Title

A conserved BAH module within mammalian BAHD1 connects H3K27me3 to Polycomb gene silencing

Authors

Huitao Fan[1,2†], Yiran Guo[1,3†], Yi-Hsuan Tsai[1], Aaron J. Storey[4], Arum Kim[1,2], Weida Gong[1], Ricky D. Edmondson[4], Samuel G. Mackintosh[4], Haitao Li[5], Stephanie D. Byrum[4], Alan J. Tackett[4], Ling Cai[1,6], Gang Greg Wang[1,2,3*]

[1]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599, USA.
[2]Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599, USA.
[3]Curriculum in Genetics and Molecular Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.
[4]Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA.
[5]Tsinghua-Peking Center for Life Sciences, Beijing Advanced Innovation Center for Structural Biology, and Beijing Frontier Research Center for Biological Structure, Tsinghua University, Beijing 100084, China.
[6]Department of Genetics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599, USA.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.
* To whom correspondence should be addressed. Tel: +1 919-966-5952; Fax: +1 919-966-9673; Email: greg_wang@med.unc.edu

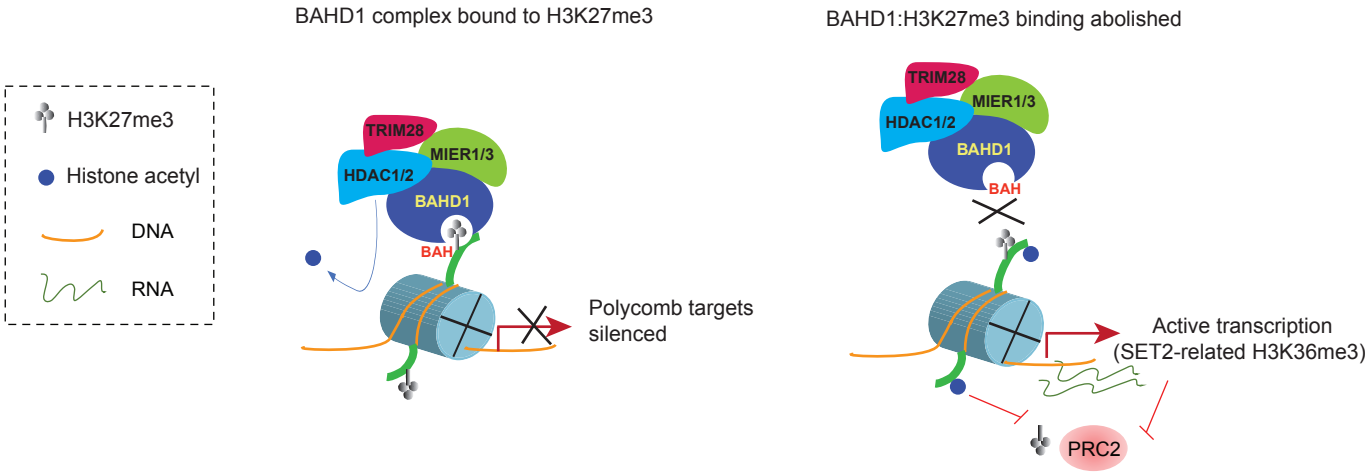
ABSTRACT

Trimethylation of histone H3 lysine 27 (H3K27me3) is important for gene silencing and imprinting, (epi)genome organization and organismal development. In a prevalent model, the functional readout of H3K27me3 in mammalian cells is achieved through the H3K27me3-recognizing chromodomain harbored within the chromobox (CBX) component of canonical Polycomb repressive complex 1 (cPRC1), which induces chromatin compaction and gene repression. Here, we report that binding of H3K27me3 by a Bromo Adjacent Homology (BAH) domain harbored within BAH domain-containing protein 1 (BAHD1) is required for overall BAHD1 targeting to chromatin and for optimal repression of the H3K27me3-demarcated genes in mammalian cells. Disruption of direct interaction between $BAHD1^{BAH}$ and H3K27me3 by point mutagenesis leads to chromatin remodeling, notably, increased histone acetylation, at its Polycomb gene targets. Mice carrying an H3K27me3-interaction-defective mutation of $Bahd1^{BAH}$ causes marked embryonic lethality, showing a requirement of this pathway for normal development. Altogether, this work demonstrates an H3K27me3-initiated signaling cascade that operates through a conserved BAH “reader” module within BAHD1 in mammals.

Key Points **3 bullet points summarizing the manuscript's contribution to the field (100 characters max per point).**

(1) $BAHD1^{BAH}$ is a functionally validated mammalian "reader" of H3K27me3, mediating BAHD1 targeting for gene silencing.
(2) $BAHD1^{BAH}$ connects H3K27me3 together with histone deacetylation, an integral step of gene silencing.
(3) $BAHD1^{BAH}$-mediated functional readout of H3K27me3 is essential for organismal development.

Graphic abstract.



Caption:
A mammalian H3K27me3-transduction pathway operates through an H3K27me3-specific 'reader' module (BAH) of BAHD1, which assembles a complex with corepressors (HDACs and others) for suppressing histone acetylation and repressing expression at Polycomb target genes.

INTRODUCTION

Trimethylation of histone H3 lysine 27 (H3K27me3), catalyzed by Polycomb repressive complex 2 (PRC2), is widely viewed to be crucial for regulation of gene silencing and imprinting, epigenomic and cellular states, cell fate determination, and embryogenesis (1-5). As a support, mutation of the histone H3K27 site or its modifying enzymes occur recurrently in a range of human diseases, such as cancer and developmental syndrome (6-9). A prevalent model for functional transduction of H3K27me3 in mammals is that H3K27me3 is read and bound by chromodomain harbored within the chromobox (CBX) component of canonical Polycomb repressive complex 1 (cPRC1), which can induce gene silencing at least partly through mono-ubiquitination of histone H2A lysine 119 (H2AK119ub1) (2,3,10,11) and/or a phase separation-mediated chromatin compaction mechanism (12-14). However, emerging evidence has shown that, in mammalian cells, H3K27me3 is also 'sensed' and bound by a second class of "reader" modules termed Bromo-Adjacent Homology (BAH), which exist within chromatin regulators such as BAH Domain Containing 1 (BAHD1) (15,16) and BAH and Coiled-coil Domain-containing Protein 1 (BAHCC1) (17). In particular, it was recently demonstrated that disrupting the direct binding of the BAHCC1 BAH motif ($BAHCC1^{BAH}$) to H3K27me3 via point mutagenesis leads to significant de-repression of the H3K27me3-demacated genes (17). Mechanistically, BAHCC1 interacts with histone deacetylases (HDACs), thereby linking H3K27me3 with histone deacetylation, an integral step of gene repression (17). Interestingly, BAHD1 also interacts with HDACs and mediates heterochromatin formation in mammalian cells (15,18). Although the BAH domain within BAHD1 ($BAHD1^{BAH}$) was previously shown to bind H3K27me3 directly (16), the requirement of such an interaction for target gene regulation yet remains uncharacterized to date; a multifunctional scaffold role of BAHD1 was previously proposed for inducing gene repression, which involves the H3K9me3-specific readers and other BAHD1-associated silencing factors (15,18). It shall be noted that, in plant and fungus, the cPRC1 pathway gene(s) is absent and, instead, a set of the plant/fungus-specific BAH-containing proteins were recently identified, which directly engage H3K27me3 for mediating Polycomb gene silencing (19-24). These observations highlight biological significance and cross-species conservation of the BAH module-based functional readout of H3K27me3. Dissecting the molecular underpinning of chromatin modification-initiated signal transduction shall help understand the fundamentals of (epi)genomic and transcriptomic regulation.

Using independent mammalian model systems such as 293 cells, primary mouse embryonic fibroblast (MEF) cells and the genetically modified knockin mice, we here aim to determine the role of the H3K27me3-binding $BAHD1^{BAH}$ in target gene regulation. Our findings, based on mutagenesis of $BAHD1^{BAH}$, substantiate a critical requirement of $BAHD1^{BAH}$:H3K27me3 interaction for overall BAHD1 targeting to chromatin and for maintaining local histone deacetylation, resulting in optimal repression of the BAHD1-bound Polycomb target genes. Importantly, the readout of H3K27me3 by $BAHD1^{BAH}$ is essential for embryogenesis and survival in mice. We thus favor a view that the functional transduction of H3K27me3 via BAH-directed molecular pathways is evolutionarily conserved across the domain of Eukaryota, including animal, which operate to ensure appropriate epigenomic and transcriptomic patterns during organismal development.

MATERIALS AND METHODS

Plasmids

Full-length cDNA of human BAHD1 (a kind gift of H. Bierne, L'Institut Micalis, France), together with a 3×FLAG tag fused to its C-terminal, was cloned into the XhoI and NotI site of the pPyCAGIP vector (carrying IRES-Puro; a kind gift of I. Chambers). Point mutation was generated by site-directed mutagenesis. Information of primers used in the work is listed in Supplementary Table 6. Correctness of all plasmids was verified by Sanger sequencing before use.

Cell lines and tissue culture

HEK293 cells (American Tissue Culture Collection, #CRL-1573) were maintained under the vendor-recommended culture condition. Cells were transfected by Lipofectamine 2000 (Invitrogen) with the pPyCAGIP empty vector or that carrying wild-type or mutant 3×FLAG-BAHD1. Two days post-transfection, HEK293 cells were subjected to drug selection in the medium with 2 μg/ml puromicin (Invitrogen) for 14 days. The stable expression lines were continuously maintained in the medium with puromycin. Authentication of cell line identity, including parental and their derived lines, was ensured by the Tissue Culture Facility (TCF) affiliated to UNC Lineberger Comprehensive Cancer Center with the genetic signature profiling and fingerprinting analysis (25). Each month, a routine examination of cells for mycoplasma contamination was conducted by using the detection kit (Lonza). Cells with less than 10 times of passages were used in the study.

RNAi for gene knockdown (KD)

Cells were transfected with siRNA using the Lipofectamine RNAiMAX reagent according to manufacturer's instruction. The Silencer™ Select Pre-Designed siRNA for human BAHD1 (Thermo Fisher 4392422, siRNA # s22604) was ordered and used per vendor's guideline.

Quantitative RT-PCR (RT-qPCR)

RT-qPCR was carried out as previously described (17,26), with the primer sequence listed in Supplementary Table 6.

Co-immunoprecipitation (CoIP) and peptide pulldown

Whole cell lysates were prepared by using the NP40 lysis buffer (50mM Tris-HCl pH 8.0, 150mM NaCl, 1% IGEPAL® CA-630, and 1× protease inhibitor cocktail added freshly), followed by brief sonication and high-speed centrifugation at 4°C. For CoIP, cell lysates were incubated with antibodies on a rotator overnight in a cold room, followed by addition of protein A/G beads for additional 2 hrs. Beads were washed three times in NP40 lysis buffer, resuspended in 50 μl of 1× SDS sample buffer, and boiled at 90 degree for 5 min before loading onto SDS-PAGE gel. We conducted the biotinylated peptide pulldown using total cell lysate according to a previously described protocol (17,26).

Antibodies and immunoblotting

The information of antibodies used for immunoblotting, immunofluorescence, CoIP and ChIP, ChIP-seq and CUT&RUN can be found in Supplementary Table 6. Protein samples, either total cell lysates or samples from pulldown or immunoprecipitation, were loaded onto SDS-PAGE gels for immunoblotting as described before (17,26). Western blotting was performed with standard protocols using the PVDF membrane and signals visualized with ImageQuant LAS 4000 Luminescent Image Analyzer (GE Healthcare).

Immunofluorescence

Cells were fixed in 4% paraformaldehyde for 15 min at room temperature, followed by incubation in 1× PBS containing 0.25% of Triton X-100 for 10 min to permeabilize the cells. Fixed cells were co-stained with primary antibodies and then with Alexa-488/594 conjugated secondary antibodies (Life Technologies) as described before (17,26). Nuclei were stained with 4, 6-diamino-2-phenylindole (DAPI, 0.1 μg/ ml). Fluorescent signals were detected and images were taken by an FV1000 confocal microscope (Olympus, UNC Imaging Core).

Chromatin fractionation assay

Two million of cells were collected and re-suspended in 100 μl of ice-cold CSK buffer (10 mM PIPES, pH 7.0; 0.1% Triton X-100; 300 mM sucrose; 100 mM NaCl; 3 mM $MgCl_2$), followed by incubation on ice for 20 minutes. Samples were then centrifuged at 1,300g for 5 minutes at 4°C. The supernatant was collected as the soluble fraction and the pellet as the chromatin-associated fraction.

Compound

The catalytic inhibitor of EZH2 and EZH1, UNC1999, was dissolved in dimethylsulfoxide (DMSO) as 5 mM stock solution and used as previously described (27,28). The histone deacetylase (HDAC) inhibitor, CI-994, was used as we described before (17). The cell treatment condition for UNC1999 and CI-994 in this study was 5 μM as the final concentration for three and five days, respectively.

RNA-seq and Data Analysis

Total RNA was extracted using the RNeasy Mini Kit (QIAGEN) and RNA-seq libraries were prepared using the TruSeq RNA Sample Preparation Kit v2 (Illumina) as previously described (17,29). Multiplexed RNA-seq libraries were pooled and subjected for deep sequencing on an Illumina Hi-Seq 2500 or 4000 platform (UNC High-throughput Sequencing Facility). For data analysis, RNA-seq reads were first mapped using MapSplice (30) and quantified using RSEM (31). Read counts were then upper-quantile normalized and log2 transformed. Raw read counts were used for differential gene expression analysis by DESeq (32).

Gene Ontology (GO) analysis and Gene Set Enrichment Analysis (GSEA)

GO and GSEA analyses were conducted as previously described (27,29,33). GSEA was carried out by using the GSEA_4.0.3 software (34) and the Molecular Signatures Database v6.2 (MSigDB, http://www.broadinstitute.org) for testing putative enrichment of hallmark gene sets (H), curated gene sets (C2) and oncogenic gene sets (C6).

Chromatin immunoprecipitation (ChIP) followed by qPCR (ChIP-qPCR)

ChIP-qPCR was carried out as previously described (35,36). In brief, one million of cells were harvested and cross-linked with 1% of formaldehyde at room temperature for 10 minutes, followed by adding 2.5 M of glycine to quench crosslinking. Cells were sonicated by Bioruptor (Diagenode) and the cleared chromatin samples incubated with antibody-conjugated Dynabeads (Invitrogen) at 4 degree for overnight. Beads bound with chromatin were subject to extensive washing, followed by elution, de-crosslink overnight at 65 degree, and sequential digestion first by proteinase K (Thermo Fisher Scientific) and then RNase A (Roche). The ChIP DNA was then purified with MinElute PCR Purification Kit (QIAGEN). The samples of ChIP DNA and matched input DNA (similarly processed) were diluted in water and used for ChIP-qPCR. ChIP signals, generated from three independent experiments, were normalized to input signals and presented as mean

196 ± SD. Student's t-test was used for data analysis. The sequence of primers used for ChIP-
197 qPCR is available in supplementary table.

198 **ChIP followed by sequencing (ChIP-seq) and data analysis**

199 ChIP-seq was performed as previously described (27,29,33), with a multi-step procedure
200 that involves the cell fixation and sonication, antibody incubation, washing, ChIP DNA
201 recovery, library preparation, and deep sequencing (UNC High-throughput Sequencing
202 Facility). The standard data analysis of ChIP-Seq data alignment, filtration, peak calling
203 and assignment, and cross-sample comparison was performed as described before
204 (27,29,33) with slight modification. In brief, ChIP-seq reads were aligned to human
205 genome build GRCh37/hg19 using STAR version 2.7.1a (37). Alignment files in the bam
206 format were transformed into read coverage files (bigwig format) using DeepTools (38).
207 Genomic binding profiles were generated using the deepTools "bamCompare" functions
208 with options [--operation ratio ---pseudocount 1 -binSize 10 --extendReads 250] and
209 normalized to the matched input profile. The resulting bigWig files were visualized in the
210 Integrative Genome Viewer (IGV). Heatmaps were generated using the deepTools
211 "computeMatrix" and "plotHeatmap" functions. The genomic regions covering from the
212 transcription start site (TSS) to transcription end site (TES) of all genes annotated in the
213 human genome (NCBI Reference Sequences [RefSeq]; GRCh37/hg19) were obtained
214 from UCSC table browser and used for heatmap and averaged signal plotting.

215 **CUT&RUN followed by sequencing**

216 CUT&RUN (39) was performed with a commercially available kit according to
217 manufacturer's instruction (EpiCypher CUTANA™ pAG-MNase for ChIC/CUT&RUN,
218 Cat# 15-1116). Briefly, one million of cells were first collected, washed in the CUT&RUN
219 wash buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 1× Complete
220 Protease Inhibitor Cocktail), and then bound to the activated ConA beads (Bangs
221 Laboratories BP531). Next, the cell:bead sample was incubated with antibodies against
222 the protein target (1:100 dilution) and then permeabilized in the digitonin-containing buffer
223 (CUT&RUN wash buffer plus 0.01% digitonin), which was then followed by washing in the
224 digitonin buffer, incubation with pAG-MNase, and another washing in the digitonin buffer
225 to remove the unbound pAG-MNase. Following the final wash, cells were subjected to
226 digestion following the pAG-MNase activation by addition of the pAG-MNase digestion
227 buffer (digitonin buffer plus 2mM CaCl2), followed by incubation on nutator for 2 hours at
228 4ºC. Solubilized chromatin was then released using the CUT&RUN stop buffer (340 mM
229 NaCl, 20 mM EDTA, 4 mM EGTA, 50 µg/ml RNase A, 50 µg/ml glycogen) and DNA
230 purification was carried out with the PCR cleanup kit (NEB, Monarch PCR & DNA Cleanup
231 Kit, T1030). About 10ng of the purified CUT&RUN DNA was used for preparation of
232 multiplexed libraries with the NEB Ultra II DNA Library Prep Kit per manufacturer's
233 instruction (NEB #E7103). Sequencing was conducted using an Illumina NextSeq
234 500 Sequencing System (available from the core facility of UNC Pharmacology
235 Department).

236 **CUT&RUN data analysis**

237 Fastq files were mapped to the reference genome (hg19) using bowtie2.3.5 (40). The
238 non-primary alignment, PCR duplicates, or blacklist regions were removed from aligned
239 data by Samtools (v1.9), Picard "MarkDuplicates" function (ver 2.20.4), and bedtools
240 (v2.28.0), respectively. Peak calling was performed using MACS2 (macs2 callpeak -f
241 BAMPE -g hs/mm --keep-dup 1) (41) . Distribution of peaks was calculated by

242 “annotatePeaks.pl” function of HOMER (Hypergeometric Optimization of Motif
243 Enrichment) (42). Deeptools (v3.3.0) was used to generate bigwig files, heatmaps, and
244 averaged plotting of CUT&RUN signals (38). Genomic binding profiles were generated
245 using the deepTools “bamCompare” functions.

246 **Mass spectrometry–based analysis of BAHD1-containing protein complexes**

247 As previously described (17,26), nuclear extract was prepared using the Dignam protocol
248 from the nuclei of HEK293 cells that stably express 3xFLAG-tagged BAHD1, with parental
249 HEK293 cells as a negative control. Following dialysis against low-salt buffer (150 mM
250 NaCl, 20mM Hepes, pH 7.90, 0.025% IGEPAL® CA-630, 0.2 mM EDTA, 10% glycerol, 1
251 mM DTT, 0.2 mM PMSF), nuclear extracts were centrifuged at high speed to remove
252 precipitation and supernatants incubated with FLAG M2 beads (Sigma). After extensive
253 washing, the resin and bound proteins from both experimental and negative control cells
254 were subjected to mass spectrometry-based protein identification (carried out at
255 Proteomics Core facility, University of Arkansas for Medical Sciences).

256 **Generation of the *Bahd1^{W659G}* knock-in allele by CRISPR/cas9-mediated gene**
257 **editing in the C57BL/6J mouse strain**

258 Guide RNAs, which target the site close to the *Bahd1* W659 codon in exon 5, were
259 identified using Benchling software. Three guide RNAs were selected for initial activity
260 testing. Here, the selected guide RNAs were cloned into a T7 promoter vector followed
261 by in vitro transcription and spin column purification. Functional testing was performed by
262 cotransfecting a mouse embryonic fibroblast cell line with guide RNA and Cas9 protein.
263 Cas9 protein was expressed and purified by the UNC Protein Expression Core Facility.
264 The guide RNA target site was amplified from transfected cells and analyzed by T7
265 Endonuclease I digestion of re-annealed PCR products (New England Biolabs). The
266 guide RNA eventually selected for genome editing in embryos was Bahd1-sg47B
267 (protospacer sequence 5’- GCCTGTAATACCACAGG -3’). A donor oligonucleotide
268 (Bahd1-W659G-T), with the sequence of 5’-
269 CTAGCCATGACCTCTTCTCTTGTCCCCTAGGAGAGCTGATGATGAGCCTCCTG**GG**
270 **C**TATTACAGGCCTGAGCACTTACAGGGAGGTCGCAGCCCCAGCATGCATGAGG-3’,
271 was designed for homologous recombination to insert the TGG to GCC (bold/underlined
272 in oligo sequence) mutation of Tryptophan 659 to Glycine.

273 C57BL/6J zygotes were microinjected with 400 nM Cas9 protein, 20 ng/ul guide
274 RNA and 50 ng/ul donor oligonucleotide. In vitro transcribed, purified guide RNA was
275 diluted in microinjection buffer (5 mM Tris-HCl pH 7.5, 0.1 mM EDTA), heated at 95
276 degree for 3 min, and placed on ice prior to addition of Cas9 protein. The mixture was
277 then incubated at 37 degree for 5 min and placed on ice, after which the donor
278 oligonucleotide was added. The final mixture was filtered through a 0.2 uM filter and held
279 on ice prior to microinjection.

280 All animal experiments were approved by and performed in accord with the
281 guideline of Institutional Animal Care and Use Committee at the University of North
282 Carolina (UNC) at Chapel Hill, USA. Microinjected embryos were implanted in recipient
283 pseudopregnant females. Resulting pups were screened by PCR and sequencing for the
284 presence of the mutant allele. Of 17 pups born that survived to genotyping, 5 showed
285 clear evidence of the W659G mutant allele. Independent male founders were mated to
286 wild-type C57BL/6J females, and transmitted the mutant allele through the germline to
287 offspring.

Generation of primary mouse embryonic fibroblast (MEF) cells

The *Bahd1*^{W659G} heterozygous mice were set up for breeding and the E13.5 embryos harvested. The isolation and establishment of primary MEF cells were conducted as previously described (43). In brief, unwanted organs of embryos were removed, followed by tissue homogenization and enzymatic dissociation of cells. Genotypes of cells were determined by PCR and direct sequencing. Cells were then cultured in the DMEM base medium supplemented with 15% of FBS for 25 passages to establish stable MEF cell lines.

Statistical analysis

Experimental data are presented as mean ± SD of three independent experiments. Statistical analysis was performed with Student's t test for comparing two sets of data with assumed normal distribution. $P < 0.05$ was considered significant. Statistical significance levels are denoted as follows: *P<0.05; **P<0.01; ***P<0.001; ****P<0.0001. No statistical methods were used to predetermine sample size.

RESULTS

Specific interaction between the BAH domain of BAHD1 (BAHD1^{BAH}) and H3K27me3 is required for overall binding of BAHD1 to chromatin.

A conserved BAH module located at the C-terminus of BAHD1 (BAHD1^{BAH}; Supplementary Figure 1A) was shown to bind H3K27me3 directly (16), an interaction that we confirmed by immunofluorescence (IF) of H3K27me3 and 3xFLAG-tagged BAHD1 (f-BAHD1) in the HEK293 stable expression cells (Supplementary Figure 1B-C). When compared to BAHCC1^{BAH}, another H3K27me3-specific "reader" module (17), BAHD1^{BAH} displays considerable overall sequence divergence but retains conservation at so-called H3K27me3-"caging" residues (see stars in Supplementary Figure 1A). By CoIP, we further showed that BAHD1 associates with cellular H3K27me3 and not the other examined histone trimethylation, including H3K4me3, H3K9me3, H3K36me3, H3K79me3 and H4K20me3 (Figure 1A). To assess the requirement of such a BAHD1^{BAH}:H3K27me3 interaction for tethering BAHD1 to chromatin, we established the HEK293 stable cells that expressed the comparable level of BAHD1, either wildtype (WT) or with point mutation of a conserved "caging" residue within BAHD1^{BAH} (W667G or Y669A, see also stars in Supplementary Figure 1A; Figure 1B, top). Relative to WT, either "caging" mutation of BAHD1^{BAH} disrupted the H3K27me3 binding in the histone peptide pulldown assay (Figure 1B, bottom). Additionally, immuoblotting with the cell fractionation samples detected a significant decrease in the chromatin association by BAHD1 and a simultaneous increase of BAHD1 in the soluble fraction (Figure 1C). Thus, direct engagement of H3K27me3 by BAHD1^{BAH} is required for overall chromatin targeting of BAHD1 in cells.

BAHD1^{BAH} mediates BAHD1 targeting to the H3K27me3-marked Polycomb targets.

Next, we performed H3K27me3 ChIP-seq and mapped the genome-wide binding of f-BAHD1 with CUT&RUN in HEK293 cells. The called peaks of f-BAHD1 are mostly located at promoter, intergenic and intron regions (Figure 2A) whereas a negative control experiment of CUT&RUN by using parental HEK293 cells without f-BAHD1 expression

did not produce the enriched signals (Supplementary Figure 2A-B). There is a significant overlap between overall bindings of BAHD1 and H3K27me3 (Figure 2B-C). Heatmaps showed that genes with the more BAHD1 binding also tend to carry more H3K27me3 (Figure 2C), as exemplified by what was observed at the classic Polycomb target genes such as IGF2, FOXB1 and CCND2 (Figure 2D, the top and bottom tracks). Compared to WT, overall genome binding of the H3K27me3-binding-defective mutant, either BAHD1^{W667G} or BAHD1^{Y669A}, was found significantly reduced (Figure 2E and the two middle tracks of Figure 2D), in agreement with the decreased chromatin association of these BAH mutants (Figure 1C). These results lend strong support for a critical involvement of BAHD1^{BAH} in targeting BAHD1 to the H3K27me3-associated genomic loci.

BAHD1^{BAH} is critical for the BAHD1-mediated Polycomb gene repression in HEK293 cells.

We next sought to determine the requirement of an H3K27me3-binding BAHD1^{BAH} motif for regulation of target gene expression. Towards this end, we first identified siRNA that allowed an almost complete depletion of BAHD1 in HEK293 cells (Figure 3A; Figure 3B, lane 2 vs. 1). To assess the role of BAHD1^{BAH} per se, we next generated a BAHD1 cDNA for rescuing the endogenous BAHD1 knockdown (KD) due to its resistance to the used BAHD1-targeting siRNA (Figure 3C), as demonstrated by immunoblotting post-transfection of siRNA into cells with stable expression of such exogenous f-BAHD1, either WT or an H3K27me3-binding-defective mutant (Figure 3B, lanes 4, 6 and 8 versus lane 2). Utilizing such a validated rescue system, we subsequently performed RNA-seq profiling, which revealed significant transcriptomic alterations caused by the BAH mutation (W667G; Figure 3D and Supplementary Table 1). Relative to WT, BAHD1^{W667G} predominantly caused gene expression upregulation (Figure 3D, red). Both Gene Set Enrichment Analysis (GSEA; Figure 3E) and Platform for Integrative Analysis of Omics data (PIANO (44); Supplementary Figure 3A) further uncovered that, relative to WT, BAHD1^{W667G} is most significantly associated with de-repression of genesets previously known to be marked by H3K27me3, repressed by Polycomb factor (PRC2 or PRC1), or related to organismal development (Figure 3E-F and Supplementary Figure 3A-B). By RT-qPCR, we verified the marked reactivation of the tested Polycomb targets, which included a classic imprinted gene (IGF2), development genes (SOX8, WNT7B and MEIS3) and IGFBP2 (H3K27me3-marked), post-KD of endogenous BAHD1 relative to mock (Figure 3G, red vs. black); such a change was found to be reversed due to pre-rescue with exogenous WT BAHD1, but not either one of the H3K27me3-binding-defective mutants (BAHD1^{W667G} or BAHD1^{Y669A}; Figure 3G, cyan and blue vs. yellow and black). Meanwhile, neither depletion of BAHD1 nor expression of its mutant altered the protein level of Polycomb proteins or global H3K27me3 (Figure 3H and Supplementary Figure 3C). Based on genomic profiling data (CUT&RUN and RNA-seq), we have also defined those BAHD1/H3K27me3 co-bound target genes that exhibit a repression pattern in a BAHD1^{BAH}-dependent manner (Figure 3I and Supplementary Table 2), and this gene signature includes a number of classic Polycomb gene targets such as lineage-specific transcription factors and developmental regulators. Altogether, these observations highlight a role of direct BAHD1^{BAH}:H3K27me3 interaction in the BAHD1-related Polycomb gene silencing.

H3K27me3 and BAHD1 co-repress a set of Polycomb target genes in HEK293 cells. To further define genes repressed by H3K27me3 per se, we employed a catalytic inhibitor of PRC2, UNC1999 (27,28), to 'erase' global H3K27me3 in HEK293 cells (Figure 4A), followed by RNA-seq (Supplementary Table 3). UNC1999 mainly induced gene derepression (Figure 4B). As expected, these UNC1999-activated transcripts were most enriched with those previously known to be the H3K27me3-demarcated, Polycomb-targeted, and/or development-associated ones (Supplementary Figure 4A-B). Repressive effect by PRC2-catalyzed H3K27me3 was further verified by RT-qPCR of several Polycomb targets post-treatment of UNC1999 (Figure 4C). Importantly, differentially expressed genes (DEGs) reactivated by UNC1999 significantly overlapped with those derepressed due to the H3K27me3-interaction-disrupting mutation of BAHD1^{BAH}, relative to WT (Figure 4D). In consistence, UNC1999 treatment also significantly dissociated BAHD1 off the chromatin (Figure 4E). These results thus lent a strong support for an axis of H3K27me3→BAHD1^{BAH} →repression.

Disruption of the BAHD1^{BAH}:H3K27me3 interaction results in chromatin remodeling, notably the histone acetylation increase, at the BAHD1/H3K27me3 co-targeted genes.

We also performed f-BAHD1 pulldown in the HEK293 stable expression cells and the subsequent mass spectrometry-based analysis revealed a complex that comprises HDAC1, HDAC2 and MIER1/3 (Figure 5A and Supplementary Table 4), consistent with a previous report (18). Both CoIP and IF demonstrated a strong association of BAHD1 with HDACs (Figure 5B-C). Additionally, the aromatic "cage" mutations of BAHD1^{BAH} that disrupted direct H3K27me3 binding by BAHD1 did not affect the BAHD1:HDAC interaction (Figure 5D, WT versus WG or YA), indicative of different protein interfaces involved in the two interaction events. Notably, GSEA analyses of RNA-seq profiles also showed a positive correlation of HDAC target gene reactivation with BAHD1^{W667G} relative to WT (Figure 5E; Supplementary Figure 3A), as well as UNC1999 treatment relative to mock (Supplementary Figure 4A). By ChIP-qPCR, we detected the HDAC2 binding at the examined BAHD1 target genes (Figure 5F, red vs black). Such binding of HDAC2 to BAHD1 targets was found to be significantly reduced in HEK293 cells expressing the BAH mutant of BAHD1 (W667G or Y669A), compared to WT (Figure 5F, blue/yellow vs red). Treatment of HEK293 cells with the HDAC inhibitor CI-994 led to significant reactivation of BAHD1 target genes (Figure 5G). These observations highlight an important role for the BAHD1:HDAC complex in mediating Polycomb gene silencing.

Next, we aimed to define chromatin changes caused by the disrupted interaction between BAHD1^{BAH} and H3K27me3. Towards this end, ChIP-seq was performed for both gene-activation-related histone acetylation (H3K27ac and H3K9ac) and gene-repression-related histone methylation (H3K27me3 and H3K9me3) in the BAHD1-depleted HEK293 cells that were rescued with exogenous BAHD1, WT or BAHD1^{W667G} (Supplementary Figure 5A). At the BAHD1/H3K27me3 co-targeted genes, overall H3K27ac was found increased in cells expressing BAHD1^{W667G}, compared to WT, whereas overall H3K27me3 at these regions was markedly decreased (Figure 6A-B). Meanwhile, while H3K9ac showed a mild increase (Figure 6A-B), H3K9me3 was found to be generally non-existing or very low at these BAHD1-targeted genes, especially TSS-proximal regions (Figure 6A, the bottom panel). Using global ChIP-seq profiles, we also

426 observed a closer correlation between BAHD1 and H3K27me3, in comparison to
427 H3K9me3 (Supplementary Figure 5B). The remodeling of histone modifications (i.e., gain
428 of H3K9ac/K27ac and concurrent reduction of H3K27me3), as well as a lack of H3K9me3,
429 at the BAHD1-bound genomic sites are more evident at a set of BAHD1-repressed genes
430 such as IGF2, TGFB1, SOX8, FOXB1 and CCND1 (Figure 6C and Supplementary Figure
431 5C). These observations also suggest that H3K27me3, rather than H3K9me3, is more
432 directly involved in targeting BAHD1 to these Polycomb-targeted loci, in contrast to the
433 previous working model (15,18). ChIP-qPCR further verified the increase of histone
434 acetylation at promoters of several BAHD1 target genes (such as IGF2), but not a control
435 house-keeping gene (GAPDH), in cells with BAHD1^{W667G} versus WT (Supplementary
436 Figure 5D). Collectively, we demonstrate that direct binding of H3K27me3 by BAHD1^{BAH}
437 is essential for maintaining a repressive state of histone deacetylation at its Polycomb
438 gene targets.
439
440 **A germline mutation of *Bahd1^{W659G}* causes severe neonatal lethality in mice.**
441 BAHD1 is essential for animal development. Its global knockout (KO) in mice led to
442 defects in placental and fetal growth (18) whereas the *Bahd1*+/- heterozygous mice
443 showed an anxiety-like behavior (45), indicating a pleiotropic function of BAHD1 among
444 various tissue types and cell lineages. To further assess potential involvement of the
445 BAHD1^{BAH}:H3K27me3 interaction during normal development, we generated mice
446 harboring a germline mutation of *Bahd1^{W659G}* (Supplementary Figure 6A), equivalent to
447 human *BAHD1^{W667G}*. Breeding of the *Bahd1^{W659G}* heterozygous mice produced only 2.7%
448 (2 out of a total of 74) of the *Bahd1^{W659G}* homozygous pups at birth (Figure 7A), which is
449 significantly lower than the expected frequency among both male and female pups
450 (Figure 7B). Additionally, a few of the surviving *Bahd1^{W659G}* homozygous mice exhibited
451 a dwarfism phenotype, compared to the matched WT or heterozygous littermates (Figure
452 7C-D and Supplementary Figure 6B). Using such a murine model, we further established
453 primary MEF cells that carry either WT *Bahd1* or *Bahd1^{W659G}* homozygous alleles. The
454 subsequent RNA-seq profiling revealed the significant enrichment of Polycomb and
455 H3K27me3 target genes among transcripts upregulated in MEFs with Bahd1^{W659G},
456 relative to WT (Figure 7E-F, Supplementary Figure 6C-D, and Supplementary Table 5),
457 which is in agreement with what was observed in the HEK293 cell model. These results
458 thus demonstrate a requirement of the BAHD1^{BAH}-mediated H3K27me3 interaction for
459 early embryogenesis and appropriate postnatal development, which was previously
460 unreported.
461
462
463 **DISCUSSION**
464
465 It was postulated that chromatin modification such as H3K27me3 serves as a context-
466 dependent docking site for recruitment of the so-called reader/effector and associated
467 protein complexes, which impart their respective gene- and genome-regulatory effect
468 (5,46,47). While it is viewed that mammalian cPRC1 binds to H3K27me3 for mediating
469 Polycomb gene silencing, such a pathway is generally absent from the plant and fungus
470 (19,20,22,48). A flurry of recent studies have collectively demonstrated that H3K27me3
471 is also recognized and directly bound by a set of BAH modules that exist among various

organisms covering from Arabidopsis thaliana (19-21,23,24), animal (15-17) and fungus (Neurospora crassa) (22), suggesting a biological significance of the BAH-based functional transduction of H3K27me3. Here, we reported the crucial role for BAHD1^{BAH} in the mammalian Polycomb gene silencing and animal development, which was not explored before. Indeed, although BAHD1^{BAH} was shown to bind H3K27me3 directly (16), it was previously proposed that BAHD1 acts as a multifunctional "scaffold" and its gene-repressive effect involves its associations with H3K9me3 readers and/or other related corepressors (15,18). Through extensive genomic profiling (ChIP-seq, CUT&RUN and RNA-seq) and BAHD1^{BAH} mutagenesis (using 293 cells and a mutation knockin murine model and derived MEF cells), we demonstrate that (i) an intact H3K27me3-engaging module (BAHD1^{BAH}) is required for overall BAHD1 binding onto chromatin targets (as shown by chromatin fractionation/association assay and CUT&RUN of BAHD1, WT versus BAH-mutated), (ii) that the direct interaction between BAHD1^{BAH} and H3K27me3 is essential for silencing of a subset of the H3K27me3-demarcated Polycomb targets, which include developmental transcription factors (such as FOXB1 and MEIS3) and imprinted gene (such as IGF2); (iii) that the direct H3K27me3 binding by BAHD1^{BAH} is also required for maintaining histone deacetylation at its Polycomb targets, in agreement with the observed BAHD1:HDAC interaction; and lastly, (iv) that abolishing the BAHD1^{BAH}:H3K27me3 engagement causes severe developmental defects in mice such as embryonic lethality and dwarfism, consistent to the known requirement of Polycomb factors for organismal development and cell lineage specification. Notably, the observed chromatin remodeling due to mutation of BAHD1^{BAH} includes increase of H3K27ac/H3K9ac and decrease of H3K27me3. For the latter, it is known that H3K27ac directly antagonizes the PRC2-mediated methylation of the same site and that, following gene reactivation, the productive transcriptional elongation leads to accumulation of the SETD2-mediated H3K36me3, which also suppresses H3K27me3 formation in cis (49).

A previously proposed model emphasizes a multifaceted interaction of BAHD1 with the H3K9me3-specific readers such as HP1 or TRIM28/KAP1 and other repressive partners such as the H3K9 methyltransferase (15,18). Here, we observed that H3K9me3 is generally absent or low at the BAHD1-bound genomic sites, especially TSS-proximal regions (see the bottom panel of Figure 6A and examples in Figure 6C and Supplementary Figure 5C). Based on these results and various phenotypes caused by the H3K27me3-binding-defective mutation of BAHD1^{BAH}, we thus favor a view that the direct binding of H3K27me3 by BAHD1^{BAH} provides a primary mechanism for tethering BAHD1 onto its Polycomb gene targets. Besides the H3K27me3-binding BAH motif, BAHD1 can use a N-terminal region as an interface for protein-protein interaction (50) and may form interactions with other gene-repressive pathways of H3K9me3 (15) or DNA methylation (51). Please note that the most frequent and consistent hits of BAHD1 interactors, as identified previously (18) and from this study (Supplementary Table 4), include at least MIER3, MIER1, HDAC1, HDAC2 and TRIM28/KAP1 (in a relatively less frequency), indicating that this set of cofactors form a more stable "core" complex with BAHD1. Please also note that we detected existence of H3K9me3 several to dozens of Kb away from a subset of those BAHD1-bound regions (as exemplified by IGF2, FOXB1 and CCND1 shown in Figure 6C and Supplementary Figure 5C); therefore, putative involvement of TRIM28/KAP1, a BAHD1 partner and H3K9me3-specific reader, is possible at certain BAHD1 targets— for example, it is conceivable that, following initial

518 tethering of BAHD1 to the TSS-proximal regions marked by H3K27me3, a long-range
519 looping interaction between TSS and the distal H3K9me3-demarcated region may form,
520 due to physical association between BAHD1 and the H3K9me3-bound TRIM28/KAP1
521 (15,18), which possibly mediates further stabilization and/or spreading of the BAHD1
522 complex on chromatin. While further studies are warranted to delineate the molecular
523 details, the BAHD1-induced heterochromatin formation is likely to be context-dependent
524 (such as chromatin environment and interactors) and involves a multilevel mechanism, of
525 which the direct binding of H3K27me3 by BAHD1^{BAH} represents one of the critical,
526 possibly primary, events.
527 A significant finding of this report lies in a clear demonstration of BAHD1^{BAH} as a
528 functional H3K27me3 reader in the mammalian cells, which acts to maintain appropriate
529 Polycomb gene-silencing program for sustaining normal development. Considering the
530 existence of cPRC1 and recent identification of BAHCC1^{BAH} as H3K27me3-specific
531 readers (17), the chromodomain (cPRC1)- and BAH-mediated functional readout of
532 H3K27me3 can coexist in mammalian cells, imparting their respective gene-silencing
533 effects through distinctive molecular pathways. Besides a catalytic function for inducing
534 H2AK119ub1, cPRC1 induces chromatin compaction via cofactor-mediated phase
535 separation and condensation formation (12-14). On the other hand, BAHCC1 (17) and
536 BAHD1 both interact with HDACs, thus linking H3K27me3 together with histone
537 deacetylation, an integral step of gene repression. Conceivably, the chromodomain- and
538 BAH-containing effectors of H3K27me3 may potentially cooperate, synergize and/or
539 compete, the detail of which awaits additional analysis. It is also possible that, due to
540 differential expression patterns, their overall contributions to Polycomb gene silencing
541 vary among various cell lineages and tissue types. The mRNA levels of Bahd1 and
542 Bahcc1 are generally lower than those of cPRC1 genes in embryonic stem cells (data not
543 shown) whereas their expression can be more comparable in other cell types such as
544 293 cells. Overall, the BAH module–based transduction of H3K27me3 is not only
545 functionally important but also evolutionarily ancient, existing in the fungus, animal and
546 plant.
547 Over dozens of the chromatin-associated proteins from different organisms carry
548 a BAH motif (52). It is noteworthy to mention that these different BAH domains
549 demonstrate functional divergence. In contrast to a set of the H3K27me3-"reading" BAHs
550 mentioned above in fungus, plant and animal, the BAH domain of ORC1 preferentially
551 recognizes histone H4 lysine 20 dimethylation (H4K20me2) in metazoan species (53). In
552 maize, a methyltransferase termed ZMET2 uses its BAH domain to recognize and bind
553 H3K9me2, thereby establishing a crosstalk between the repressive histone modification
554 and DNA methylation (54). In yeast, the BAH motif of Rsc2 was shown to bind
555 recombinant H3 in vitro whereas that of Sir3 contacts with nucleosome to induce
556 heterochromatin formation (55). The molecular basis underlying the observed functional
557 divergence among BAH modules remains to be fully characterized. Importantly,
558 deregulation of the BAH-containing proteins is often associated with pathogenesis. For
559 instance, we reported that germ-line mutation of either Bahcc1^{BAH} (17) or Bahd1^{BAH} in
560 mice caused embryonic lethality and dwarfism. The germ-line mutation of ORC1's
561 H4K20me2-binding BAH domain is responsible for Meier-Gorlin syndrome, a form of
562 primordial dwarfism (53,56). Somatic mutation at the BAH domains of Polybromo-1 was

detected in clear cell renal carcinoma (57). Altogether, BAH evolves as a versatile module family critically involved in various biological processes, which merits further studies.

DATA AVAILABILITY

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors. Raw and processed NGS data have been deposited with the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) under accession number GSE160354.

SUPPLEMENTARY DATA STATEMENT

Supplementary Data are available at NAR Online.

FUNDING

This work was supported by National Institutes of Health grants (R01-CA215284 and R01-CA218600 to G.G.W.), grants of Gilead Sciences Research Scholars Program in haematology/oncology (to G.G.W.), When Everyone Survives (WES) Leukemia Research Foundation (to G.G.W.) and the UCRF Stimulus Initiative Grant of UNC Lineberger Comprehensive Cancer Center (to L.C.). We also acknowledge proteomics support from the University of Arkansas for Medical Sciences Proteomics Core Facility, the IDeA National Resource for Proteomics, the Translational Research Institute through the National Center for Advancing Translational Sciences of the National Institutes of Health, and the Center for Translational Pediatric Research Bioinformatics Core Resource with support through TL1TR003109, P20GM121293, P20GM103625, P20GM103429, S10OD018445, UL1TR003107, and R01CA236209. The cores affiliated to UNC Cancer Center are supported in part by the UNC Lineberger Comprehensive Cancer Center Core Support Grant P30-CA016086. G.G.W. is an American Cancer Society (ACS) Research Scholar and a Leukemia and Lymphoma Society (LLS) Scholar.

Conflict of interest statement.

G.G.W. is inventor of a patent application filed by University of North Carolina at Chapel Hill. G.G.W. received research fund from the Deerfield Management/Pinnacle Hill Company.

ACKNOWLEDGMENTS

We graciously thank H. Bierne for providing the reagent and the Wang Laboratory members, T. Ptacek, and J. Simon for helpful discussion and technical support. We thank UNC's facilities, including Imaging Core, High-throughput Sequencing Facility (HTSF), Bioinformatics Core, Animal Models Core Facility and Tissue Culture Facility for their professional assistance of this work. The cores affiliated to UNC Cancer Center are supported in part by the UNC Lineberger Comprehensive Cancer Center Core Support Grant P30-CA016086.

Author contributions

H.F., Y.G., A.R.K. and L.C., performed experiments. Y.T., W.G., and Y.G., performed genomic data analysis under direction of G.G.W. A.S., R.E., S.B., S.M., and A.T.

609 conducted and analyzed mass spectrometry experiments. H.F., Y.G., H.L., L.C., A.T. and
610 G.G.W. interpreted the data. G.G.W. conceived the project. H.F., Y.G., and G.G.W.
611 prepared the manuscript with inputs of all other authors.
612
613
614 **Reference:**

615 1. Yu, J.R., Lee, C.H., Oksuz, O., Stafford, J.M. and Reinberg, D. (2019) PRC2 is high
616 maintenance. *Genes & development*, **33**, 903-935.
617 2. Di Croce, L. and Helin, K. (2013) Transcriptional regulation by Polycomb group proteins.
618 *Nature structural & molecular biology*, **20**, 1147-1155.
619 3. Blackledge, N.P., Rose, N.R. and Klose, R.J. (2015) Targeting Polycomb systems to
620 regulate gene expression: modifications to a complex story. *Nat Rev Mol Cell Biol*, **16**,
621 643-649.
622 4. Chammas, P., Mocavini, I. and Di Croce, L. (2020) Engaging chromatin: PRC2 structure
623 meets function. *Br J Cancer*, **122**, 315-328.
624 5. Guo, Y., Zhao, S. and Wang, G.G. (2021) Polycomb Gene Silencing Mechanisms: PRC2
625 Chromatin Targeting, H3K27me3 'Readout', and Phase Separation-Based Compaction.
626 *Trends Genet*.
627 6. Schwartzentruber, J., Korshunov, A., Liu, X.Y., Jones, D.T., Pfaff, E., Jacob, K., Sturm,
628 D., Fontebasso, A.M., Quang, D.A., Tonjes, M. *et al.* (2012) Driver mutations in histone
629 H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature*, **482**, 226-231.
630 7. Lewis, P.W., Muller, M.M., Koletsky, M.S., Cordero, F., Lin, S., Banaszynski, L.A.,
631 Garcia, B.A., Muir, T.W., Becher, O.J. and Allis, C.D. (2013) Inhibition of PRC2 activity
632 by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science (New York,*
633 *N.Y*, **340**, 857-861.
634 8. Xu, B., Konze, K.D., Jin, J. and Wang, G.G. (2015) Targeting EZH2 and PRC2 dependence
635 as novel anticancer therapy. *Exp Hematol*, **43**, 698-712.
636 9. Cyrus, S., Burkardt, D., Weaver, D.D. and Gibson, W.T. (2019) PRC2-complex related
637 dysfunction in overgrowth syndromes: A review of EZH2, EED, and SUZ12 and their
638 syndromic phenotypes. *Am J Med Genet C Semin Med Genet*.
639 10. Min, J., Zhang, Y. and Xu, R.M. (2003) Structural basis for specific binding of Polycomb
640 chromodomain to histone H3 methylated at Lys 27. *Genes & development*, **17**, 1823-1828.
641 11. Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S. and
642 Zhang, Y. (2004) Role of histone H2A ubiquitination in Polycomb silencing. *Nature*, **431**,
643 873-878.
644 12. Plys, A.J., Davis, C.P., Kim, J., Rizki, G., Keenen, M.M., Marr, S.K. and Kingston, R.E.
645 (2019) Phase separation of Polycomb-repressive complex 1 is governed by a charged
646 disordered region of CBX2. *Genes Dev*, **33**, 799-813.
647 13. Tatavosian, R., Kent, S., Brown, K., Yao, T., Duc, H.N., Huynh, T.N., Zhen, C.Y., Ma, B.,
648 Wang, H. and Ren, X. (2019) Nuclear condensates of the Polycomb protein chromobox 2
649 (CBX2) assemble through phase separation. *J Biol Chem*, **294**, 1451-1463.
650 14. Seif, E., Kang, J.J., Sasseville, C., Senkovich, O., Kaltashov, A., Boulier, E.L., Kapur, I.,
651 Kim, C.A. and Francis, N.J. (2020) Phase separation by the polyhomeotic sterile alpha
652 motif compartmentalizes Polycomb Group proteins and enhances their activity. *Nat*
653 *Commun*, **11**, 5609.

654 15. Bierne, H., Tham, T.N., Batsche, E., Dumay, A., Leguillou, M., Kerneis-Golsteyn, S.,
655 Regnault, B., Seeler, J.S., Muchardt, C., Feunteun, J. *et al.* (2009) Human BAHD1
656 promotes heterochromatic gene silencing. *Proceedings of the National Academy of*
657 *Sciences of the United States of America*, **106**, 13826-13831.
658 16. Zhao, D., Zhang, X., Guan, H., Xiong, X., Shi, X., Deng, H. and Li, H. (2016) The BAH
659 domain of BAHD1 is a histone H3K27me3 reader. *Protein Cell*, **7**, 222-226.
660 17. Fan, H., Lu, J., Guo, Y., Li, D., Zhang, Z.M., Tsai, Y.H., Pi, W.C., Ahn, J.H., Gong, W.,
661 Xiang, Y. *et al.* (2020) BAHCC1 binds H3K27me3 via a conserved BAH module to
662 mediate gene silencing and oncogenesis. *Nature genetics*, **52**, 1384-1396.
663 18. Lakisic, G., Lebreton, A., Pourpre, R., Wendling, O., Libertini, E., Radford, E.J., Le
664 Guillou, M., Champy, M.F., Wattenhofer-Donze, M., Soubigou, G. *et al.* (2016) Role of
665 the BAHD1 Chromatin-Repressive Complex in Placental Development and Regulation of
666 Steroid Metabolism. *PLoS Genet*, **12**, e1005898.
667 19. Yang, Z., Qian, S., Scheid, R.N., Lu, L., Chen, X., Liu, R., Du, X., Lv, X., Boersma, M.D.,
668 Scalf, M. *et al.* (2018) EBS is a bivalent histone reader that regulates floral phase transition
669 in Arabidopsis. *Nature genetics*, **50**, 1247-1253.
670 20. Li, Z., Fu, X., Wang, Y., Liu, R. and He, Y. (2018) Polycomb-mediated gene silencing by
671 the BAH-EMF1 complex in plants. *Nature genetics*, **50**, 1254-1261.
672 21. Qian, S., Lv, X., Scheid, R.N., Lu, L., Yang, Z., Chen, W., Liu, R., Boersma, M.D., Denu,
673 J.M., Zhong, X. *et al.* (2018) Dual recognition of H3K4me3 and H3K27me3 by a plant
674 histone reader SHL. *Nat Commun*, **9**, 2425.
675 22. Wiles, E.T., McNaught, K.J., Kaur, G., Selker, J.M.L., Ormsby, T., Aravind, L. and Selker,
676 E.U. (2020) Evolutionarily ancient BAH-PHD protein mediates Polycomb silencing. *Proc*
677 *Natl Acad Sci U S A*, **117**, 11614-11623.
678 23. Zhang, Y.Z., Yuan, J., Zhang, L., Chen, C., Wang, Y., Zhang, G., Peng, L., Xie, S.S., Jiang,
679 J., Zhu, J.K. *et al.* (2020) Coupling of H3K27me3 recognition with transcriptional
680 repression through the BAH-PHD-CPL2 complex in Arabidopsis. *Nat Commun*, **11**, 6212.
681 24. Qian, F., Zhao, Q.Y., Zhang, T.N., Li, Y.L., Su, Y.N., Li, L., Sui, J.H., Chen, S. and He,
682 X.J. (2021) A histone H3K27me3 reader cooperates with a family of PHD finger-
683 containing proteins to regulate flowering time in Arabidopsis. *J Integr Plant Biol*.
684 25. Yu, M., Selvaraj, S.K., Liang-Chu, M.M., Aghajani, S., Busse, M., Yuan, J., Lee, G., Peale,
685 F., Klijn, C., Bourgon, R. *et al.* (2015) A resource for cell line authentication, annotation
686 and quality control. *Nature*, **520**, 307-311.
687 26. Cai, L., Rothbart, S.B., Lu, R., Xu, B., Chen, W.Y., Tripathy, A., Rockowitz, S., Zheng,
688 D., Patel, D.J., Allis, C.D. *et al.* (2013) An H3K36 methylation-engaging Tudor motif of
689 polycomb-like proteins mediates PRC2 complex targeting. *Mol Cell*, **49**, 571-582.
690 27. Xu, B., On, D.M., Ma, A., Parton, T., Konze, K.D., Pattenden, S.G., Allison, D.F., Cai, L.,
691 Rockowitz, S., Liu, S. *et al.* (2015) Selective inhibition of EZH2 and EZH1 enzymatic
692 activity by a small molecule suppresses MLL-rearranged leukemia. *Blood*, **125**, 346-357.
693 28. Konze, K.D., Ma, A., Li, F., Barsyte-Lovejoy, D., Parton, T., Macnevin, C.J., Liu, F., Gao,
694 C., Huang, X.P., Kuznetsova, E. *et al.* (2013) An orally bioavailable chemical probe of the
695 Lysine Methyltransferases EZH2 and EZH1. *ACS Chem Biol*, **8**, 1324-1334.
696 29. Cai, L., Tsai, Y.H., Wang, P., Wang, J., Li, D., Fan, H., Zhao, Y., Bareja, R., Lu, R.,
697 Wilson, E.M. *et al.* (2018) ZFX Mediates Non-canonical Oncogenic Functions of the
698 Androgen Receptor Splice Variant 7 in Castrate-Resistant Prostate Cancer. *Molecular cell*.

699 30. Wang, K., Singh, D., Zeng, Z., Coleman, S.J., Huang, Y., Savich, G.L., He, X.,
700 Mieczkowski, P., Grimm, S.A., Perou, C.M. *et al.* (2010) MapSplice: accurate mapping of
701 RNA-seq reads for splice junction discovery. *Nucleic acids research*, **38**, e178.
702 31. Li, B. and Dewey, C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq
703 data with or without a reference genome. *BMC Bioinformatics*, **12**, 323.
704 32. Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data.
705 *Genome Biol*, **11**, R106.
706 33. Lu, R., Wang, P., Parton, T., Zhou, Y., Chrysovergis, K., Rockowitz, S., Chen, W.Y.,
707 Abdel-Wahab, O., Wade, P.A., Zheng, D. *et al.* (2016) Epigenetic Perturbations by
708 Arg882-Mutated DNMT3A Potentiate Aberrant Stem Cell Gene-Expression Program and
709 Acute Leukemia Development. *Cancer cell*, **30**, 92-107.
710 34. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A.,
711 Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. *et al.* (2005) Gene set enrichment
712 analysis: a knowledge-based approach for interpreting genome-wide expression profiles.
713 *Proceedings of the National Academy of Sciences of the United States of America*, **102**,
714 15545-15550.
715 35. Wang, G.G., Cai, L., Pasillas, M.P. and Kamps, M.P. (2007) NUP98-NSD1 links H3K36
716 methylation to Hox-A gene activation and leukaemogenesis. *Nat Cell Biol*, **9**, 804-812.
717 36. Wang, G.G., Song, J., Wang, Z., Dormann, H.L., Casadio, F., Li, H., Luo, J.L., Patel, D.J.
718 and Allis, C.D. (2009) Haematopoietic malignancies caused by dysregulation of a
719 chromatin-binding PHD finger. *Nature*, **459**, 847-851.
720 37. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P.,
721 Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner.
722 *Bioinformatics*, **29**, 15-21.
723 38. Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S.,
724 Dundar, F. and Manke, T. (2016) deepTools2: a next generation web server for deep-
725 sequencing data analysis. *Nucleic acids research*, **44**, W160-165.
726 39. Skene, P.J., Henikoff, J.G. and Henikoff, S. (2018) Targeted in situ genome-wide profiling
727 with high efficiency for low cell numbers. *Nat Protoc*, **13**, 1006-1019.
728 40. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-
729 efficient alignment of short DNA sequences to the human genome. *Genome Biol*, **10**, R25.
730 41. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum,
731 C., Myers, R.M., Brown, M., Li, W. *et al.* (2008) Model-based analysis of ChIP-Seq
732 (MACS). *Genome Biol*, **9**, R137.
733 42. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre,
734 C., Singh, H. and Glass, C.K. (2010) Simple combinations of lineage-determining
735 transcription factors prime cis-regulatory elements required for macrophage and B cell
736 identities. *Molecular cell*, **38**, 576-589.
737 43. Garfield, A.S. (2010) Derivation of primary mouse embryonic fibroblast (PMEF) cultures.
738 *Methods Mol Biol*, **633**, 19-27.
739 44. Varemo, L., Nielsen, J. and Nookaew, I. (2013) Enriching the gene set analysis of genome-
740 wide data by incorporating directionality of gene expression and combining statistical
741 hypotheses and methods. *Nucleic acids research*, **41**, 4378-4391.
742 45. Pourpre, R., Naudon, L., Meziane, H., Lakisic, G., Jouneau, L., Varet, H., Legendre, R.,
743 Wendling, O., Selloum, M., Proux, C. *et al.* (2020) BAHD1 haploinsufficiency results in
744 anxiety-like phenotypes in male mice. *PLoS ONE*, **15**, e0232789.

745 46. Strahl, B.D. and Allis, C.D. (2000) The language of covalent histone modifications.
746 *Nature*, **403**, 41-45.
747 47. Chi, P., Allis, C.D. and Wang, G.G. (2010) Covalent histone modifications--miswritten,
748 misinterpreted and mis-erased in human cancers. *Nat Rev Cancer*, **10**, 457-469.
749 48. Schuettengruber, B., Bourbon, H.M., Di Croce, L. and Cavalli, G. (2017) Genome
750 Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell*, **171**, 34-57.
751 49. Schmitges, F.W., Prusty, A.B., Faty, M., Stutzer, A., Lingaraju, G.M., Aiwazian, J., Sack,
752 R., Hess, D., Li, L., Zhou, S. *et al.* (2011) Histone methylation by PRC2 is inhibited by
753 active chromatin marks. *Molecular cell*, **42**, 330-341.
754 50. Lebreton, A., Job, V., Ragon, M., Le Monnier, A., Dessen, A., Cossart, P. and Bierne, H.
755 (2014) Structural basis for the inhibition of the chromatin repressor BAHD1 by the
756 bacterial nucleomodulin LntA. *MBio*, **5**, e00775-00713.
757 51. Libertini, E., Lebreton, A., Lakisic, G., Dillies, M.A., Beck, S., Coppee, J.Y., Cossart, P.
758 and Bierne, H. (2015) Overexpression of the Heterochromatinization Factor BAHD1 in
759 HEK293 Cells Differentially Reshapes the DNA Methylome on Autosomes and X
760 Chromosome. *Front Genet*, **6**, 339.
761 52. Yang, N. and Xu, R.M. (2013) Structure and function of the BAH domain in chromatin
762 biology. *Crit Rev Biochem Mol Biol*, **48**, 211-221.
763 53. Kuo, A.J., Song, J., Cheung, P., Ishibe-Murakami, S., Yamazoe, S., Chen, J.K., Patel, D.J.
764 and Gozani, O. (2012) The BAH domain of ORC1 links H4K20me2 to DNA replication
765 licensing and Meier-Gorlin syndrome. *Nature*, **484**, 115-119.
766 54. Du, J., Zhong, X., Bernatavichute, Y.V., Stroud, H., Feng, S., Caro, E., Vashisht, A.A.,
767 Terragni, J., Chin, H.G., Tu, A. *et al.* (2012) Dual binding of chromomethylase domains to
768 H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell*, **151**, 167-180.
769 55. Chambers, A.L., Pearl, L.H., Oliver, A.W. and Downs, J.A. (2013) The BAH domain of
770 Rsc2 is a histone H3 binding domain. *Nucleic Acids Res*, **41**, 9168-9182.
771 56. Bicknell, L.S., Walker, S., Klingseisen, A., Stiff, T., Leitch, A., Kerzendorfer, C., Martin,
772 C.A., Yeyati, P., Al Sanna, N., Bober, M. *et al.* (2011) Mutations in ORC1, encoding the
773 largest subunit of the origin recognition complex, cause microcephalic primordial
774 dwarfism resembling Meier-Gorlin syndrome. *Nature genetics*, **43**, 350-355.
775 57. Porter, E.G. and Dykhuizen, E.C. (2017) Individual Bromodomains of Polybromo-1
776 Contribute to Chromatin Association and Tumor Suppression in Clear Cell Renal
777 Carcinoma. *The Journal of biological chemistry*, **292**, 2601-2610.
778
779

Figure legends

Figure 1. H3K27me3 binding by the BAHD1 BAH module (BAHD1^{BAH}) contributes significantly to overall chromatin association of BAHD1.

(A) Co-IP between FLAG-tagged BAHD1 (f-BAHD1) and the indicated histone methylation in HEK293 cells.

(**B**) Top: anti-FLAG immunoblotting using total cell extracts of HEK293 cells, either parental or those with stable expression of empty vector (EV) or f-BAHD1, WT or BAH-mutated (W667G or Y669A).

Bottom: Pulldown using the biotinylated histone H3 peptide, either unmodified or harboring H3K27me3, and total cell extracts that contain WT or BAH-mutated f-BAHD1, followed by anti-FLAG immunoblot.

(**C**) Immunoblotting for f-BAHD1 using the chromatin-bound (right) and soluble (left) fractions prepared from the HEK293 cells expressing either WT or BAH-mutated (W667G or Y669A) f-BAHD1. GAPDH and H3 served as the control of fractionation.

Figure 2. BAHD1^{BAH}-mediated H3K27me3 engagement is required for BAHD1's recruitment to Polycomb target genes.

(A) Pie chart showing distribution of the called f-BAHD1 CUT&RUN peaks among the indicated genomic features in HEK293 cells.

(B-C) Plotting of averaged intensity **(B)** and heatmaps **(C)** of f-BAHD1 CUT&RUN and H3K27me3 ChIP-seq signals across ±5 kb from transcription start site (TSS) and transcription end site (TES) of target genes in HEK293 cells.

(**D**) IGV views showing the called peaks of WT or BAH-mutated (W667G or Y669A) f-BAHD1, as well as H3K27me3, at the indicated gene in HEK293 cells.

(E) Averaged CUT&RUN signals (y-axis) of the indicated WT and mutant f-BAHD1 across ±5 kb from the peak center (x-axis) in HEK293 stable expression cells.

Figure 3. Transcriptomic profiling demonstrates an essential involvement of BAHD1^{BAH} in BAHD1-induced gene repression in HEK293 cells.

(A) RT-qPCR of endogenous *BAHD1* after siRNA-mediated knockdown (KD) in HEK293 cells (n = 3 biologically independent samples). Two different sets of BAHD1 primers were used. Data are presented as mean ± SD.

(**B**) Immunoblotting of exogenous f-BAHD1, either WT or with a mutation of BAHD1^{BAH} (W667G or Y669A), post-transduction of the BAHD1-targeting siRNA (siBAHD1) or scramble control (siCtl). Left and right panels showed the HEK293 cell lines with stable expression of f-BAHD1 that was efficiently depleted by siBAHD1 (left) or designed to be siBAHD1-resistant ("si-r"; right, see also panel C).

(**C**) The BAHD1 cDNA sequence targeted by the used BAHD1-targeting siRNA (siBAHD1 as shown in blue on the top; see also lanes 2 vs 1 in panel B), and a designed siBAHD1-resistant BAHD1 form that harbors seven silent point mutations of the siBAHD1-targeting site (bottom; see also lanes 3-8 of panel B). Each box indicates a codon.

(**D**) Heatmap showing the relative expression of differentially expressed genes (DEGs) identified in HEK293 cells with rescue of WT BAHD1, compared to its W667G mutant, followed by endogenous BAHD1 KD by siBAHD1 (n=3 biological replicates per group). The thresholds of DEGs are adjusted *P* value (adj. p) < 0.01, fold-change (FC) >2.0 and transcript baseMean

826 value > 10 based on RNA-seq. Color bar, log2(FC).
827 (**E**) Summary of GSEA using RNA-seq profiles of HEK293 cells that expressed an H3K27me3-
828 binding-defective mutant of BAHD1 (W667G), compared to WT. NES, normalized enrichment
829 score.
830 (**F**) GSEA revealing that, relative to WT, the BAH mutation (W667G) of BAHD1 is positively
831 correlated to reactivation of the indicated Polycomb gene signature associated with H3K27me3,
832 PRC2 (EZH2) or PRC1 (BMI1).
833 (**G**) RT-qPCR of the indicated BAHD1-repressed gene 48 hours post-transduction of the
834 scramble control (mock) or BAHD1-targeting siRNA (siBAHD1) into parental HEK293 cells
835 (black and red), as well as post-transduction of siBAHD1 into HEK293 cell lines pre-rescued with
836 exogenous siBAHD1-resistant BAHD1, either WT or BAH-mutated (W667G or Y669A). Data of
837 three independent experiments are presented as mean ± SD after normalization to GAPDH and
838 then to mock. n.s., not significant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.
839 (**H**) Immunoblotting for Polycomb proteins (EZH2, SUZ12 and RING1B) in the indicated HEK293
840 cells used in the panel **G**.
841 (**I**) Venn diagram showing overlap between the BAHD1/H3K27me3 co-bound genes (left) and
842 DEGs derepressed due to the BAHD1 W667G mutation, relative to WT (right), in HEK293 cells.
843 Below lists the Polycomb target examples.
844
845 **Figure 4. H3K27me3 and BAHD1 co-repress Polycomb target genes.**
846
847 (**A**) H3K27me3 and total H3 immunoblots using HEK293 cells post-treatment with 5 uM of DMSO
848 or UNC1999 for three days.
849 (**B**) Heatmap showing the relative expression of DEGs in HEK293 cells treated with 5 uM of
850 UNC1999 (right) relative to DMSO (left) for three days, as revealed by RNA-seq (n=3 biological
851 replicates per group).
852 (**C**) RT-qPCR of BAHD1 and the indicated BAHD1-repressed target gene in HEK293 cells post-
853 treatment with 5 uM of UNC1999 versus DMSO for three days (n = 3 biologically independent
854 samples). Data are presented as mean ± SD after normalization to GAPDH and then to mock-
855 treated. n.s., not significant; ** $P < 0.01$; *** $P < 0.001$.
856 **(D)** Venn diagram showing significant overlap between the DEGs upregulated due to the
857 W667G-mutated BAHD1 relative to WT (left) and those upregulated due to UNC1999 treatment
858 relative to DMSO (right) in HEK293 cells.
859 (**E**) Immunoblotting for f-BAHD1 using the total extract (left) or the chromatin-bound and soluble
860 fractions (right) of the HEK293 cells post-treatment of UNC1999, relative to DMSO. GAPDH and
861 H3 served as the control of fractionation.
862
863 **Figure 5. BAHD1 recruits HDACs, mediating BAHD1 target gene repression.**
864 **(A)** Volcano plot, based on quantitative mass spectrometry data, demonstrated the enrichment
865 score and statistical significance of the BAHD1 interactors identified from FLAG pulldown of the
866 f-BAHD1 stable expression cells, compared with that of parental HEK293 cells. The x-axis and
867 y axis represent fold-change value (f-BAHD1/parental) and $-\log_{10}$ (P value) showing statistical
868 significance, respectively.
869 (**B-C**) CoIP (**B**) and IF (**C**) showing both interaction and co-localization of endogenous HDACs
870 with f-BAHD1 in HEK293 cells.
871 (**D**) CoIP for HDAC2 interaction with f-BAHD1, either WT or BAH-mutated, in HEK293 cells. WG

and YA indicate W667G and Y669A, respectively.
(**E**) GSEA revealing a positive correlation between the W667G mutation of BAHD1 and reactivation of the indicated HDAC target signature.
(**F**) ChIP-qPCR for detecting HDAC2 binding to the indicated BAHD1 gene target in HEK293 cells with BAHD1^{W667G} or BAHD1^{Y669A}, compared to WT BAHD1.
(**G**) RT-qPCR for the indicated BAHD1 target gene in HEK293 cells treated with 5uM of CI-994, relative to DMSO control, for 5 days. Paired t-test was used (n.s., P >0.05; *, P < 0.05; **, P < 0.01).

Figure 6. BAHD1 binding to H3K27me3 is required for maintaining histone deacetylation at Polycomb gene targets.
(A-B) Averaged ChIP-seq signals **(A)** and heatmap **(B)** of H3K27ac, H3K27me3, H3K9ac and H3K9me3 over ±5 kb from TSS to TES of the BAHD1-targeted Polycomb genes in HEK293 cells with expression of either WT or W667G-mutated BAHD1. ChIP-seq reads were normalized to input and sequencing depth. TSS, transcription start site; TES, transcription end site. Color bar, log2(ChIP/Input).
(C) CUT&RUN signals of WT BAHD1 (FLAG; top) and ChIP-seq profiles of H3K27ac (black), H3K9ac (purple), H3K27me3 (dark blue) and H3K9me3 (light blue) at the indicated Polycomb target in HEK293 cells with expression of either WT or W667G-mutated BAHD1.

Figure 7. Engagement of H3K27me3 by BAHD1^{BAH} is essential for normal development in mice.
(**A**) Bar-chart showing the ratio of newly born pups, which were either *Bahd1* WT (w/w; n= 30) or heterozygous (w/m; n= 42) or homozygous (m/m; n= 2) for *Bahd1^{W659G}*, from breeding of heterozygous pairs.
(**B**) Chi-square test of the genotypes and sex distribution of the indicated genotype among newly born pups from breeding of the *Bahd1^{W659G}* heterozygous pairs. Note that the female pups or the *Bahd1^{W659G}* homozygous mutant pups were found to be significantly under-represented.
(**C**) Images showing body size of the littermates with the indicated *Bahd1^{W659G}* genotype eight weeks post-birth.
(D) Measurement of body size (top) and body weight (bottom) at the indicated time point post-birth showing that the *Bahd1^{W659G}* homozygous mice (female or male; n=1 and 1) were smaller than its heterozygous (n=2 and 3) or WT (n=2 and 5) littermates.
(**E-F**) Summary (**E**) and examples (**F**) of GSEA revealed reactivation of the indicated Polycomb gene signature in MEF cells carrying the *Bahd1^{W659G}* homozygous mutation, relative to WT.

Fig 1.

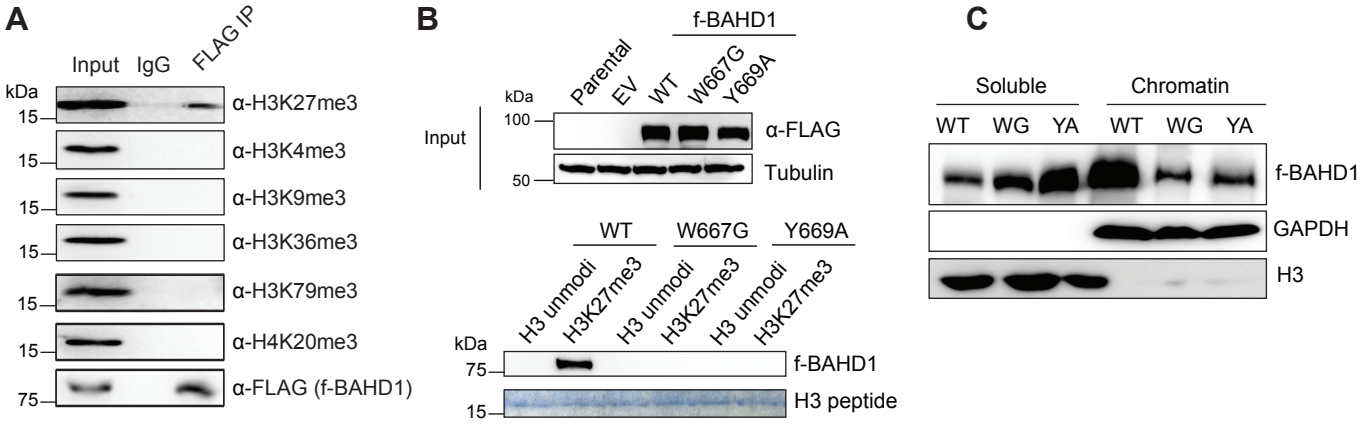
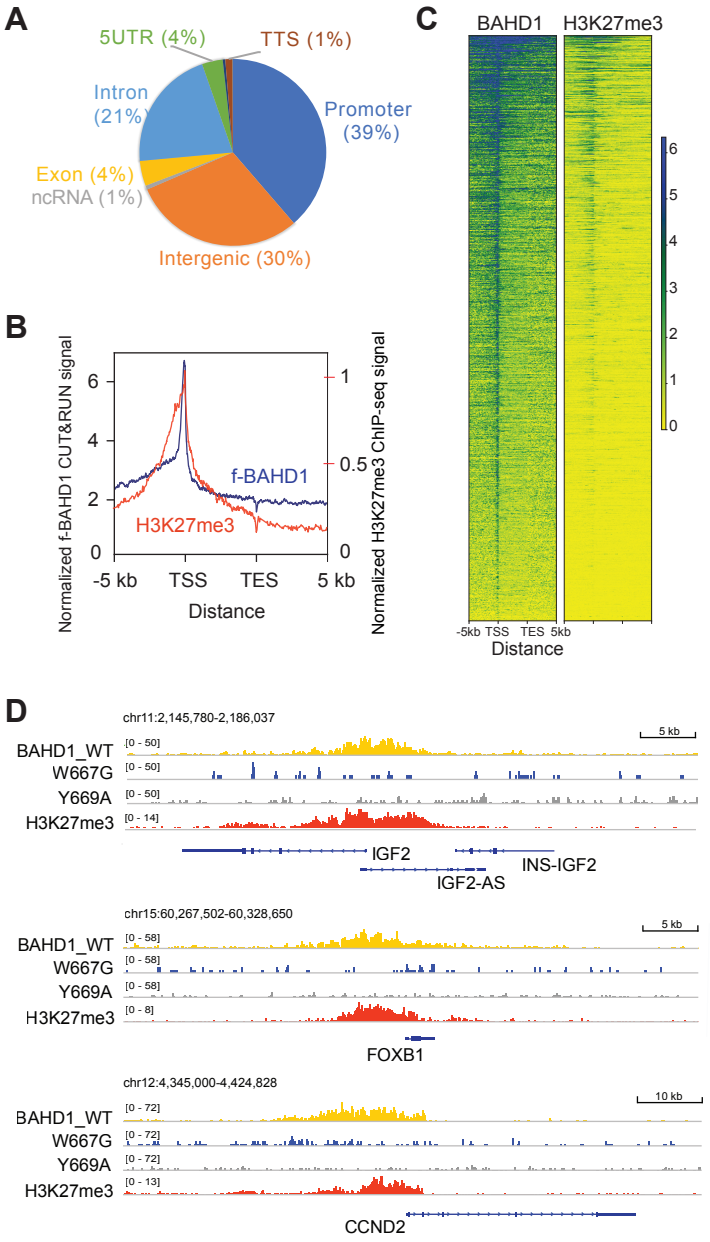

A
Input
IgG
FLAG IP
kDa
15
α-H3K27me3
15
α-H3K4me3
15
α-H3K9me3
15
α-H3K36me3
15
α-H3K79me3
15
α-H4K20me3
75
α-FLAG (f-BAHD1)
B
f-BAHD1
Parental
EV
WT
W667G
Y669A
kDa
Input
100
α-FLAG
50
Tubulin
WT
W667G
Y669A
H3 unmodi
H3K27me3
H3 unmodi
H3K27me3
H3 unmodi
H3K27me3
kDa
75
f-BAHD1
15
H3 peptide
C
Soluble
Chromatin
WT
WG
YA
WT
WG
YA
f-BAHD1
GAPDH
H3

Fig 2


A
5UTR (4%)
TTS (1%)
Intron (21%)
Promoter (39%)
Exon (4%)
ncRNA (1%)
Intergenic (30%)
B
Normalized f-BAHD1 CUT&RUN signal
Normalized H3K27me3 ChIP-seq signal
f-BAHD1
H3K27me3
-5 kb
TSS
TES
5 kb
Distance
C
BAHD1
H3K27me3
-5kb TSS TES 5kb
Distance
D
chr11:2,145,780-2,186,037
BAHD1_WT
W667G
Y669A
H3K27me3
IGF2
INS-IGF2
IGF2-AS
chr15:60,267,502-60,328,650
FOXB1
chr12:4,345,000-4,424,828
CCND2

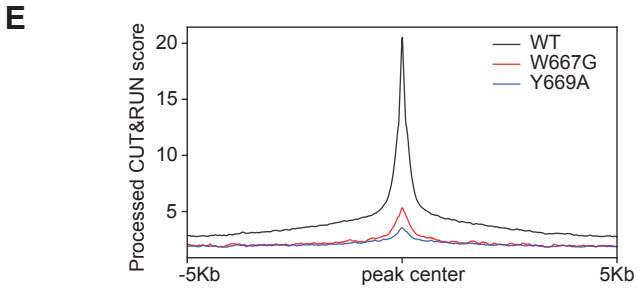
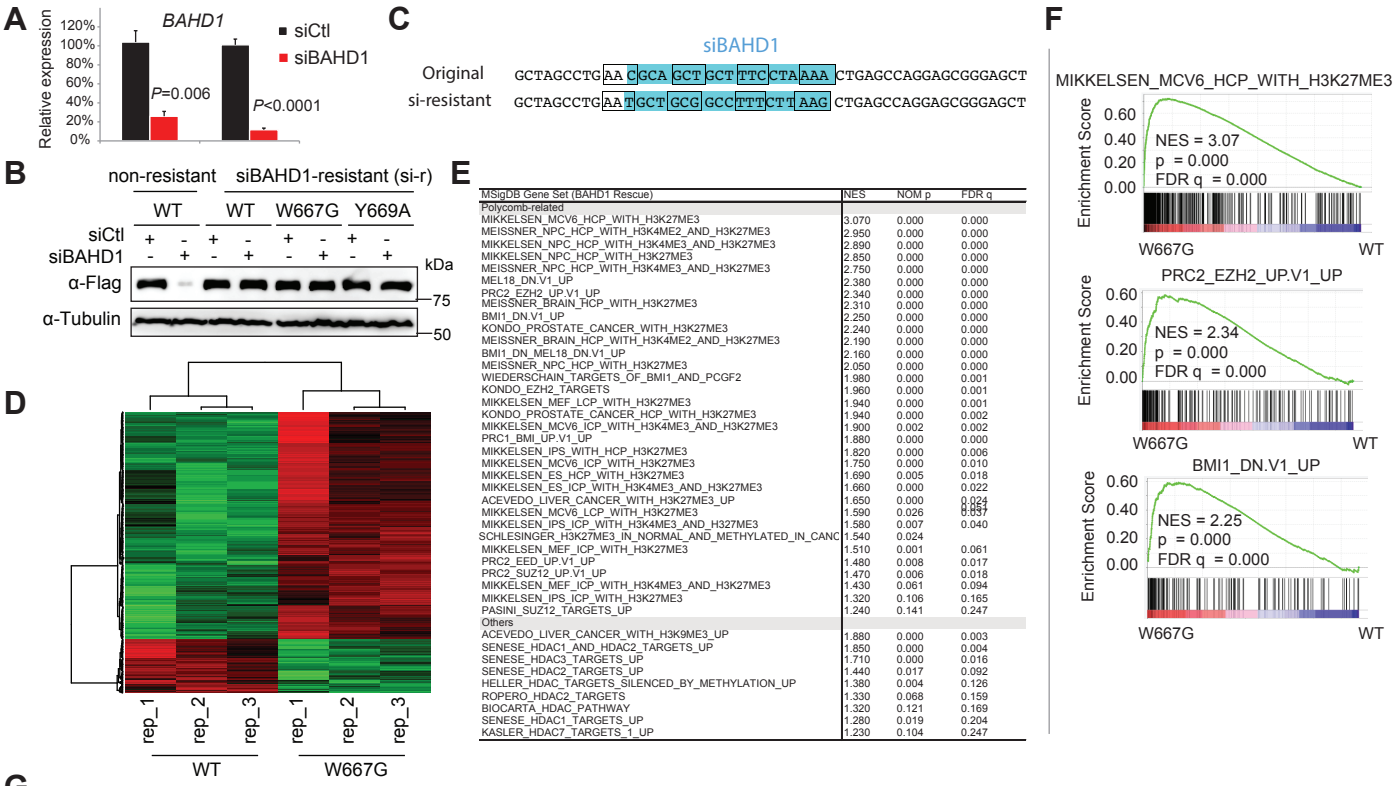

E
Processed CUT&RUN score
WT
W667G
Y669A
-5Kb
peak center
5Kb

Fig 3.



MSigDB Gene Set (BAHD1 Rescue)	NES	NOM p	FDR q
Polycomb-related			
MIKKELSEN_MCV6_HCP_WITH_H3K27ME3	3.070	0.000	0.000
MEISSNER_NPC_HCP_WITH_H3K4ME2_AND_H3K27ME3	2.950	0.000	0.000
MIKKELSEN_NPC_HCP_WITH_H3K4ME3_AND_H3K27ME3	2.890	0.000	0.000
MIKKELSEN_NPC_HCP_WITH_H3K27ME3	2.850	0.000	0.000
MEISSNER_NPC_HCP_WITH_H3K4ME3_AND_H3K27ME3	2.750	0.000	0.000
MEL18_DN.V1_UP	2.380	0.000	0.000
PRC2_EZH2_UP.V1_UP	2.340	0.000	0.000
MEISSNER_BRAIN_HCP_WITH_H3K27ME3	2.310	0.000	0.000
BMI1_DN.V1_UP	2.250	0.000	0.000
KONDO_PROSTATE_CANCER_WITH_H3K27ME3	2.240	0.000	0.000
MEISSNER_BRAIN_HCP_WITH_H3K4ME2_AND_H3K27ME3	2.190	0.000	0.000
BMI1_DN_MEL18_DN.V1_UP	2.160	0.000	0.000
MEISSNER_NPC_HCP_WITH_H3K27ME3	2.050	0.000	0.000
WIEDERSCHAIN_TARGETS_OF_BMI1_AND_PCGF2	1.980	0.000	0.001
KONDO_EZH2_TARGETS	1.960	0.000	0.001
MIKKELSEN_MEF_LCP_WITH_H3K27ME3	1.940	0.000	0.001
KONDO_PROSTATE_CANCER_HCP_WITH_H3K27ME3	1.940	0.000	0.002
MIKKELSEN_MCV6_ICP_WITH_H3K4ME3_AND_H3K27ME3	1.900	0.002	0.002
PRC1_BMI_UP.V1_UP	1.880	0.000	0.000
MIKKELSEN_IPS_WITH_HCP_H3K27ME3	1.820	0.000	0.006
MIKKELSEN_MCV6_ICP_WITH_H3K27ME3	1.750	0.000	0.010
MIKKELSEN_ES_HCP_WITH_H3K27ME3	1.690	0.005	0.018
MIKKELSEN_ES_ICP_WITH_H3K4ME3_AND_H3K27ME3	1.660	0.000	0.022
ACEVEDO_LIVER_CANCER_WITH_H3K27ME3_UP	1.650	0.000	0.024
MIKKELSEN_MCV6_LCP_WITH_H3K27ME3	1.590	0.026	0.037
MIKKELSEN_IPS_ICP_WITH_H3K4ME3_AND_H327ME3	1.580	0.007	0.040
SCHLESINGER_H3K27ME3_IN_NORMAL_AND_METHYLATED_IN_CANC	1.540	0.024	
MIKKELSEN_MEF_ICP_WITH_H3K27ME3	1.510	0.001	0.061
PRC2_EED_UP.V1_UP	1.480	0.008	0.017
PRC2_SUZ12_UP.V1_UP	1.470	0.006	0.018
MIKKELSEN_MEF_ICP_WITH_H3K4ME3_AND_H3K27ME3	1.430	0.061	0.094
MIKKELSEN_IPS_ICP_WITH_H3K27ME3	1.320	0.106	0.165
PASINI_SUZ12_TARGETS_UP	1.240	0.141	0.247
Others			
ACEVEDO_LIVER_CANCER_WITH_H3K9ME3_UP	1.880	0.000	0.003
SENESE_HDAC1_AND_HDAC2_TARGETS_UP	1.850	0.000	0.004
SENESE_HDAC3_TARGETS_UP	1.710	0.000	0.016
SENESE_HDAC2_TARGETS_UP	1.440	0.017	0.092
HELLER_HDAC_TARGETS_SILENCED_BY_METHYLATION_UP	1.380	0.004	0.126
ROPERO_HDAC2_TARGETS	1.330	0.068	0.159
BIOCARTA_HDAC_PATHWAY	1.320	0.121	0.169
SENESE_HDAC1_TARGETS_UP	1.280	0.019	0.204
KASLER_HDAC7_TARGETS_1_UP	1.230	0.104	0.247

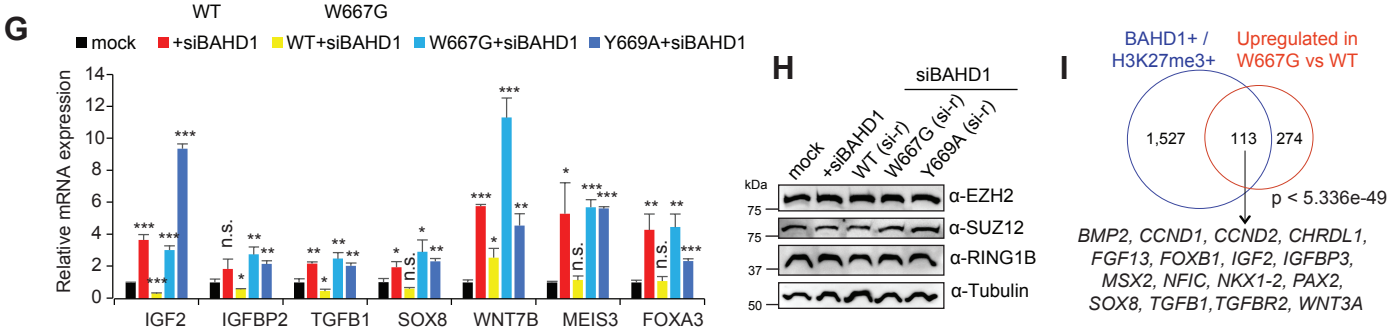
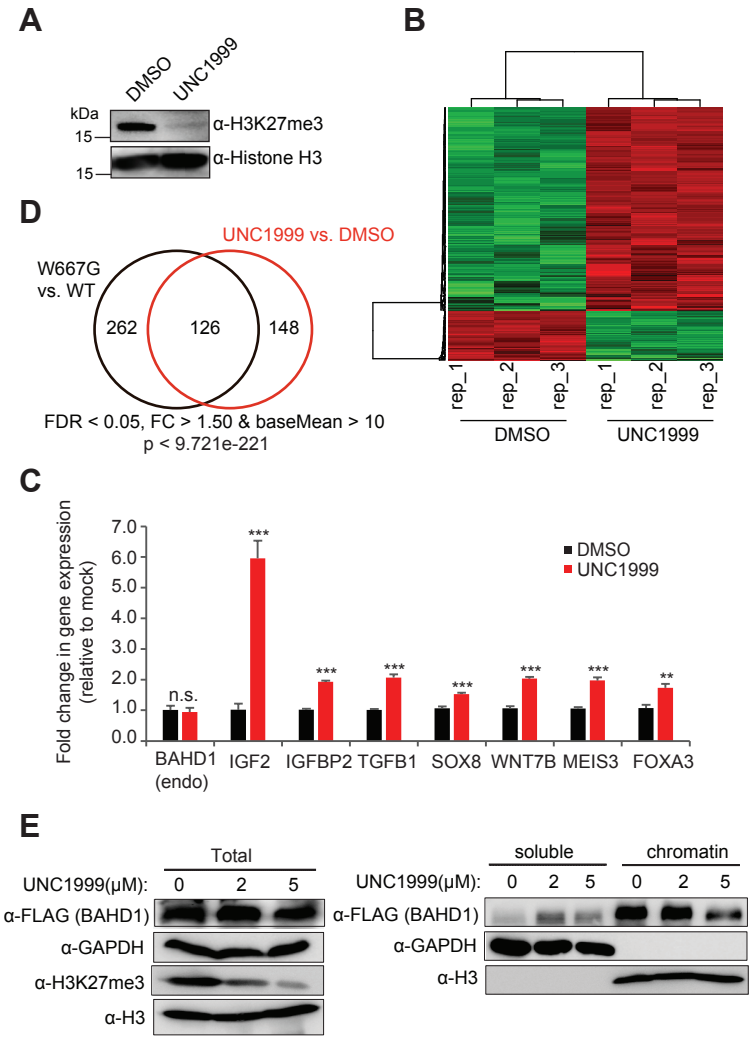
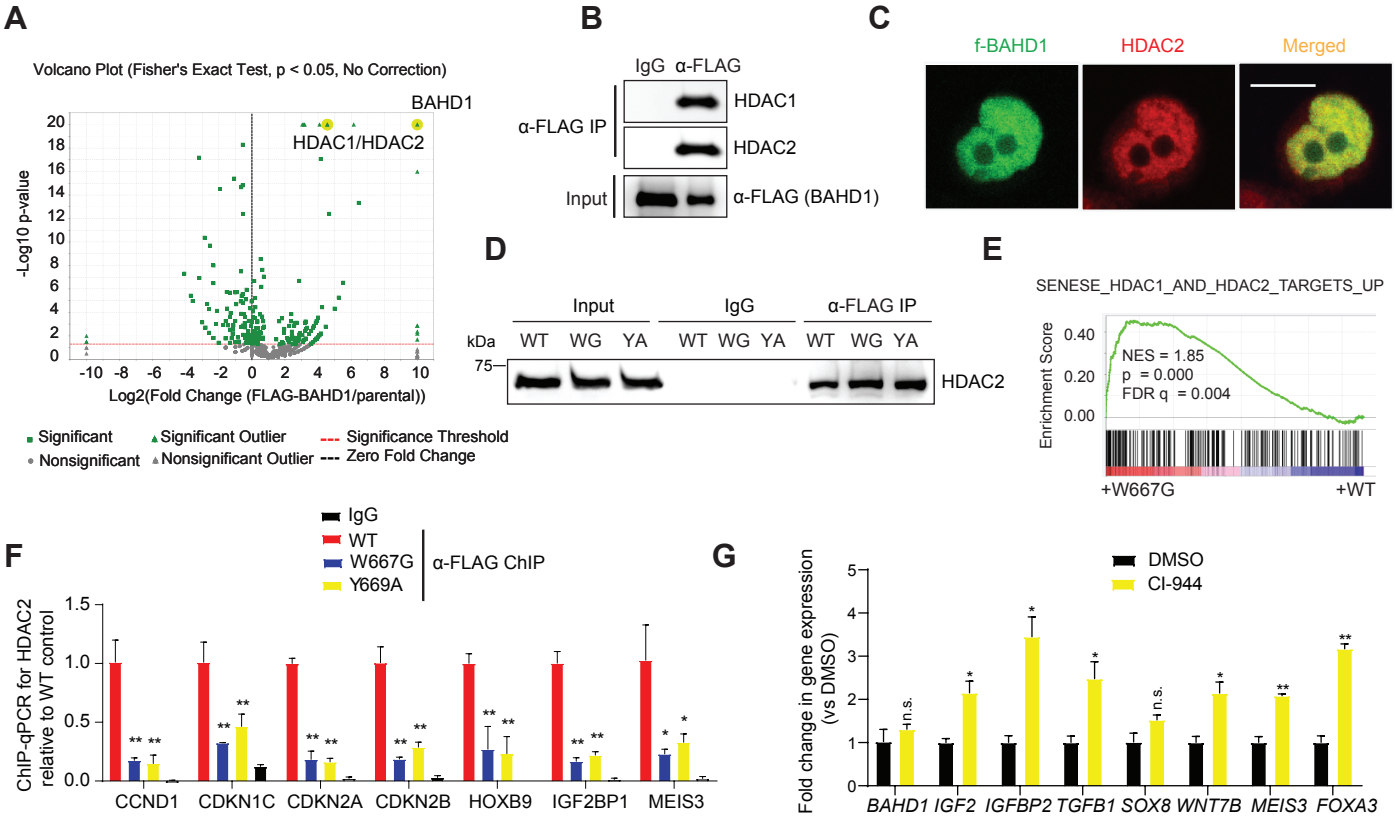


Fig 4.


A
DMSO
UNC1999
kDa
15
15
α-H3K27me3
α-Histone H3
B
rep_1
rep_2
rep_3
rep_1
rep_2
rep_3
DMSO
UNC1999
D
UNC1999 vs. DMSO
W667G
vs. WT
262
126
148
FDR < 0.05, FC > 1.50 & baseMean > 10
p < 9.721e-221
C
Fold change in gene expression
(relative to mock)
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
DMSO
UNC1999
n.s.
***
***
***
***
***
***
**
BAHD1
(endo)
IGF2
IGFBP2
TGFB1
SOX8
WNT7B
MEIS3
FOXA3

E
Total
UNC1999(μM):
0
2
5
α-FLAG (BAHD1)
α-GAPDH
α-H3K27me3
α-H3
soluble
chromatin
UNC1999(μM):
0
2
5
0
2
5
α-FLAG (BAHD1)
α-GAPDH
α-H3

Fig 5.


A
Volcano Plot (Fisher's Exact Test, p < 0.05, No Correction)
BAHD1
HDAC1/HDAC2
-Log10 p-value
Log2(Fold Change (FLAG-BAHD1/parental))
Significant
Nonsignificant
Significant Outlier
Nonsignificant Outlier
Significance Threshold
Zero Fold Change
B
IgG α-FLAG
α-FLAG IP
HDAC1
HDAC2
Input
α-FLAG (BAHD1)
C
f-BAHD1
HDAC2
Merged
D
Input
IgG
α-FLAG IP
kDa
WT WG YA WT WG YA WT WG YA
75
HDAC2
E
SENESE_HDAC1_AND_HDAC2_TARGETS_UP
Enrichment Score
NES = 1.85
p = 0.000
FDR q = 0.004
+W667G
+WT

F
IgG
WT
W667G
Y669A
α-FLAG ChIP
ChIP-qPCR for HDAC2 relative to WT control
CCND1
CDKN1C
CDKN2A
CDKN2B
HOXB9
IGF2BP1
MEIS3

G
DMSO
CI-944
Fold change in gene expression (vs DMSO)
n.s.
BAHD1
IGF2
IGFBP2
TGFB1
SOX8
WNT7B
MEIS3
FOXA3

Fig 6.

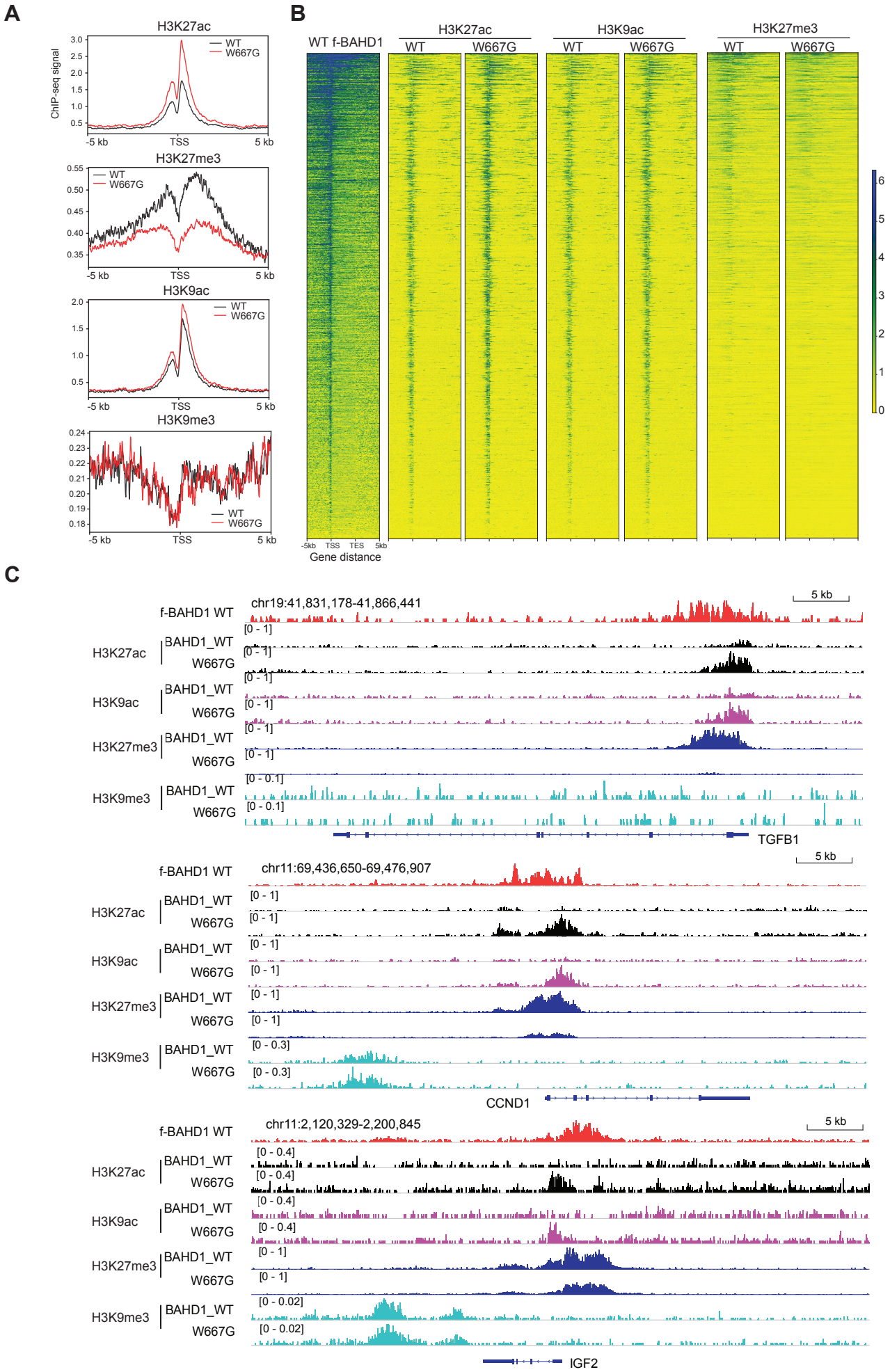
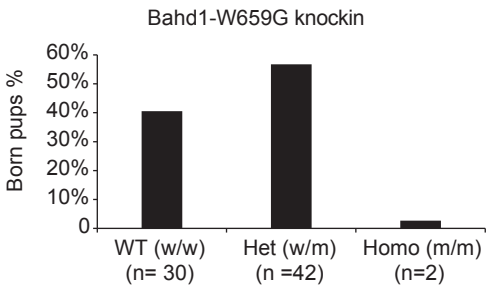

A
H3K27ac
H3K27me3
H3K9ac
H3K9me3
WT
W667G
ChIP-seq signal
-5 kb
TSS
5 kb
B
WT f-BAHD1
H3K27ac
H3K9ac
H3K27me3
WT
W667G
-5kb
TSS
TES
5kb
Gene distance
C
f-BAHD1 WT
chr19:41,831,178-41,866,441
5 kb
H3K27ac
H3K9ac
H3K27me3
H3K9me3
BAHD1_WT
W667G
[0 - 1]
[0 - 0.1]
TGFB1
chr11:69,436,650-69,476,907
[0 - 0.3]
CCND1
chr11:2,120,329-2,200,845
[0 - 0.4]
[0 - 0.02]
IGF2

Fig 7.

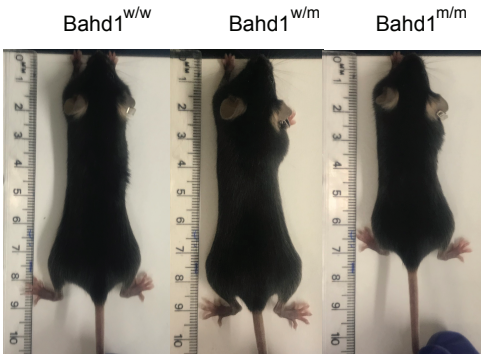
A



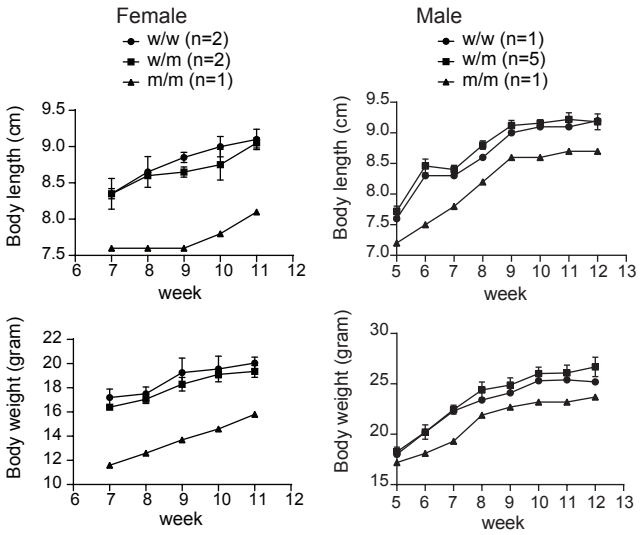
B

sex	WT	Het	Homo	total	p (χ2)
male	26	33	3	62	0.0002
female	14	18	3	35	0.031
p (χ2)		—		0.006	—

C



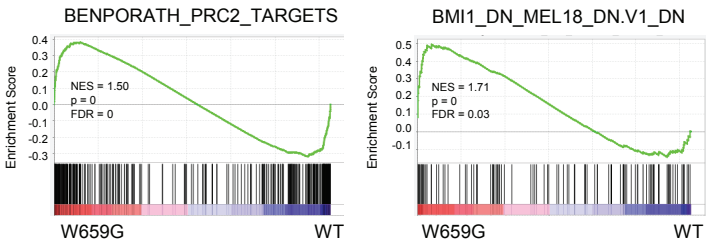
D



E

NAME	NES	FDR q
MIKKELSEN_MCV6_HCP_WITH_H3K27ME3	1.963	0.017
MEISSNER_NPC_HCP_WITH_H3K4ME2_&_H3K27ME3	1.634	0.162
PASINI_SUZ12_TARGETS_DN	1.553	0.231
MEISSNER_BRAIN_HCP_WITH_H3K27ME3	1.598	0.194
MIKKELSEN_IPS_WITH_HCP_H3K27ME3	1.544	0.238
BENPORATH_PRC2_TARGETS	1.499	0
BMI1_DN_MEL18_DN.V1_DN	1.707	0.032
MEL18_DN.V1_DN	1.460	0.104
PRC2_EZH2_UP.V1_UP	1.384	0.128
BMI1_DN_MEL18_DN.V1_UP	1.393	0.135
ESC_V6.5_UP_LATE.V1_UP	1.356	0.136
ESC_V6.5_UP_LATE.V1_DN	1.354	0.131
BMI1_DN.V1_UP	1.314	0.157
MEL18_DN.V1_UP	1.329	0.145

F



Supplementary Material

A conserved BAH module within mammalian BAHD1 connects H3K27me3 to Polycomb gene silencing

Huitao Fan[1,2†], Yiran Guo[1,3†], et al

Supplementary Tables

Supplementary Table 1. RNA-seq identified the differentially expressed genes (DEGs) that showed significant up- or down-regulation in HEK293 cells with knockdown (KD) of endogenous BAHD1 and rescue by exogenous BAH-mutated BAHD1 (W667G), in comparison to wildtype (WT) BAHD1.

Supplementary Table 2. List of the BAHD1/H3K27me3 co-bound gene targets that exhibited significant upregulation due to the W667G mutation of BAHD1, relative to WT, in HEK293 cells.

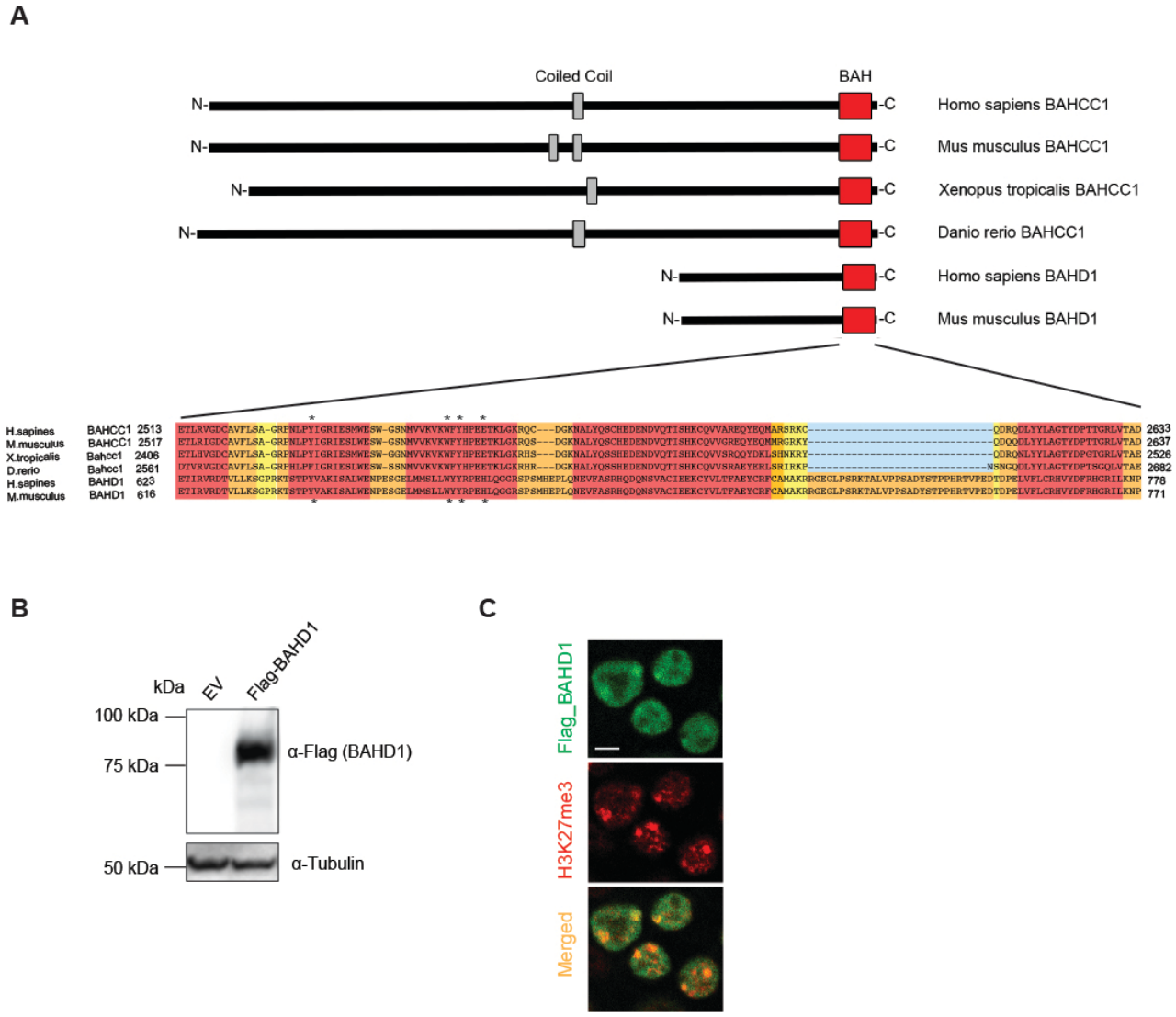
Supplementary Table 3. RNA-seq identified the DEGs that showed significant up- or down-regulation post-treatment with 5 μM of UNC1999, relative to mock (DMSO), for 3 days in HEK293 cells.

Supplementary Table 4. Summary of proteins identified by mass spectrometry following the FLAG pulldown in HEK293 cells, either parental (as a negative control) or with the FLAG-BAHD1 stable expression.

Supplementary Table 5. RNA-seq identified the DEGs that showed significant up- or down-regulation in primary mouse embryonic fibroblast (MEFs) cells carrying the W659G homozygous mutation of BAHD1, relative to WT.

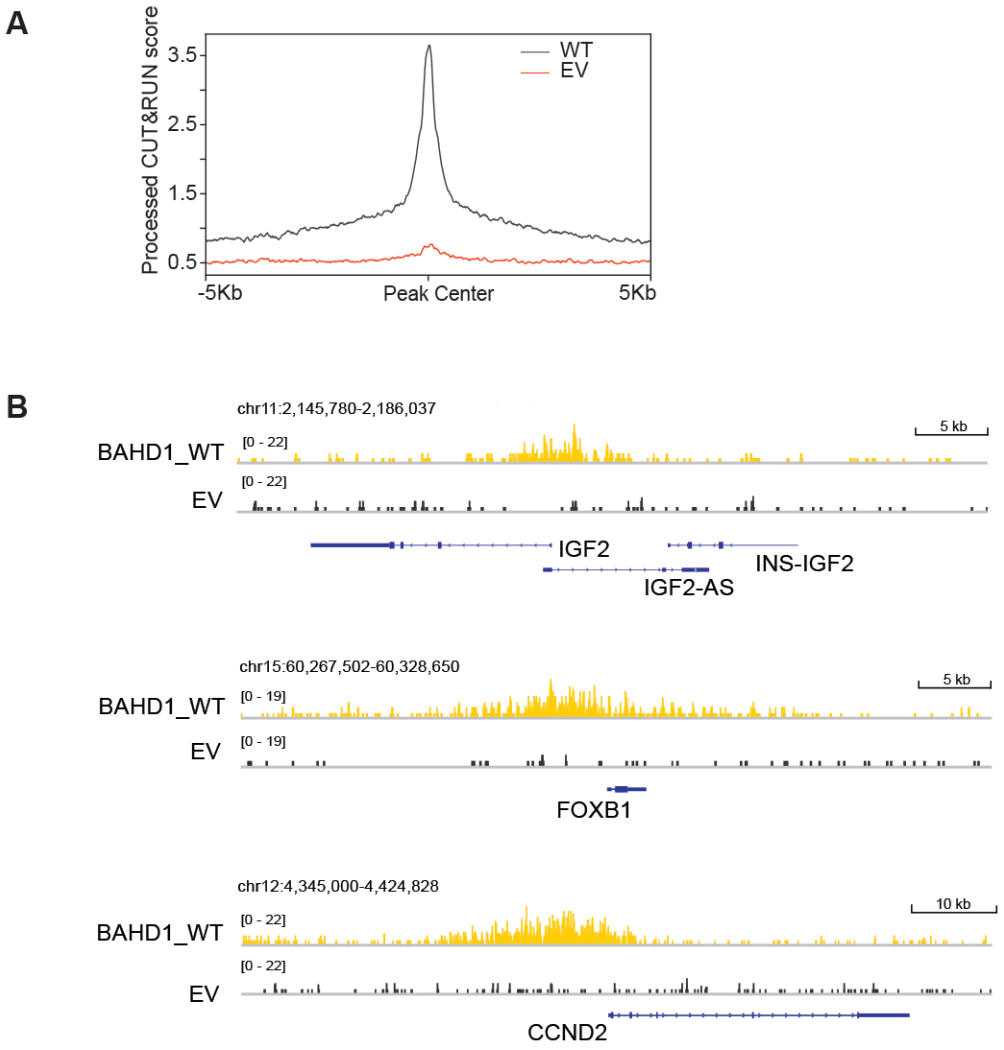
Supplementary Table 6. Information of the used reagents including primers, antibodies, and kits.

Supplementary Figure 1. A BAH module of BAHD1 (BAHD1^{BAH}) acts as an H3K27me3-specific 'reader' in mammalian cells.



(**A**) Alignment using the amino acid sequence of the BAH motif embedded in BAHD1 and BAHCC1 of the indicated species. The conserved residues are highlighted in red. The H3K27me3-'caging' residues are designated by stars.
(**B**) FLAG immunoblotting of the exogenously expressed BAHD1-3xFLAG (detected as a ~85kD protein) post-transfection into HEK293 cells.
(**C**) Representative images of confocal immunofluorescence microscopy showing that BAHD1 (green) colocalizes with H3K27me3 (red) in HEK293 cells. Scale bar, 15 μm.

Supplementary Figure 2. A functional BAHD1^{BAH} motif contributes significantly to BAHD1 binding to the H3K27me3-marked genes in HEK293 cells.



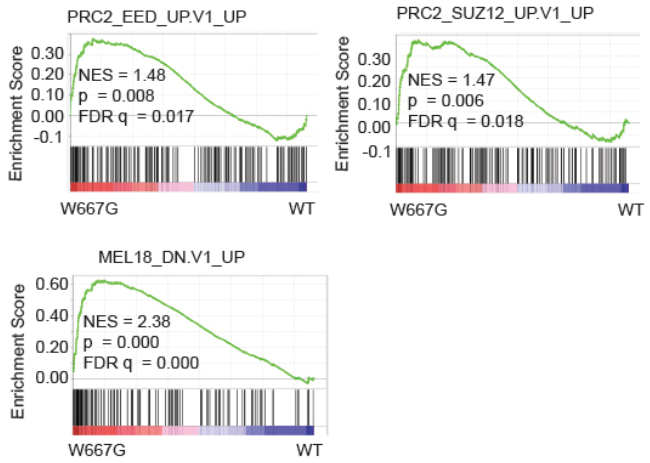
(A) Averaged FLAG CUT&RUN signals (y-axis) in 293 cells transduced with either empty vector (EV; which serves as a negative control) or WT f-BAHD1. The two samples were processed in parallel and the libraries were subjected to single-end sequencing. The libraries of CUT&RUN samples shown in main Figure 2 were subjected to paired-end sequencing and thus their data analysis were conducted separately.
(B) IGV views showing the called WT f-BAHD1 peak at the indicated gene in HEK293 cells transduced with EV or WT f-BAHD1.

Supplementary Figure 3. Transcriptomic profiling demonstrates an essential requirement of BAHD1^{BAH} for the BAHD1-mediated Polycomb gene repression in HEK293 cells.

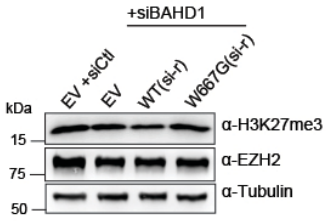
A

	Adjusted p-value
BENPORATH_ES_WITH_H3K27ME3	7.86E-90
BENPORATH_SUZ12_TARGETS	1.19E-75
MIKKELSEN_MCV6_HCP_WITH_H3K27ME3	1.61E-71
BENPORATH_EED_TARGETS	1.23E-60
GO_NEUROGENESIS	5.40E-54
BENPORATH_PRC2_TARGETS	3.03E-44
MIKKELSEN_MEF_HCP_WITH_H3K27ME3	3.47E-44
GO_POSITIVE_REGULATION_OF_DEVELOPMENTAL_PROCESS	1.23E-43
GO_NEURON_DIFFERENTIATION	1.70E-42
GO_MULTICELLULAR_ORGANISMAL_PROCESS	1.07E-39
GO_REGULATION_OF_CELL_DIFFERENTIATION	2.19E-39
MEISSNER_NPC_HCP_WITH_H3K4ME2_AND_H3K27ME3	8.94E-37
GO_REGULATION_OF_NERVOUS_SYSTEM_DEVELOPMENT	7.29E-36
WANG_SMARCE1_TARGETS_UP	1.34E-34
MIKKELSEN_NPC_HCP_WITH_H3K27ME3	6.04E-29
DOUGLAS_BMI1_TARGETS_UP	3.01E-23
GO_CIRCULATORY_SYSTEM_DEVELOPMENT	4.28E-22
GO_EPITHELIUM_DEVELOPMENT	5.61E-22
GO_CELLULAR_COMPONENT_MORPHOGENESIS	1.25E-21
HELLER_HDAC_TARGETS_SILENCED_BY_METHYLATION_UP	9.50E-20
GO_BLOOD_VESSEL_MORPHOGENESIS	3.80E-19
GO_NEPHRON_DEVELOPMENT	2.15E-18
GO_HEAD_DEVELOPMENT	2.70E-18
GOZGIT_ESR1_TARGETS_DN	3.16E-18
GO_VASCULATURE_DEVELOPMENT	4.84E-18
GO_EMBRYO_DEVELOPMENT	1.64E-15
MEISSNER_BRAIN_HCP_WITH_H3K27ME3	5.03E-15
NUYTTEN_EZH2_TARGETS_UP	1.87E-14
SCHLESINGER_METHYLATED_DE_NOVO_IN_CANCER	6.37E-14
GO_NEGATIVE_REGULATION_OF_CELL_DIFFERENTIATION	1.81E-12
GO_REGULATION_OF_GROWTH	2.03E-12
GO_SKELETAL_SYSTEM_DEVELOPMENT	2.49E-11
SENESE_HDAC1_AND_HDAC2_TARGETS_UP	5.00E-11
HELLER_SILENCED_BY_METHYLATION_UP	1.83E-10
HELLER_HDAC_TARGETS_UP	7.29E-10
WIEDERSCHAIN_TARGETS_OF_BMI1_AND_PCGF2	1.02E-08

B



C



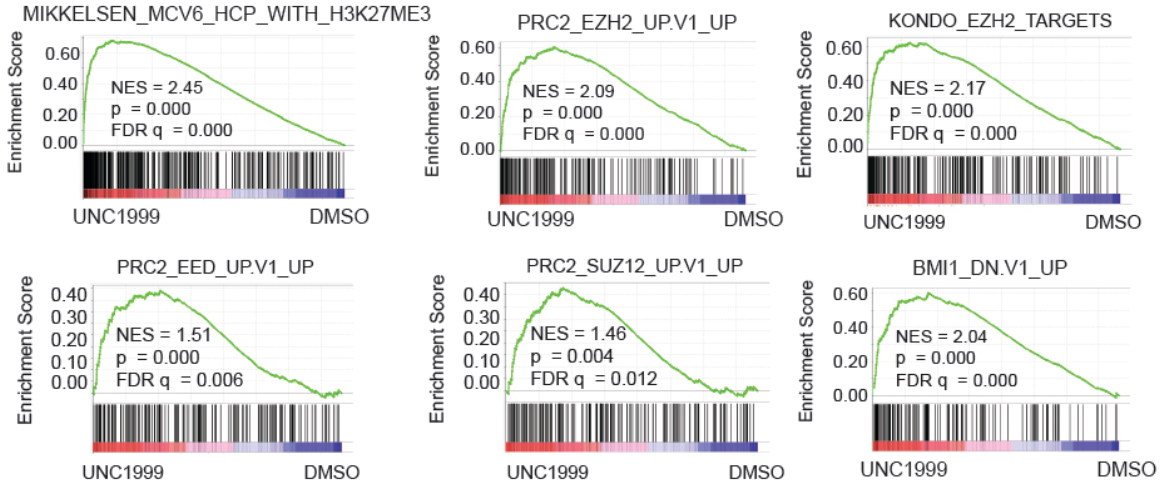
(A) Platform for Integrative Analysis of Omics data (PIANO) analyses of RNA-seq profiles revealed that, relative to WT, BAHD1^{W667G} is positively associated with derepression of genesets marked by H3K27me3, repressed by PRC2 or PRC1, or related to development.
(B) GSEA shows that, relative to WT, the H3K27me3-binding-defective mutation of BAHD1^{BAH} is positively correlated to reactivation of gene signatures related to PRC2 or PRC1.
(C) Immunoblotting for EZH2 and H3K27me3 in the indicated HEK293 cells, which were pre-rescued with either empty vector (EV) or an exogenous siBAHD1-resistant (si-r) BAHD1, WT or BAH-mutated (W667G), followed by transduction of control siRNA (siCtl, lane 1) or siBAHD1 (lanes 2-4) for 48 hours.

Supplementary Figure 4. H3K27me3 and BAHD1, which assembles a complex with corepressors such as HDAC1/2, co-repress a set of Polycomb target genes.

A

MSigDB Gene Set (BAHD1 UNC1999)	NES	NOM p	FDR q
Polycomb-related			
MIKKELSEN_MCV6_HCP_WITH_H3K27ME3	2.450	0.000	0.000
MIKKELSEN_NPC_HCP_WITH_H3K4ME3_AND_H3K27ME3	2.440	0.000	0.000
MEISSNER_NPC_HCP_WITH_H3K4ME2_AND_H3K27ME3	2.390	0.000	0.000
MIKKELSEN_NPC_HCP_WITH_H3K27ME3	2.370	0.000	0.000
MEISSNER_NPC_HCP_WITH_H3K4ME3_AND_H3K27ME3	2.330	0.000	0.000
KONDO_EZH2_TARGETS	2.170	0.000	0.000
KONDO_PROSTATE_CANCER_WITH_H3K27ME3	2.130	0.000	0.000
PRC2_EZH2_UP.V1_UP	2.090	0.000	0.000
BMI1_DN.V1_UP	2.040	0.000	0.000
BMI1_DN_MEL18_DN.V1_UP	2.030	0.000	0.000
MEISSNER_NPC_HCP_WITH_H3K27ME3	2.030	0.000	0.000
MIKKELSEN_ES_ICP_WITH_H3K4ME3_AND_H3K27ME3	2.010	0.000	0.000
MIKKELSEN_MEF_LCP_WITH_H3K27ME3	1.980	0.000	0.000
MIKKELSEN_MCV6_ICP_WITH_H3K4ME3_AND_H3K27ME3	1.950	0.000	0.000
MEL18_DN.V1_UP	1.940	0.000	0.000
MEISSNER_BRAIN_HCP_WITH_H3K27ME3	1.910	0.000	0.001
MIKKELSEN_IPS_ICP_WITH_H3K4ME3_AND_H327ME3	1.900	0.000	0.001
ACEVEDO_LIVER_CANCER_WITH_H3K27ME3_UP	1.900	0.000	0.001
KONDO_PROSTATE_CANCER_HCP_WITH_H3K27ME3	1.890	0.000	0.001
WIEDERSCHAIN_TARGETS_OF_BMI1_AND_PCGF2	1.890	0.000	0.001
MEISSNER_BRAIN_HCP_WITH_H3K4ME2_AND_H3K27ME3	1.780	0.000	0.004
MIKKELSEN_MEF_ICP_WITH_H3K27ME3	1.760	0.000	0.005
MIKKELSEN_ES_HCP_WITH_H3K27ME3	1.710	0.003	0.009
MIKKELSEN_MEF_ICP_WITH_H3K4ME3_AND_H3K27ME3	1.690	0.001	0.011
MIKKELSEN_MCV6_ICP_WITH_H3K27ME3	1.630	0.001	0.020
SCHLESINGER_H3K27ME3_IN_NORMAL_AND_METHYLATED_IN_CANC	1.620	0.010	0.021
MIKKELSEN_IPS_ICP_WITH_H3K27ME3	1.610	0.009	0.023
MIKKELSEN_MCV6_LCP_WITH_H3K27ME3	1.520	0.035	0.046
PRC2_EED_UP.V1_UP	1.510	0.000	0.006
PRC2_SUZ12_UP.V1_UP	1.460	0.004	0.012
PRC1_BMI_UP.V1_UP	1.390	0.012	0.028
Others			
HELLER_HDAC_TARGETS_SILENCED_BY_METHYLATION_UP	1.820	0.000	0.002
SENESE_HDAC1_AND_HDAC2_TARGETS_UP	1.720	0.000	0.008
SENESE_HDAC3_TARGETS_UP	1.620	0.000	0.021
HELLER_HDAC_TARGETS_SILENCED_BY_METHYLATION_DN	1.550	0.000	0.038
HELLER_HDAC_TARGETS_UP	1.410	0.001	0.094
BIOCARTA_HDAC_PATHWAY	1.400	0.081	0.100
SENESE_HDAC1_TARGETS_UP	1.370	0.000	0.119
SENESE_HDAC2_TARGETS_UP	1.250	0.123	0.230

B



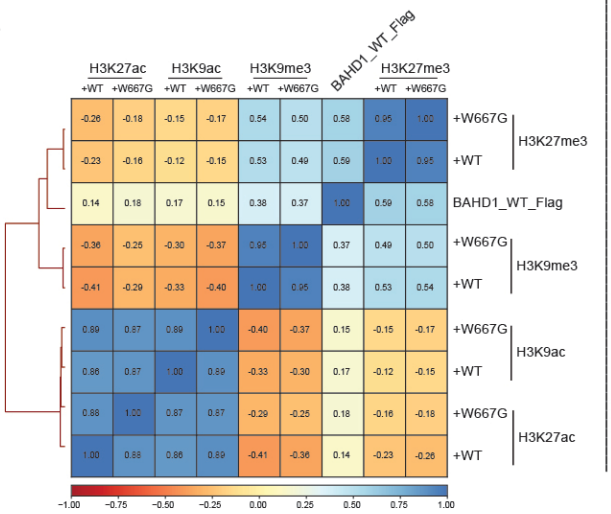
(A) Summary of GSEA using RNA-seq profiles of HEK293 cells post-treatment of UNC1999, relative to DMSO. Representative genesets are categorized into Polycomb (PRC2 and PRC1)-related or others.
(B) GSEA demonstrates that UNC1999 treatment is correlated with derepression of genesets suppressed by PRC1 or PRC2 or the H3K27me3-bound genes.

Supplementary Figure 5. BAHD1 binding to H3K27me3 is required for maintenance of histone deacetylation at Polycomb gene targets.

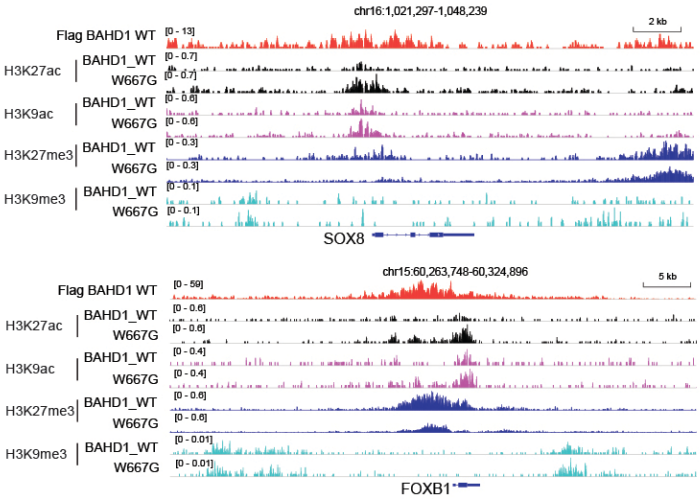
A

Sample ID	Epitope	Raw read counts	Uniquely aligned read counts
HEK293_siBAHD1+WT	Input	16,675,719	15,474,688
HEK293_siBAHD1+W667G	Input	22,819,626	21,179,652
HEK293_siBAHD1+WT	H3K27ac_ChIP	16,675,719	15,474,688
HEK293_siBAHD1+W667G	H3K27ac_ChIP	15,871,141	14,530,199
HEK293_siBAHD1+WT	H3K9ac_ChIP	14,314,621	12,771,888
HEK293_siBAHD1+W667G	H3K9ac_ChIP	18,978,077	16,468,905
HEK293_siBAHD1+WT	H3K27me3_ChIP	28,876,028	26,755,547
HEK293_siBAHD1+W667G	H3K27me3_ChIP	81,146,845	73,727,447
HEK293_siBAHD1+WT	H3K9me3_ChIP	59,837,532	40,305,403
HEK293_siBAHD1+W667G	H3K9me3_ChIP	47,851,890	33,294,800

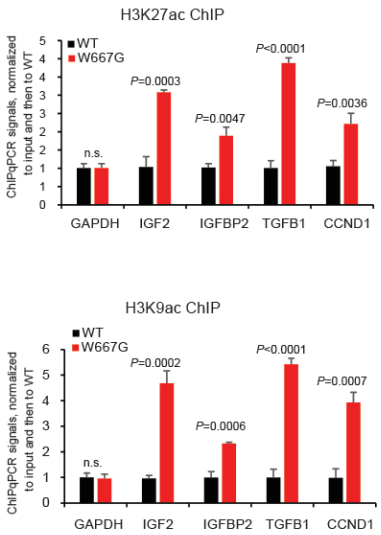
B



C

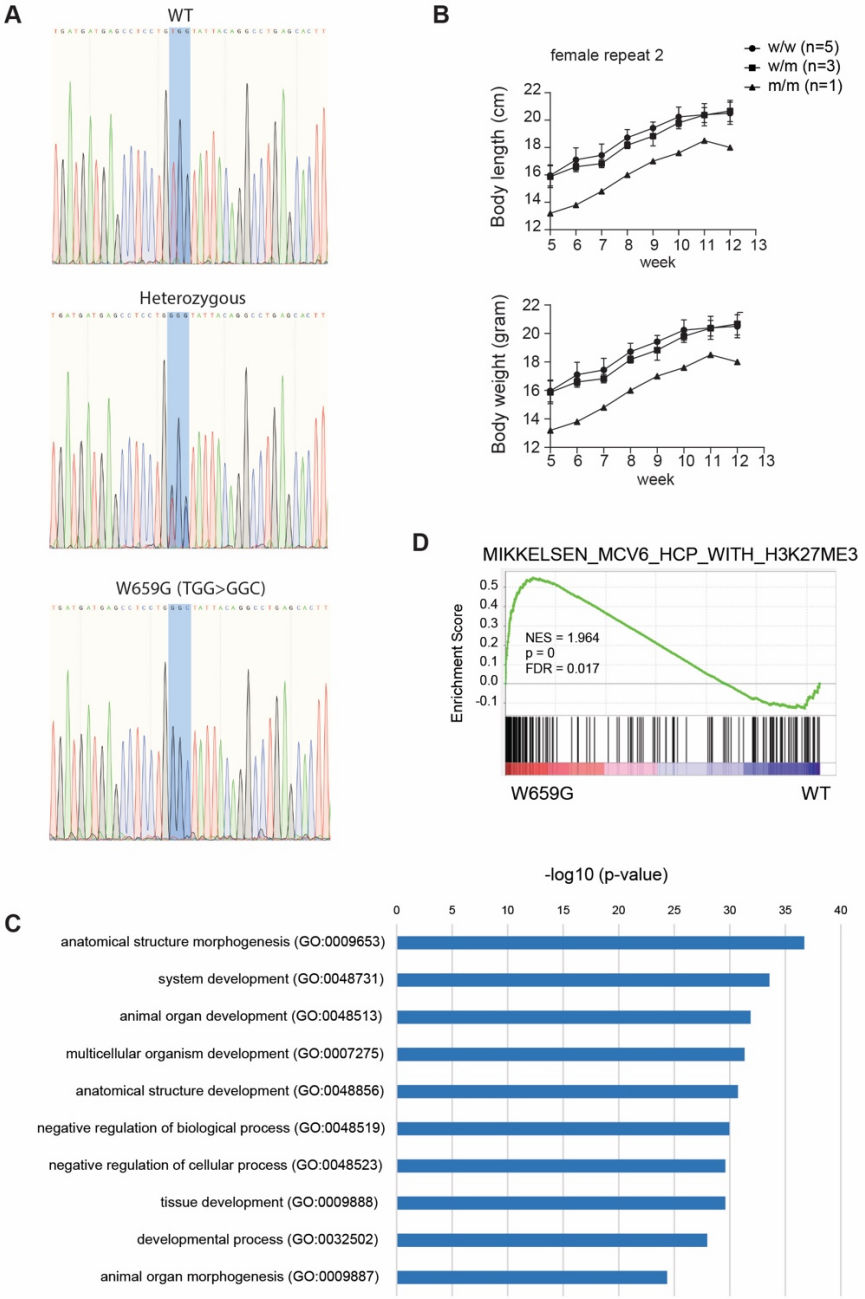


D



(**A**) Summary of read counts of input, H3K27ac, H3K9ac, H3K27me3 and H3K9me3 ChIP-seq using HEK293 cells (with endogenous BAHD1 depleted) carrying the exogenous WT or W667G-mutated BAHD1. (**B**) Correlation analysis between BAHD1 signals and the indicated histone modification, based on global genomic profiles. (**C**) ChIP-seq profiles of WT BAHD1 (red; FLAG) and H3K27ac (black), H3K9ac (purple), H3K27me3 (dark blue) and H3K9me3 (light blue) at SOX8 and FOXB1 in HEK293 cells (with endogenous BAHD1 depleted) carrying exogenous WT or W667G-mutated BAHD1. (**D**) ChIP-qPCR of H3K27ac and H3K9ac at the indicated gene promoter in HEK293 cells (with endogenous BAHD1 depleted) carrying the exogenous WT or W667G-mutated BAHD1 (n = 3 biologically independent samples). Data are presented as mean ± SD.

Supplementary Figure 6. The H3K27me3 engagement by BAHD1^{BAH} is essential for embryonic development in mice.



(A) Sanger sequencing confirmation of genotype of Bahd1, either WT or with a heterozygous or homozygous W659G (TGG>GGC) mutation, using tissue genomic DNA prepared from mice. **(B)** Measurements of body size and body weight at the indicated time point post-birth showed that a *Bahd1^{W659G}* homozygous female mouse (n=1) is smaller relative to its heterozygous (n=5) and WT (n=1) female littermates. (**C**) GO results using the genes upregulated in MEF cells carrying the *Bahd1^{W659G}* homozygous mutation, relative to WT. (**D**) GSEA results showed upregulation of the indicated gene in MEF cells carrying the *Bahd1^{W659G}* homozygous mutation, relative to WT.