procedures were performed by Pfizer (Groton, CT) and at the Brigham & Women's Hospital (Boston, MA). E14.5 has been established as the optimal time point for mouse developmental disorder phenotype analyses<sup>18</sup>. In short, tissues were collected immediately after E14.5 embryo collection, fixed in 10% buffered formalin and embedded in paraffin. Four micrometer serial sections were cut and mounted on glass slides. One section from each lesion was stained with hematoxylin and eosin (H&E). Other slices were stained for CD31 (ab28364) or VEGFR2 (ab2349) with hematoxylin counterstaining following Abcam's protocols.

**Cell culture.** E14 mESCs (E14TG2a) were provided from the lab of Jacqueline Mermoud (University of Marburg, Germany). SAMD1 KO E14 mESCs were established previously<sup>4</sup>. The WT and SAMD1 KO mESCs were cultured in Dulbecco's modified Eagle's medium (DMEM) and GlutaMAX (Gibco; 61965-026), 15% fetal calf serum (FCS) (Biochrom; S0115, Lot: 1247B), 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 LIF (1000 U/ml; Millipore; ESG1107, lot: 3060038) on 0.2% gelatin-coated plates.

**Neuronal differentiation of mouse ES cells: embryoid body system.** mESCs were plated at a density of 1×10<sup>6</sup> on 0.1% gelatin-coated and dried 100 mm cell culture dishes for 3 days in mESC medium, and the medium was changed every 24 h. To start neurodifferentiation, 1×10<sup>6</sup> dissociated ES cells were incubated on nonadhesive bacterial dishes (100 mm) for 6–8 days in mESC medium without LIF, and the medium was changed every 2 d. Embroid bodies (EBs) were collected and redistributed to 4 wells of a 6-well plate containing 2 ml differentiation medium containing DMEM (Gibco; 11960-044), 20% FCS, 2 mM L-glutamine (Gibco; 25030-081), 1× penicillin/streptomycin, 1× nonessential amino acids, 50 µM β-mercaptoethanol, and 0.5 µM *trans*-retinoic acid (Millipore; 554720) and incubated for 2 days. EBs from two wells each were collected and redistributed to a 100 mm cell culture dish, coated with 20 µg/ml poly-L-ornithine hydrobromide in PBS (Sigma-Aldrich; P3655; 4 h to overnight at room temperature, rinsed twice with water) and incubated for 7 days in neurobasal medium (Gibco; 21103-049) with 1× L-glutamine, 1× penicillin/streptomycin, and 1× B-27 supplement (Gibco; 17504-044)<sup>55</sup>.

The established neural precursor cells (NPCs) were cultured in Euromed-N medium (EuroClone; ECM0883L) supplemented with 1× N-2 (Gibco; 17502-048), 1× L-glutamine, 1× penicillin/streptomycin, 50 µg/ml BSA

(Sigma-Aldrich; A9418), 20 µg/ml insulin (Merck; 11376497,001), 10 ng/ml EGF (Peprotech; 315–09), and 10 ng/ml FGF (Peprotech; 100-18B) on 0.1% gelatin-coated and subsequently 5 µg/ml laminin-coated (Sigma-Aldrich; L2020) six-well tissue culture plates and passaged with Accutase (Sigma-Aldrich; A6964) (adapted from<sup>56</sup>).

To promote the neurogenic capacity of NPCs and allow the formation of mature neurons, WT and SAMD1 KO NPCs were seeded at high density ( $6 \times 10^5$ ) into gelatinized (0.1% in H<sub>2</sub>O) and poly-L-ornithine hydrobromide-coated (20 µg/ml in PBS) six-well tissue culture plates for differentiation with N2B27 medium, consisting of DMEM/F12 (Gibco; 11320-033) and neurobasal medium (1:1), with  $1 \times N-2$ ,  $1 \times B-27$ , 100 µM β-mercaptoethanol and the addition of 10 ng/ml FGF (Peprotech; 100-18B) (adapted from<sup>57</sup>). Analysis of the cells was performed with immunofluorescence staining and RT-qPCR at chosen time points.

**RT-qPCR experiments.** 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 mActb and mGapdh expression. The qPCR primers used are presented in Supplementary Table S4.

**Antibodies for neuronal studies.** The following commercial antibodies were used: H3K4me2 (Diagenode; C15410035), TUJ1/TUBB3 (BioLegend; 801201), and GFAP (Dako; Z0334). The antibody for SAMD1 is custom-made. It is directed against the SAM domain of human SAMD1 (amino acids 452 to 538) and was produced as described previously<sup>4</sup>.

**Proliferation assay.** To determine proliferation rates, cells were seeded on 6-well plates at a density of  $1 \times 10^5$  cells per well. Cell viability was determined at 1, 3, and 6 days after seeding using the MTT assay by adding 90 µl of 5 mg/ml thiazolyl blue  $\geq 98\%$  (Carl Roth; 4022) to each well. The medium was aspirated after 1 h, and stained cells were dissolved in 400 µl of lysis buffer (80% isopropanol, 10% 1 M HCl, 10% Triton X-100). 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.

**Immunofluorescence staining.** WT or SAMD1 KO mESCs were seeded on 0.2% gelatin-coated coverslips. WT or SAMD1 KO differentiated neural precursor and neural cells were seeded on 0.1% gelatin-coated and 20 µg/ml poly-L-ornithine hydrobromide-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 antibodies, using Alexa Fluor 488 goat anti-rabbit IgG (H + L) and Alexa Fluor 546 goat anti-mouse IgG (H + L) (Thermo Fisher Scientific; A-11008 and A-11003), at a 1:2000 dilution each. Following three washing steps, the coverslips were mounted onto microscopy slides using VECTASH-IELD® Antifade Mounting Medium with DAPI (Vector Laboratories; H-1200) and sealed. Microscopy was performed using a Leica DM5500 microscope, and data were analyzed using ImageJ (Fiji).

**Immunofluorescence quantification.** To determine the cellular fluorescence in microscopy images using ImageJ, cells of interest were outlined and measured for their area, integrated intensity and mean gray value. Adjacent background regions were measured for the same parameters to calculate the corrected total cell fluorescence (CTCF) according to the following formula:  $CTCF = \text{integrated density} - (\text{area of selected cell} \times \text{mean gray value of background reading})$ . Average CTCF values and standard deviations for 5 cells each in 3 different fluorescence microscopy images per sample are presented in a graph.

**Bioinformatics analysis.** Gene expression data from mouse embryo tissues were obtained from ENCODE (<https://www.encodeproject.org/>, under “Mouse Development”)<sup>24</sup> and from GSE45719<sup>23</sup>. GSEA was performed with standard settings<sup>58</sup> using previously published data (GSE144396)<sup>4</sup>.

**Statistical analysis.** Quantitative results are expressed as mean  $\pm$  standard deviation of the mean (SD). Deviation from the mendelian ratio was evaluated using Chi Square tests. The significance of differences between two independently measured groups were evaluated using a two-tailed unpaired t-test. Two-way ANOVA followed by Tukey's postdoc test was applied for multiple group comparisons. Significance of the Lipoprint analysis was evaluated using a two-tailed Mann–Whitney *U* test. The p-values were calculated using the GraphPad Prism Software. The p-values of the GSEA analysis were obtained from the GSEA software<sup>58</sup>. A value of  $p < 0.05$  was considered statistically significant.

### Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding authors on reasonable request. Gene expression data analyzed during the study are available from ENCODE<sup>24</sup> (<https://www.encodeproject.org/>, under “Mouse Development”) and from the GEO repository (GSE45719<sup>23</sup> and GSE144396<sup>4</sup>).

Received: 29 September 2022; Accepted: 10 February 2023

Published online: 21 February 2023

## References

- Lees, A. M., Deconinck, A. E., Campbell, B. D. & Lees, R. S. Atherin: A newly identified, lesion-specific, LDL-binding protein in human atherosclerosis. *Atherosclerosis* **182**, 219–230. <https://doi.org/10.1016/j.atherosclerosis.2005.01.041> (2005).
- Campbell, B., Bourassa, P. & Aiello, R. SAMD1 distribution patterns in mouse atherosclerosis models suggest roles in LDL retention, antigen presentation, and cell phenotype modulation. *Biorxiv* <https://doi.org/10.1101/2021.09.12.459413> (2021).
- Tian, S. *et al.* The miR-378c-Samd1 circuit promotes phenotypic modulation of vascular smooth muscle cells and foam cells formation in atherosclerosis lesions. *Sci. Rep.* **11**, 10548. <https://doi.org/10.1038/s41598-021-89981-z> (2021).
- Stielow, B. *et al.* The SAM domain-containing protein 1 (SAMD1) acts as a repressive chromatin regulator at unmethylated CpG islands. *Sci. Adv.* **7**, 2229. <https://doi.org/10.1126/sciadv.abf2229> (2021).
- Bartke, T. *et al.* Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* **143**, 470–484. <https://doi.org/10.1016/j.cell.2010.10.012> (2010).
- Engelen, E. *et al.* Proteins that bind regulatory regions identified by histone modification chromatin immunoprecipitations and mass spectrometry. *Nat. Commun.* **6**, 7155. <https://doi.org/10.1038/ncomms8155> (2015).
- Villaseñor, R. *et al.* ChromID identifies the protein interactome at chromatin marks. *Nat. Biotechnol.* **38**, 728–736. <https://doi.org/10.1038/s41587-020-0434-2> (2020).
- Weber, L. *et al.* The histone acetyltransferase KAT6A is recruited to unmethylated CpG islands via a DNA binding winged helix domain. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkac1188> (2022).
- Simon, C. *et al.* The CpG island-binding protein SAMD1 contributes to an unfavorable gene signature in HepG2 hepatocellular carcinoma cells. *Biology* <https://doi.org/10.3390/biology11040557> (2022).
- Zhang, J. *et al.* SFMBT1 functions with LSD1 to regulate expression of canonical histone genes and chromatin-related factors. *Genes Dev.* **27**, 749–766. <https://doi.org/10.1101/gad.210963.112> (2013).
- Barnes, C. E., English, D. M., Broderick, M., Collins, M. O. & Cowley, S. M. Proximity-dependent biotin identification (BioID) reveals a dynamic LSD1-CoREST interactome during embryonic stem cell differentiation. *Mol. Omics* **18**, 31–44. <https://doi.org/10.1039/d1mo00236h> (2022).
- Malovannaya, A. *et al.* Streamlined analysis schema for high-throughput identification of endogenous protein complexes. *Proc. Natl. Acad. Sci. USA* **107**, 2431–2436. <https://doi.org/10.1073/pnas.0912599106> (2010).
- Stielow, B., Simon, C. & Liefke, R. Making fundamental scientific discoveries by combining information from literature, databases, and computational tools: An example. *Comput. Struct. Biotechnol. J.* **19**, 3027–3033. <https://doi.org/10.1016/j.csbj.2021.04.052> (2021).
- Norman, T. M. *et al.* Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. *Science* **365**, 786–793. <https://doi.org/10.1126/science.aax4438> (2019).
- Dungan, C. M. *et al.* Muscle-specific cellular and molecular adaptations to late-life voluntary concurrent exercise. *Function* <https://doi.org/10.1093/function/zqac027> (2022).
- El-Brolosy, M. A. & Stainier, D. Y. R. Genetic compensation: A phenomenon in search of mechanisms. *PLoS Genet.* **13**, e1006780. <https://doi.org/10.1371/journal.pgen.1006780> (2017).
- Mohun, T. *et al.* Deciphering the mechanisms of developmental disorders (DMDD): A new programme for phenotyping embryonic lethal mice. *Dis. Model Mech.* **6**, 562–566. <https://doi.org/10.1242/dmm.011957> (2013).
- Geyer, S. H. *et al.* A staging system for correct phenotype interpretation of mouse embryos harvested on embryonic day 14 (E14.5). *J. Anat.* **230**, 710–719. <https://doi.org/10.1111/joa.12590> (2017).
- Jain, R. K. Molecular regulation of vessel maturation. *Nat. Med.* **9**, 685–693. <https://doi.org/10.1038/nm0603-685> (2003).
- Korn, C. & Augustin, H. G. Mechanisms of vessel pruning and regression. *Dev. Cell.* **34**, 5–17. <https://doi.org/10.1016/j.devcel.2015.06.004> (2015).
- Whitehead, L., Watson, E. C. & Grant, Z. L. Analyzing vessel regression and endothelial apoptosis as a component of angiogenic vessel remodeling. *Methods Mol. Biol.* **2441**, 29–39. [https://doi.org/10.1007/978-1-0716-2059-5\\_3](https://doi.org/10.1007/978-1-0716-2059-5_3) (2022).
- Yang, Y. & Oliver, G. Development of the mammalian lymphatic vasculature. *J. Clin. Invest.* **124**, 888–897. <https://doi.org/10.1172/JCI17609> (2014).
- Deng, Q., Ramskold, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193–196. <https://doi.org/10.1126/science.1245316> (2014).
- Davis, C. A. *et al.* The Encyclopedia of DNA elements (ENCODE): Data portal update. *Nucleic Acids Res.* **46**, D794–D801. <https://doi.org/10.1093/nar/gkx1081> (2018).
- Bancroft, J. D., Layton, C. & Suvarna, S. K. *Bancroft's Theory and Practice of Histological Techniques* 8th edn, 126–138 (Elsevier, 2019).
- Übil, E. *et al.* Mesenchymal-endothelial transition contributes to cardiac neovascularization. *Nature* **514**, 585–590. <https://doi.org/10.1038/nature13839> (2014).
- Schipani, E. *et al.* Hypoxia in cartilage: HIF-1 $\alpha$  is essential for chondrocyte growth arrest and survival. *Genes Dev.* **15**, 2865–2876. <https://doi.org/10.1101/gad.934301> (2001).
- Nagao, M. *et al.* Vascular endothelial growth factor in cartilage development and osteoarthritis. *Sci. Rep.* **7**, 13027. <https://doi.org/10.1038/s41598-017-13417-w> (2017).
- Chen, F. *et al.* Atrioventricular conduction and arrhythmias at the initiation of beating in embryonic mouse hearts. *Dev. Dyn.* **239**, 1941–1949. <https://doi.org/10.1002/dvdy.22319> (2010).
- Ross, C. & Boroviak, T. E. Origin and function of the yolk sac in primate embryogenesis. *Nat. Commun.* **11**, 3760. <https://doi.org/10.1038/s41467-020-17575-w> (2020).
- Kingsley, P. D., Malik, J., Fantauzzo, K. A. & Palis, J. Yolk sac-derived primitive erythroblasts enucleate during mammalian embryogenesis. *Blood* **104**, 19–25. <https://doi.org/10.1182/blood-2003-12-4162> (2004).
- Shibuya, M. Vascular endothelial growth factor (VEGF) and its receptor (VEGFR) signaling in angiogenesis: A crucial target for anti- and pro-angiogenic therapies. *Genes Cancer* **2**, 1097–1105. <https://doi.org/10.1177/1947601911423031> (2011).
- Cao, G. *et al.* Involvement of human PECAM-1 in angiogenesis and in vitro endothelial cell migration. *Am. J. Physiol. Cell Physiol.* **282**, C1181–1190. <https://doi.org/10.1152/ajpcell.00524.2001> (2002).
- Sato, T. N. *et al.* Distinct roles of the receptor tyrosine kinases Tie-1 and Tie-2 in blood vessel formation. *Nature* **376**, 70–74. <https://doi.org/10.1038/376070a0> (1995).
- Akazawa, H. & Komuro, I. Roles of cardiac transcription factors in cardiac hypertrophy. *Circ. Res.* **92**, 1079–1088. <https://doi.org/10.1161/01.RES.0000072977.86706.23> (2003).
- Hou, S. *et al.* Heterogeneity in endothelial cells and widespread venous arterialization during early vascular development in mammals. *Cell Res.* **32**, 333–348. <https://doi.org/10.1038/s41422-022-00615-z> (2022).
- Keller, G. Embryonic stem cell differentiation: Emergence of a new era in biology and medicine. *Genes Dev.* **19**, 1129–1155. <https://doi.org/10.1101/gad.1303605> (2005).

38. Hu, J., Zhang, Z., Shen, W. J. & Azhar, S. Cellular cholesterol delivery, intracellular processing and utilization for biosynthesis of steroid hormones. *Nutr. Metab.* **7**, 47. <https://doi.org/10.1186/1743-7075-7-47> (2010).
39. Anuka, E., Gal, M., Stocco, D. M. & Orly, J. Expression and roles of steroidogenic acute regulatory (StAR) protein in “non-classical”, extra-adrenal and extra-gonadal cells and tissues. *Mol. Cell Endocrinol.* **371**, 47–61. <https://doi.org/10.1016/j.mce.2013.02.003> (2013).
40. Garcia, M. D. & Larina, I. V. Vascular development and hemodynamic force in the mouse yolk sac. *Front. Physiol.* **5**, 308. <https://doi.org/10.3389/fphys.2014.00308> (2014).
41. Carvalho, R. L. *et al.* Compensatory signalling induced in the yolk sac vasculature by deletion of TGFbeta receptors in mice. *J. Cell Sci.* **120**, 4269–4277. <https://doi.org/10.1242/jcs.013169> (2007).
42. Croy, B. A. *et al.* Imaging of vascular development in early mouse decidua and its association with leukocytes and trophoblasts. *Biol. Reprod.* **87**, 125. <https://doi.org/10.1095/biolreprod.112.102830> (2012).
43. Elmore, S. A. *et al.* Histology atlas of the developing mouse placenta. *Toxicol. Pathol.* **50**, 60–117. <https://doi.org/10.1177/01926233211042270> (2022).
44. Baron, M. H., Vacaru, A. & Nieves, J. Erythroid development in the mammalian embryo. *Blood Cells Mol. Dis.* **51**, 213–219. <https://doi.org/10.1016/j.bcmd.2013.07.006> (2013).
45. Yamane, T. Mouse yolk sac hematopoiesis. *Front. Cell Dev. Biol.* **6**, 80. <https://doi.org/10.3389/fcell.2018.00080> (2018).
46. Hou, S. *et al.* Embryonic endothelial evolution towards first hematopoietic stem cells revealed by single-cell transcriptomic and functional analyses. *Cell Res.* **30**, 376–392. <https://doi.org/10.1038/s41422-020-0300-2> (2020).
47. Heinke, J., Patterson, C. & Moser, M. Life is a pattern: Vascular assembly within the embryo. *Front. Biosci.* **4**, 2269–2288. <https://doi.org/10.2741/541> (2012).
48. Drews, B., Landaverde, L. F., Kuhl, A. & Drews, U. Spontaneous embryo resorption in the mouse is triggered by embryonic apoptosis followed by rapid removal via maternal sterile purulent inflammation. *BMC Dev. Biol.* **20**, 1. <https://doi.org/10.1186/s12861-019-0201-0> (2020).
49. Cooper, D. N., Krawczak, M., Polychronakos, C., Tyler-Smith, C. & Kehrer-Sawatzki, H. Where genotype is not predictive of phenotype: Towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum. Genet.* **132**, 1077–1130. <https://doi.org/10.1007/s00439-013-1331-2> (2013).
50. Copp, A. J. & Greene, N. D. Neural tube defects—disorders of neurulation and related embryonic processes. *Wiley Interdiscip. Rev. Dev. Biol.* **2**, 213–227. <https://doi.org/10.1002/wdev.71> (2013).
51. Haiman, C. A. *et al.* Genome-wide testing of putative functional exonic variants in relationship with breast and prostate cancer risk in a multiethnic population. *PLoS Genet.* **9**, e1003419. <https://doi.org/10.1371/journal.pgen.1003419> (2013).
52. Hughes, E. D. *et al.* Genetic variation in C57BL/6 ES cell lines and genetic instability in the Bruce4 C57BL/6 ES cell line. *Mamm. Genome* **18**, 549–558. <https://doi.org/10.1007/s00335-007-9054-0> (2007).
53. Durkin, M. E., Qian, X., Popescu, N. C. & Lowy, D. R. Isolation of mouse embryo fibroblasts. *Biol. Protoc.* **3**, 18. <https://doi.org/10.21769/bioprotoc.908> (2016).
54. Lipinski, C. A. & Reaume, A. G. High throughput in vivo phenotypic screening for drug repurposing: Discovery of MLR-1023 a novel insulin sensitizer and novel Lyn kinase activator with clinical proof of concept. *Bioorg. Med. Chem.* **28**, 115425. <https://doi.org/10.1016/j.bmc.2020.115425> (2020).
55. Brewer, G. J., Torricelli, J. R., Evege, E. K. & Price, P. J. Optimized survival of hippocampal neurons in B27-supplemented Neurobasal, a new serum-free medium combination. *J. Neurosci. Res.* **35**, 567–576. <https://doi.org/10.1002/jnr.490350513> (1993).
56. Conti, L. *et al.* Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol.* **3**, e283. <https://doi.org/10.1371/journal.pbio.0030283> (2005).
57. Ying, Q. L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat. Biotechnol.* **21**, 183–186. <https://doi.org/10.1038/nbt780> (2003).
58. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550. <https://doi.org/10.1073/pnas.0506580102> (2005).

## Acknowledgements

We thank Dr. Andrew Lichtman (Brigham & Women’s Hospital, Boston, MA) for guiding discussions, critical inputs, unwavering intellectual support, and for managing members of his lab throughout this research. We thank Dr. Margaret Tarrío (Brigham & Women’s Hospital, Boston, MA) for breeding, raising, performing microscopy and IHC studies on KO mice, as well as performing mouse treatment studies. We thank Dr. Bastian Stielow for creating the SAMD1 KO mouse embryonic stem cells and Iris Rohner for technical assistance. Bruce Campbell is retired.

## Author contributions

B.C., S.J.E., and R.L. wrote the manuscript. L.M.W. and R.L. performed and supervised, respectively, neuronal differentiation experiments with mESCs. S.J.E. was responsible for the generation, breeding, and phenotyping of the SAMD1 mouse line. T.R.S.O. was responsible for the histological and embryonic lethality analyses of the WT and KO embryos. P.B. performed histology and histological analyses of WT and KO embryos. R.A. performed oversight and leadership for the Pfizer research activity including planning, execution, and mentorship, and provided materials, reagents, laboratory samples, animals, instrumentation, computing resources, and other analysis tools. All authors reviewed the manuscript.

## Funding

Open Access funding enabled and organized by Projekt DEAL. This study received funding from Athex and Pfizer. The funders were not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication. R.L. acknowledges support from the German Research Foundation (DFG, TRR 81/3–109546710). Open access funding provided by the Open Access Publishing Fund of Philipps-Universität Marburg with support of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation).

## Competing interests

Authors P. Bourassa and R. Aiello are currently employed by Cybexa Therapeutics. Author S. J. Engle is currently employed by Biogen. The authors declare that the research was conducted in the absence of any commercial

or financial relationships that could be construed as a potential conflict of interest. Thus, all authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-29779-3>.

**Correspondence** and requests for materials should be addressed to B.C. or R.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

## Supplementary Information

### Investigation of SAMD1 Ablation in Mice

**Bruce Campbell<sup>1,#,\$,\*</sup>, Lisa M. Weber<sup>2,\$</sup>, Sandra J. Engle<sup>4,†</sup>, Terence R.S. Ozolinš<sup>5,†</sup>, Patricia Bourassa<sup>6,†</sup>, Robert Aiello<sup>6,†</sup>, Robert Liefke<sup>2,3,\*</sup>**

<sup>1</sup> Retired

<sup>2</sup> Institute of Molecular Biology and Tumor Research (IMT), Philipps University of Marburg, 35043 Marburg, Germany

<sup>3</sup> Department of Hematology, Oncology, and Immunology, University Hospital Giessen and Marburg, 35043 Marburg, Germany

<sup>4</sup> Biogen, Cambridge, MA 02142, USA

<sup>5</sup> Queen's University, Department of Biomedical and Molecular Sciences, Kingston, Ontario K7L 3N6, Canada

<sup>6</sup> Cybrexa Therapeutics, Groton, CT 06340, USA

# Previous affiliation: Atherex Inc., Lincoln, MA 01773, USA

† Previous affiliation: Pfizer Inc., Groton, CT 06340, USA

\$ These authors contributed equally

\* Correspondence:

Bruce Campbell

[bcampbell9516@gmail.com](mailto:bcampbell9516@gmail.com)

Robert Liefke

[robert.liefke@imt.uni-marburg.de](mailto:robert.liefke@imt.uni-marburg.de)

Embryo Genotypes and Lethality Mapping						
Embryonic Day	+/+	+/-	-/-	Total N	p-value	p-value
10.5	1 (5%)	10 (56%)	7 (39%)	18	0.12	E10.5-E12.5 p=0.45
11.5	2 (13%)	7 (47%)	6 (40%)	15	0.33	
12.5	27 (23%)	61 (54%)	25 (22%)	113	0.67	
14.5	29 (30%)	53 (54%)	16 (16%)	98	0.13	E14.5-E18.5 p=0.049
16.5	2 (11%)	10 (56%)	6 (33%)	18	0.36	
18.5	16 (46%)	15 (43%)	4 (11%)	35	0.011	
Expected Mendelian ratio	1 (25%)	2 (50%)	1 (25%)			

**Supplementary Table S1: Embryonic genotyping.**

Embryos were genotyped between E10.5 and E18.5. The decreased total number of KO embryos from E14.5 to E18.5 ( $p < 0.049$ ) suggests embryonic lethality between E12.5 and E14.5. Chi-square calculated p-values (1:2:1 expected ratio) are shown for each embryonic day, for E10.5-E12.5 combined, and for E14.5-E18.5 combined.

Embryo Lethality Mapping when Crossing: $\text{Samd1}^{+/-} \times \text{Samd1}^{+/-}$						
Sex	+/+	+/-	-/-	Total N	p-value KO Mortality (1:2:1 ratio)	p-value HET Mortality (1:2 ratio)
Male (55.2%)	35 (47%)	39 (53%)	0	74	$5.8 \times 10^{-8}$	0.01
Female (44.8%)	26 (43%)	34 (57%)	0	60	$7.5 \times 10^{-6}$	0.01
Total M+F	61 (45.5%)	73 (54.5%)	0	134	$5.1 \times 10^{-13}$	0.003
Expected Mendelian ratios	1 (25%)	2 (50%)	1 (25%)			

**Supplementary Table S2:  $\text{Samd1}^{+/-} \times \text{Samd1}^{+/-}$  crossings.**

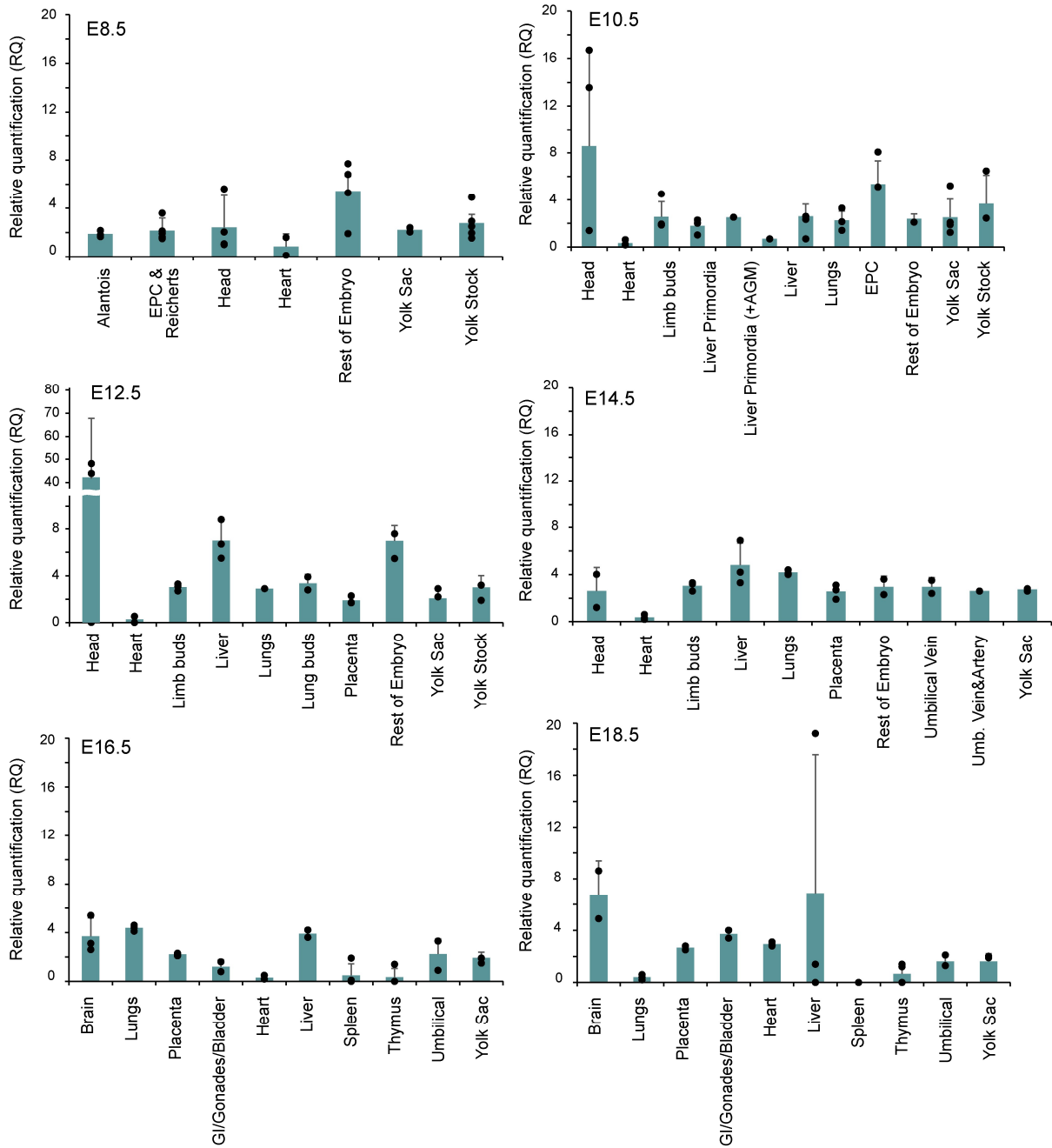
Genotyping of  $\text{Samd1}^{+/-} \times \text{Samd1}^{+/-}$  crossings at P21. P-values for KO lethality were calculated using Chi-square and the 1:2:1 Mendelian ratio. P-values for HET lethality were calculated using Chi-square for a 1:2 Mendelian ratio.



Embryo Lethality Mapping when Crossing: Samd1 <sup>+/+</sup> x Samd1 <sup>+/-</sup>				
Sex	+/+	+/-	Total N	p-value HET Mortality (1:1 ratio)
Male (50.9%)	69 (56%)	55 (44%)	124	0.21
Female (49.1%)	76 (63%)	44 (37%)	120	0.003
Total M+F	145 (59.5%)	99 (40.6%)	244	0.003
Expected Mendelian ratios	1(50%)	1 (50%)		

**Supplementary Table S3: Samd1<sup>+/+</sup> x Samd1<sup>+/-</sup> crossings.**

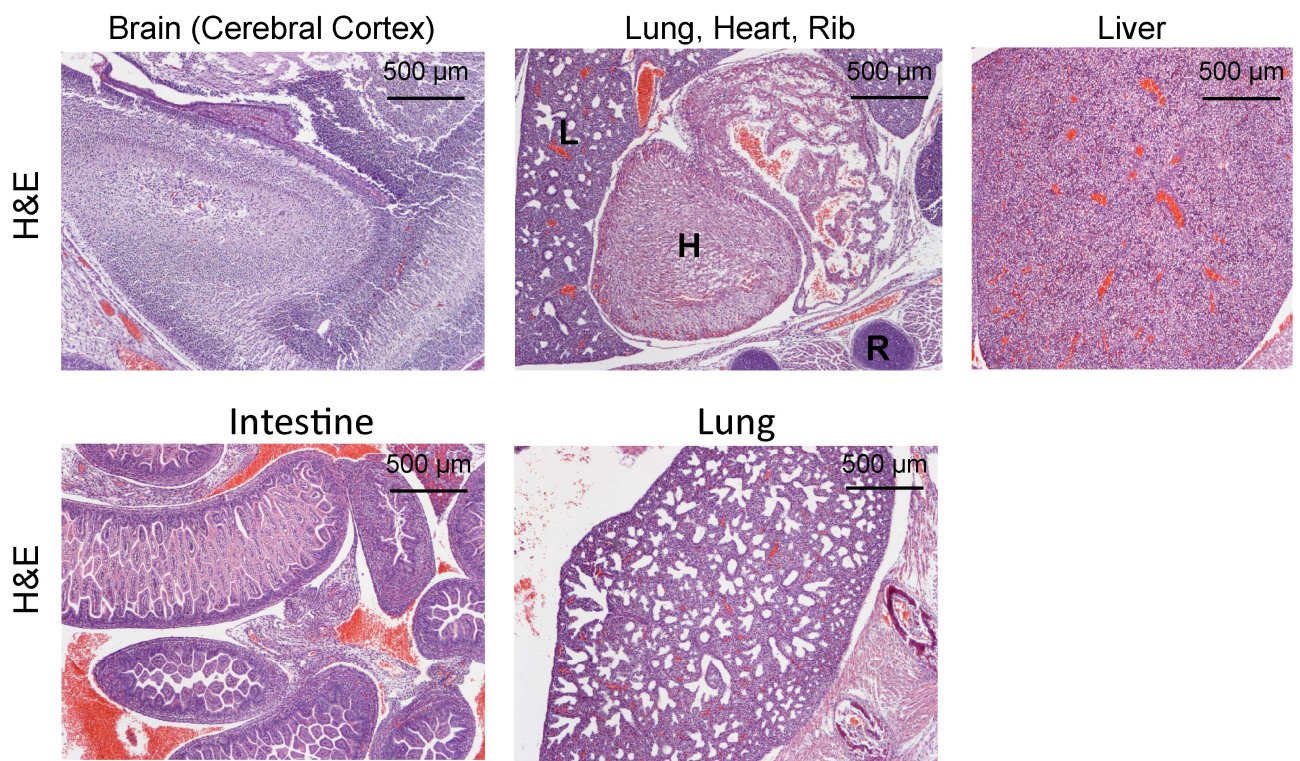
Genotyping of Samd1<sup>+/+</sup> x Samd1<sup>+/-</sup> crossings at P21. P-values for HET lethality were calculated using Chi-square for the 1:1 Mendelian ratio.



**Supplementary Figure S1: Gene expression of SAMD1 during embryogenesis.**

RT-qPCR analysis of mRNA levels of SAMD1 using samples from wild-type embryos at various stages. Data were normalized to *Gapdh*, and presented as relative quantification (QC) compared to the SAMD1 mRNA level from an entire E8.5 embryo. Data are presented as mean  $\pm$  SD.

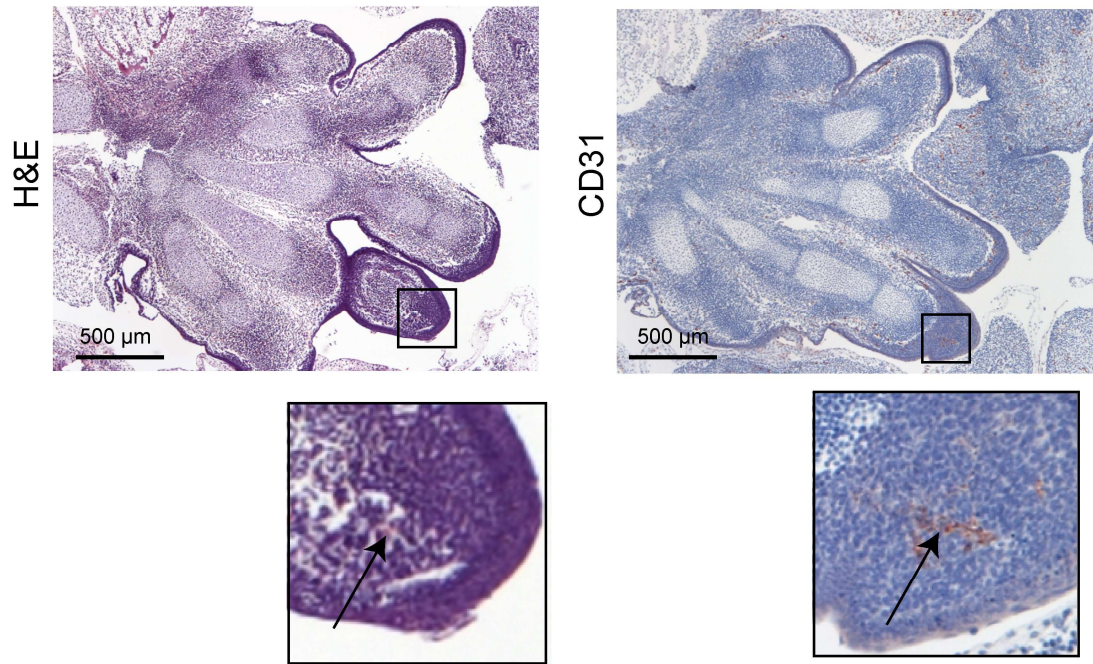
HET



**Supplementary Figure S2: Internal organs of SAMD1 HET mice are normal.**

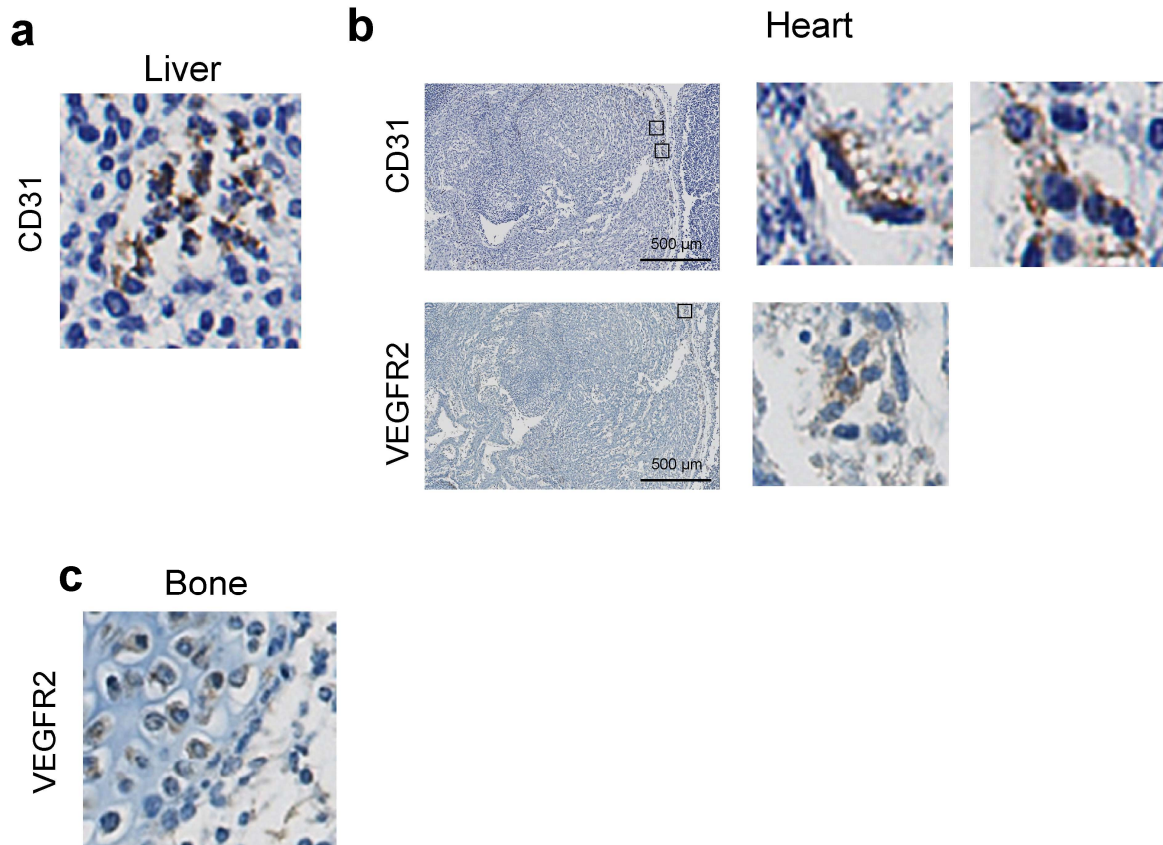
H&E staining of various organs in HET mice at day E14.5. In the upper middle panel lung (L), heart (H) and rib (R) are shown. These tissues are indistinguishable from the WT tissues shown in Figures 2 and 3.

### Forepaw



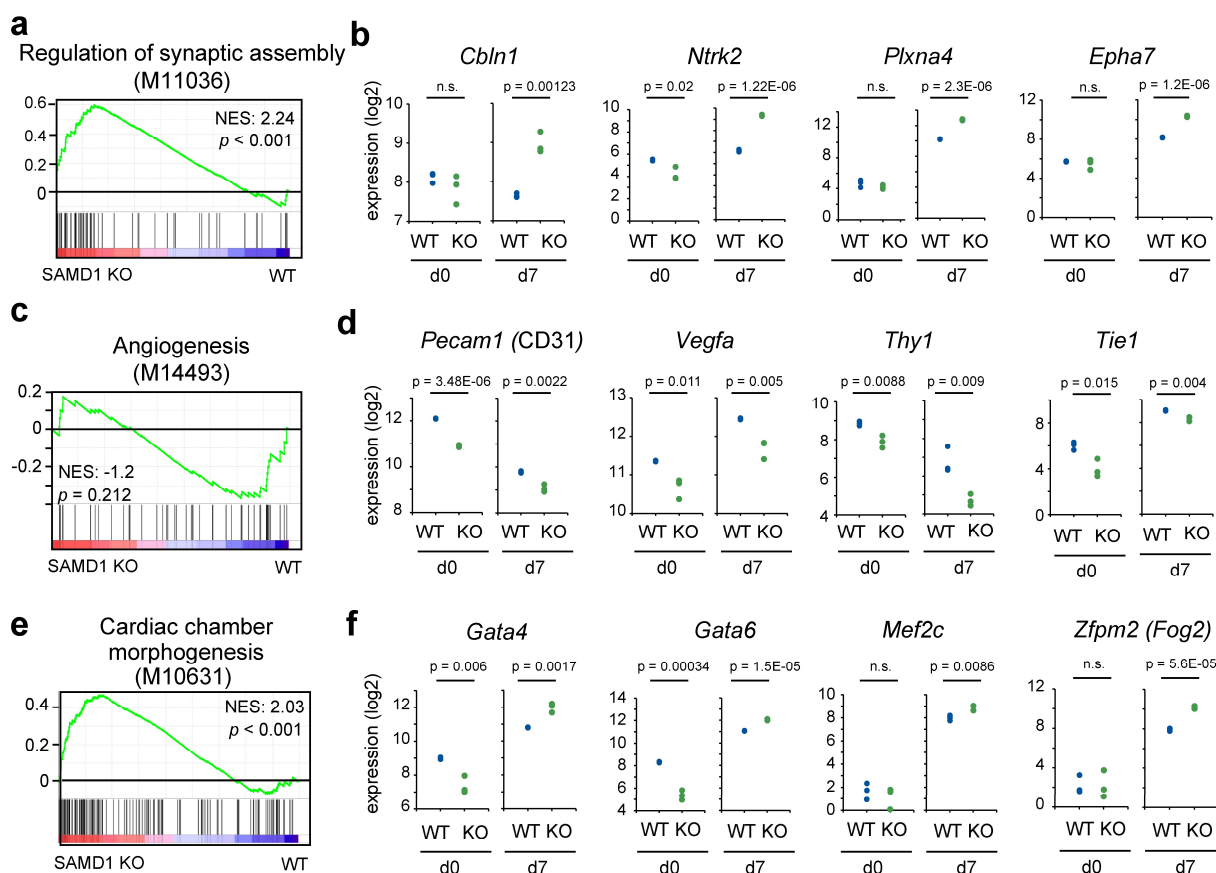
**Supplementary Figure S3: H&E and CD31 staining of SAMD1 KO embryo forepaw at E14.5.**

The arrows indicate a few scattered RBCs in the H&E images and misshapen ECs in a disorganized pattern in the CD31 image.



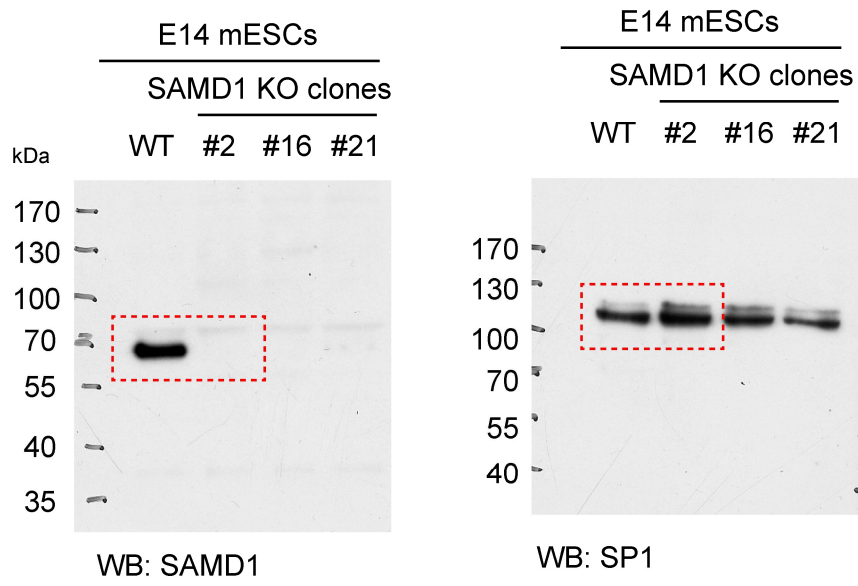
**Supplementary Figure S4: Cell fragments in liver and heart of E14.5 KO embryo.**

CD31 and VEGFR2 (brown) staining (counterstained with hematoxylin (blue)) in liver (**a**) and heart (**b**), showing examples for cell fragments and phagocytosis of ECs in SAMD1 KO mice. (**c**) VEGFR2-stained chondrocytes in the bone of an E14.5 embryo.



**Supplementary Figure S5: SAMD1 deletion impairs multiple cellular pathways during undirected ES cell differentiation.**

(a,c,e) GSEA (gene set enrichment analysis) of RNA-Seq data after 7 days of undirected ES cell differentiation<sup>4</sup>. (a) GSEA of synaptic assembly pathways. (b) Example dysregulated genes in differentiated SAMD1 KO mES cells related to neuronal pathways. (c) GSEA of angiogenesis pathways. (d) Examples of genes related to angiogenesis and arteriogenesis affected upon SAMD1 deletion in undifferentiated (d0) and differentiated (d7) ES cells. (e) GSEA of the cardiac chamber morphogenesis pathway. (f) Example dysregulated genes in SAMD1 KO cells related to heart development. Data from RNA-Seq. P-values two-tailed unpaired student's t-test. n.s. = not significant.



**Supplementary Figure S6: Full Western blots.**

## Supplementary Table S4: RT-qPCR primers

### *Housekeeping genes*

RT\_mActb\_fw GTA-CCC-AGG-CAT-TGC-TGA-CA  
RT\_mActb\_rv AGG-GTG-TAA-AAC-GCA-GCT-CAG  
RT\_mGapdh\_fw AGA-CGG-CCG-CAT-CTT-CTT-GT  
RT\_mGapdh\_rv GCC-TTG-ACT-GTG-CCG-TTG-AA

### *Target genes*

RT\_mActn1\_fw GAT-ATT-GGC-AAC-GAC-CCC-CA  
RT\_mActn1\_rv TAA-TGA-ACC-CAG-CCA-GCC-AG  
RT\_mCbln1\_fw GAG-CCG-TCC-GAG-ATG-AGT-AA  
RT\_mCbln1\_rv CAC-CCG-TTC-AAC-ATG-AGG-CT  
RT\_mCdh2\_fw TGG-GAA-TCA-GAC-GGC-TAG-AC  
RT\_mCdh2\_rv GGT-AGT-CAT-AGT-CCT-GGT-CC  
RT\_mChd7\_fw CAG-ACG-TGC-TGT-TTT-CCT-CG  
RT\_mChd7\_rv CAG-TGA-GGA-GAC-GGT-CAA-AC  
RT\_mCldn11\_fw TTG-ACA-GTT-CTC-CCC-TGC-AT  
RT\_mCldn11\_rv GTA-GCC-AAA-GCT-CAC-GAT-GG  
RT\_mFoxa2\_fw GAT-GGA-AGG-GCA-CGA-GCC  
RT\_mFoxa2\_rv GTA-TGT-GTT-CAT-GCC-ATT-CAT-CCC  
RT\_mGalnt9\_fw CTC-TAC-CCC-TGT-CAT-GGC-AT  
RT\_mGalnt9\_rv TCT-CTA-CCT-CCA-GAC-ACC-GA  
RT\_mGfap\_fw AAG-GTT-GAA-TCG-CTG-GAG-GA  
RT\_mGfap\_rv CAC-TGC-CTC-GTA-TTG-AGT-GC  
RT\_mIsl1\_fw CCC-TCT-CAG-TCC-CTT-GCA-TC  
RT\_mIsl1\_rv GCG-TTT-CTT-GTC-CTT-GCA-CC  
RT\_mKlf4\_fw AAG-AGG-GGA-AGA-AGG-TCG-TG  
RT\_mKlf4\_rv CTG-TCA-CAC-TTC-TGG-CAC-TG  
RT\_mL3mbtl3\_fw CTT-GCC-TGG-GTG-TGA-AGA-AC  
RT\_mL3mbtl3\_rv TCC-GCA-GCT-TTG-AAC-ATC-AG  
RT\_mMyh6\_fw CAC-AGA-TGC-CGC-AAT-GAT-GG  
RT\_mMyh6\_rv GTC-GTG-CAT-CTT-CTT-GGC-AC  
RT\_mMyl2\_fw TTA-TTG-TTC-CAC-AGC-AGG-GGC  
RT\_mMyl2\_rv TGC-CCT-CCG-GAT-CAA-ACA-C  
RT\_mNanog\_fw CAC-AGT-TTG-CCT-AGT-TCT-GAG-G  
RT\_mNanog\_rv GCA-AGA-ATA-GTT-CTC-GGG-ATG-AA  
RT\_mNanos1\_fw GCC-TAG-TTT-AGG-TGC-GCA-AC  
RT\_mNanos1\_rv ACG-AGG-AAG-AAC-ACC-CTC-TC  
RT\_mNes\_fw GCC-TAT-AGT-TCA-ACG-CCC-CC  
RT\_mNes\_rv AGA-CAG-GCA-GGG-CTA-GCA-AG  
RT\_mNkx2-5\_fw ATT-TTA-CCC-GGG-AGC-CTA-CG  
RT\_mNkx2-5\_rv GCT-GTC-GCT-TGC-ACT-TGT-AG  
RT\_mNppa\_fw GCT-TCG-GGG-GTA-GGA-TTG-AC  
RT\_mNppa\_rv GAG-GCA-AGA-CCC-CAC-TAG-AC  
RT\_mNrp1\_fw AGC-TTC-GGA-CGT-TTT-CAC-CT  
RT\_mNrp1\_rv GGA-AGT-CAT-CAC-CTG-TGC-CA  
RT\_mNrp2\_fw TGC-AGG-TGA-GGA-TTT-TAA-AGA-TGA  
RT\_mNrp2\_rv CAG-GTG-CAG-TAA-AGG-AGG-AGG  
RT\_mOct4\_fw TGG-AAA-GGT-GTT-CAG-CCA-GA



RT_mOct4_rv	CCT-CAC-ACG-GTT-CTC-AAT-GC
RT_mOlig1_fw	TGA-ATC-CCA-CCT-GTT-TAG-AGC-C
RT_mOlig1_rv	CGA-TGC-TCA-CGG-ATA-CGA-GAA-TAG
RT_mPax3_fw	TCC-CAT-GGT-TGC-GTC-TCT-AAG
RT_mPax3_rv	CTC-CAC-GTC-AGG-CGT-TGT-C
RT_mPax7_fw	TGG-GGT-CTT-CAT-CAA-CGG-TC
RT_mPax7_rv	ATC-GGC-ACA-GAA-TCT-TGG-AGA
RT_mPth2_fw	GAG-ACC-TGC-CAG-ATG-TCC-AG
RT_mPth2_rv	CTG-CAT-GTA-AGA-GTC-CAG-CC
RT_mSamd1_fw	CTG-ATG-GGA-CAC-CTT-TTG-GC
RT_mSamd1_rv	GTG-AGC-ACA-TCT-GTC-CGT-TG
RT_mSfmbt1_fw	CCA-ACA-AGA-GAG-ATG-CCC-AG
RT_mSfmbt1_rv	GTG-GGA-AGG-GTA-AGG-AGC-AA
RT_mSox2_fw	AGG-AGA-GAA-GTT-TGG-AGC-CC
RT_mSox2_rv	TCT-GGC-GGA-GAA-TAG-TTG-GG
RT_mTbx5_fw_2	CCC-CCT-GTA-CAG-AGC-GAG-AAT-A
RT_mTbx5_rv_2	GGT-CGT-CTG-CGG-GAA-CAA-TA
RT_mTnnt2_fw	GTG-TGC-AGT-CCC-TGT-TCA-GA
RT_mTnnt2_rv	GCA-CCA-AGT-TGG-GCA-TGA-AG
RT_mTtn_fw	ACC-CTT-TTT-GGG-CAC-TCC-TG
RT_mTtn_rv	CTG-TTG-GCT-TTA-GTC-ACG-GC
RT_mTubb3_fw	TAG-ACC-CCA-GCG-GCA-ACT-AT
RT_mTubb3_rv	GTT-CCA-GGT-TCC-AAG-TCC-ACC

## *II. Curriculum vitae*

Contains personal information of the author and is therefore excluded from the online accessible version of the dissertation.

*III. PhD Portfolio*

**Summary of PhD training – Graduate school TRR 81**

**Name of PhD student:** Lisa Marie Weber

**Affiliation:**

Institut für Molekularbiologie und Tumorforschung (IMT)

Philipps-Universität Marburg

Hans-Meerwein-Str. 2

D-35043 Marburg

Tel: +49 6421 28-65039

E-mail: lisa.weber@imt.uni-marburg.de

**Start of PhD work:** November 01, 2018

**Topic of Dissertation:**

„Characterization of the SAMD1-like winged helix domain of histone acetyltransferase KAT6A“

**Thesis committee:**

1. Dr. Robert Liefke
2. Prof. Dr. Joost Gribnau
3. Prof. Dr. Katja Nieweg

**Meetings of the thesis committee:**

- Kleinwalsertal, February 26, 2019
- Marburg, March 20, 2019
- Kleinwalsertal, February 18, 2020
- Marburg, April 12, 2021
- Kleinwalsertal, May 16, 2022

## **Seminars and workshops:**

### MARburg University Research Academy (MARA) courses:

- MARA Workshop: “Writing Lab - Scientific Writing for the Life and Natural Sciences” (Marburg, April 15 - 16, 2019)
- MARA Workshop: „GMP, GLP und GCP - Grundlagen und erweiterte Anforderungen in der pharmazeutischen Industrie” (Marburg, February 6 - 7, 2020)
- MARA Workshop: “Project Management for Scientists” (Online, March 26 - 27, 2020)
- MARA Workshop: “Good Scientific Practice - Protecting Scientific Integrity” (Online, July 1 - 2, 2020)
- MARA Workshop: „GMP Aufbaukurs” (Online, March 4, 2022)
- MARA annual PhD and Postdoc meeting: “Introduction to Science Communication” (Marburg, September 13, 2022)
- MARA Workshop: „Promotion und dann? – Perspektiven und Impulse für die Karriereplanung in den Lebens- und Naturwissenschaften” (Online, October 25 - 26, 2022)
- MARA Workshop: “Get that Job after your Doctorate! - Applying & Interviewing for Positions in Industry and Research” (Online, March 1 and 8, 2023)
- MARA Workshop: „Disputationstraining” (Online, April 20 - 21, 2023)
- MARA Workshop: „Wie funktionieren Unternehmen? - Mein Einstieg in die Wirtschaft” (Marburg, Mai 22 - 23, 2023)

### External courses:

- Sheboss Workshop: „Spiele mit der Macht” (Marburg, July 09, 2019)
- ecSeq Bioinformatics Workshop: “A Practical Introduction to NGS Data Analysis” (Online, July 14 - 16, 2021)
- Erasmus Optical Imaging Centre, Advanced Light Microscopy Course: “Functional Imaging and Super Resolution” (Rotterdam, November 4 - 8, 2019)
- LAS interactive and MARA Workshop: „Versuchstierkunde – Tierschutzgerechter Umgang mit Labormäusen” (Marburg, October, 2020)
- BioScript International Workshop: “Advanced Scientific Writing: Writing Publishable Research Articles” (Marburg, September 6 - 8, 2021)

## **Presentations:**

### Journal Club:

- “Promoter-Enhancer communication occurs primarily within insulated neighborhoods “ Sun et al., 2018, Molecular Cell (Marburg, March 14, 2019)
- “A chromatin integration labelling method enables epigenomic profiling with lower input” Harada et al., 2019, Nature Cell Biology (Marburg, October 31, 2019)

- “Corona Special“, various sources, mainly “A pneumonia outbreak associated with a new coronavirus of probable bat origin” Zhou et al., 2020, Nature (Marburg, March 12, 2020)
- “Histone modifications form a cell-type-specific chromosomal bar code that persists through the cell cycle” Halsall et al., 2021, Scientific Reports (Online, February 08, 2021)
- “Differential H4K16ac levels ensure a balance between quiescence and activation in hematopoietic stem cells” Rodrigues and Akhtar, 2021, Science Advances (Online, September 16, 2021)
- “A dynamic and combinatorial histone code drives malaria parasite asexual and sexual development” Grüning et al., 2022, Molecular and Cellular Proteomics (Online, March 17, 2022)

#### Retreats:

- 9th Winter School of the Collaborative Research Centre TRR81 “Chromatin Changes in Differentiation and Malignancies“ Marburg - Gießen - Bad Nauheim - Rotterdam (Kleinwalsertal, February 24 - March 1, 2019); Talk: “The role of EPOP in differentiation and reprogramming“
- 10th Winter School of the Collaborative Research Centre TRR81 “Chromatin Changes in Differentiation and Malignancies“ Marburg - Gießen - Bad Nauheim - Rotterdam (Kleinwalsertal, February 16 - 21, 2020); Talk: “The biological role of EPOP during embryogenesis and neurogenesis“
- 11th Winter School of the Collaborative Research Centre TRR81 “Chromatin Changes in Differentiation and Malignancies“ Marburg - Gießen - Bad Nauheim - Rotterdam (Online, February 24 - 25, 2021); Talk: “Putative function of the histone acetyltransferase KAT6A as CpG island-binding protein “
- 1st Spring School of the Collaborative Research Centre TRR81 “Chromatin Changes in Differentiation and Malignancies“ Marburg - Gießen - Bad Nauheim - Rotterdam (Kleinwalsertal, May 14 - 19, 2022); Talk: “A conserved DNA-binding winged-helix domain is required for chromatin interaction of KAT6A “

#### Conferences:

- 5th TRR81 Symposium “Chromatin Changes in Differentiation and Malignancies“ (Bad Nauheim, September 09 - 11 2019); Poster presentation
- 6th TRR81 Symposium “Chromatin Changes in Differentiation and Malignancies“ (Egmond aan Zee, Netherlands, September 27 – 29, 2022); Poster presentation

#### Invited Talks:

- Chromatin Club - Campus Germany, Seminar Series by Active Motif: “KAT6A directly interacts with unmethylated CGIs via a conserved winged helix domain” (Marburg, March 30, 2023)

**Other:**

External Training:

- Internship in the laboratory of Prof. Dr. Joost Gribnau (Department of Developmental Biology, Erasmus MC, Rotterdam, April 23-30, 2019)

- 

Supervision of student internship:

- April 26, 2021 – June 20, 2021, master student

Miscellaneous:

- Organization of the 15th TRR81 PhD Mini-Symposium “Exploring the Chromatin Landscape – New Insights into Epigenetics” (Marburg, October 20, 2022)

#### *IV. List of academic teachers*

My academic teachers at the Osnabrück University were:

Engelbrecht-Vandré, Deckers-Hebestreit, Fröhlich, Harten, Heinisch, Hensel, Holthuis, Krüppel, Kümmel, Piehler, Remy, Schmitz, Schrempf, Tabke, Ungermann, Walter, Wieczorek

My academic teachers at the Rhine-Waal University of Applied Science were:

Böhmer, Bockmühl, Fensterle, Henselder, Kleinke, Koch, Kugler, Merettig, Moenickes, Neu, Fenés, Pauling, Renner, Scholz, Simon, Weber

## V. Danksagung

An dieser Stelle möchte ich mich herzlich bei allen bedanken, die zum Gelingen dieser Arbeit beigetragen haben.

Zunächst geht mein Dank an Dr. Robert Liefke, für die Möglichkeit meine Doktorarbeit in seiner Arbeitsgruppe schreiben zu dürfen; für seine fachliche Betreuung und seine Unterstützung während der Zeit meiner Promotion.

Ein besonderer Dank geht an meine KollegInnen Iris Rohner, Dr. Bastian Stielow, Sabrina Fischer, Clara Simon und Merle Geller, für die offene und positive Arbeitsatmosphäre, die große Hilfsbereitschaft, und die Bereitwilligkeit jeden Erfolg oder Fortschritt mit einem dem Anlass entsprechenden Kuchen zu feiern.

Ich danke vor allem Prof. Dr. Zhanxin Wang und seinen MitarbeiterInnen, sowie Prof. Dr. Martha Bulyk und ihren MitarbeiterInnen, für die erfolgreiche Zusammenarbeit und ihre wertvollen Beiträge für die KAT6A Publikation. Außerdem möchte ich Bruce Campbell danken, für seinen großartigen Einsatz für die Veröffentlichung des SAMD1 Artikels.

Ich bin allen dankbar, die während meiner Promotion im Rahmen des Instituts für Molekularbiologie und Tumorforschung sowie dem Forschungsverbund TRR81 zum wissenschaftlichen und kollegialen Austausch beigetragen haben. Insbesondere weiß ich Elisabeth Schneider, Dr. Christina Stielow und Martina Rau für ihre exzellenten Organisationstalente und ihr hilfsbereites, lösungsorientiertes Handeln zu schätzen.

Meinen lieben FreundInnen und KollegInnen Sabrina, Clara, Jonathan, Igor, Satis, Joanna, Franzi, Deborah, Tjorven und Merle möchte ich danken. Danke für die gegenseitige Unterstützung, die Ablenkung vom Alltag, die Motivation und die Freude, die wir in so manchen Kaffeepausen, Labor-Tanzeinheiten, Filmeabenden und Wandertouren geteilt haben.

Abschließend richtet sich mein Dank an meine Familie, meine Freunde und Meikel, für eure Unterstützung, eure Geduld, und eure ermutigenden Worte. Meinen Eltern, die mir während der Promotion und des Studiums stets die größte Stütze waren, danke ich ganz besonders. Danke für eure Liebe, euer Verständnis, euer Zuhören und euren Zuspruch.



## VI. Ehrenwörtliche Erklärung

Ich, Lisa Marie Weber, geboren am 24.09.1992 in Siegen, erkläre ehrenwörtlich, dass ich die dem Fachbereich Medizin Marburg zur Promotionsprüfung eingereichte Arbeit mit dem Titel “Characterization of histone acetyltransferase KAT6A and transcriptional repressor SAMD1 – two proteins possessing a novel class of unmethylated DNA-binding winged helix domains” im Institut für Molekularbiologie und Tumorforschung (Geschäftsführender Direktor: Prof. Dr. Alexander Brehm) unter Leitung von Dr. Robert Liefke ohne sonstige Hilfe selbst durchgeführt und bei der Abfassung der Arbeit keine anderen als die in der Dissertation aufgeführten Hilfsmittel benutzt habe. Ich habe bisher an keinem in- oder ausländischen Medizinischen Fachbereich ein Gesuch um Zulassung zur Promotion eingereicht, noch die vorliegende oder eine andere Arbeit als Dissertation vorgelegt.

Ich versichere, dass ich sämtliche wörtlichen oder sinngemäßen Übernahmen und Zitate kenntlich gemacht habe.

Mit dem Einsatz von Software zur Erkennung von Plagiaten bin ich einverstanden.

Vorliegende Arbeit wurde in folgenden Publikationsorganen veröffentlicht:

**Weber, L. M.**, Jia, Y., Stielow, B., Gisselbrecht, S. S., Cao, Y., Ren, Y., Rohner, I., King, J., Rothman, E., Fischer, S., Simon, C., Forné, I., Nist, A., Stiewe, T., Bulyk, M. L., Wang, Z., & Liefke, R. (2023). The histone acetyltransferase KAT6A is recruited to unmethylated CpG islands via a DNA binding winged helix domain. *Nucleic Acids Research*, 51(2), 574–594.  
<https://doi.org/10.1093/nar/gkac1188>

Campbell, B., **Weber, L. M.**, Engle, S. J., Ozolinš, T. R. S., Bourassa, P., Aiello, R., & Liefke, R. (2023). Investigation of SAMD1 ablation in mice. *Scientific Reports*, 13(1), 3000.  
<https://doi.org/10.1038/s41598-023-29779-3>

Ort, Datum, Unterschrift (Lisa Marie Weber)

---

Die Hinweise zur Erkennung von Plagiaten habe ich zur Kenntnis genommen.

Ort, Datum, Unterschrift (Dr. Robert Liefke)