Title

KAT6A mutations in Arboleda-Tham syndrome drive epigenetic regulation of posterior HOXC cluster.

ORIGINAL INVESTIGATION

KAT6A mutations in Arboleda-Tham syndrome drive epigenetic regulation of posterior *HOXC* cluster

Meghna Singh[1,2,3] · Sarah J. Spendlove[1,2,3,4] · Angela Wei[1,2,3,4] · Leroy M. Bondhus[1,2,3] · Aileen A. Nava[1,2,3] · Francisca N. de L. Vitorino[5] · Seth Amano[1,2,3] · Jacob Lee[1,2,3] · Gesenia Echeverria[1,2,3] · Dianne Gomez[1,2,3] · Benjamin A. Garcia[5] · Valerie A. Arboleda[1,2,3,4]

Abstract

Arboleda-Tham Syndrome (ARTHS) is a rare genetic disorder caused by heterozygous, de novo mutations in *Lysine(K) acetyltransferase 6A (KAT6A)*. ARTHS is clinically heterogeneous and characterized by several common features, including intellectual disability, developmental and speech delay, and hypotonia, and affects multiple organ systems. KAT6A is the enzymatic core of a histone–acetylation protein complex; however, the direct histone targets and gene regulatory effects remain unknown. In this study, we use ARTHS patient ($n = 8$) and control ($n = 14$) dermal fibroblasts and perform comprehensive profiling of the epigenome and transcriptome caused by *KAT6A* mutations. We identified differential chromatin accessibility within the promoter or gene body of 23% (14/60) of genes that were differentially expressed between ARTHS and controls. Within fibroblasts, we show a distinct set of genes from the posterior *HOXC* gene cluster (*HOXC10*, *HOXC11*, *HOXC-AS3*, *HOXC-AS2*, and *HOTAIR*) that are overexpressed in ARTHS and are transcription factors critical for early development body segment patterning. The genomic loci harboring *HOXC* genes are epigenetically regulated with increased chromatin accessibility, high levels of H3K23ac, and increased gene–body DNA methylation compared to controls, all of which are consistent with transcriptomic overexpression. Finally, we used unbiased proteomic mass spectrometry and identified two new histone post-translational modifications (PTMs) that are disrupted in ARTHS: H2A and H3K56 acetylation. Our multi-omics assays have identified novel histone and gene regulatory roles of *KAT6A* in a large group of ARTHS patients harboring diverse pathogenic mutations. This work provides insight into the role of KAT6A on the epigenomic regulation in somatic cell types.

✉ Valerie A. Arboleda
   vaa2001@g.ucla.edu

1 Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, UCLA, 615 Charles E. Young Drive South, Los Angeles, CA 90095, USA

2 Department of Human Genetics, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA

3 Department of Computational Medicine, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA

4 Interdepartmental BioInformatics Program, UCLA, Los Angeles, CA, USA

5 Department of Biochemistry and Molecular Biophysics, Washington University in St. Louis, St. Louis, MO, USA

Introduction

Arboleda-Tham Syndrome (ARTHS, OMIM #616268) is a rare genetic disorder caused by de novo heterozygous mutations in the *lysine(K) acetyltransferase 6A* (*KAT6A*, a.k.a. *MYST3* and *MOZ*) gene. ARTHS is characterized by intellectual disability, developmental and speech delays, hypotonia, and congenital heart defects (Arboleda et al. 2015; Tham et al. 2015; Kennedy et al. 2019) along with less penetrant phenotypes, such as seizures, microcephaly, and autism spectrum disorder. Most of the pathogenic mutations characterized to date are protein-truncating mutations that occur throughout the length of the gene, but are most commonly observed in the last exon that makes up over half of the gene. Missense mutations to the enzymatic histone acetyltransferase domain or highly conserved C-terminus regions have also been associated with ARTHS; however, the phenotype is less severe than the protein-truncating

mutations (Kennedy et al. 2019). The majority of patients with late-truncating mutations in KAT6A, located beyond exon 16, have more severe phenotypes compared to patients with early truncating mutations in exons 1 through 15 (Kennedy et al. 2019).

KAT6A is part of the orthologous MYST family of highly conserved histone acetyltransferase (HAT) genes (Thomas and Voss 2007) that modulates gene expression through deposition of acetyl groups on the histone tail, which increases chromatin accessibility at specific loci and allows for spatial arrangement of transcription factors and transcriptional machinery local control of gene expression (Nava and Arboleda 2023). Recent studies in mammalian model systems have demonstrated that KAT6A acetylates specific lysine residues on histone H3, including lysine 9 (Arboleda et al. 2015; Voss et al. 2009) and lysine 23 (Yan et al. 2020, 2017). Lower abundance modifications, such as propionyl modifications, have also been described (Yan et al. 2020) but are near impossible to validate due to limited antibody specificity and abundance. One of the key questions that remains is the extent to which *KAT6A* coordinates and directs transcriptional activation of specific genes and pathways in human cells, especially in the context of ARTHS.

The *KAT6A* gene exhibits significant structural conservation across eukaryotes indicating a shared biological functionality throughout evolution. In model systems, the *KAT6A* homologs have been shown to influence gene silencing (Reifsnyder et al. 1996), oocyte polarity (Huang et al. 2014), hematopoiesis (Genais et al. 2020; Katsumoto et al. 2006), and neuroblast proliferation (Scott et al. 2001; Reifsnyder et al. 1996). Tissue-specific *Kat6a^{-/-}* have demonstrated a key role of Kat6a in adult hematopoietic stem cell maintenance (Sheikh et al. 2016; Perez-Campo et al. 2009) and in craniofacial patterning through nitric oxide signaling and histone acetylation (Vanyai et al. 2019; Kong et al. 2014). In mice, knockout of *Kat6a* does not globally decrease H3K9 acetylation but leads to hypoacetylation at specific loci, and reduced expression of *Hoxa* and *Hoxb* gene clusters compared to controls driven by hypoacetylation of the transcriptional start sites controlling expression of *Hoxa4, Hoxb3,* and *Hoxb4* (Hibiya et al. 2009; Voss et al. 2009). The *Hox* gene clusters play a vital role on body patterning during early developmental timepoint (Mallo et al. 2010; Hubert and Wellik 2023) and in development of the central nervous system (Philippidou and Dasen 2013; Nolte and Krumlauf 2013).

In this study, we perform an epigenomic and transcriptomic assessment on fibroblasts from 8 ARTHS patients and 14 control individuals. All patients harbor a genetic diagnosis of ARTHS and have a pathogenic *KAT6A* mutation that results in premature truncation of the protein (Table S1). Our analysis identified 60 genes that were differentially expressed in ARTHS-patient dermal fibroblast samples as compared to controls. Our multi-omic dataset demonstrates that protein-truncating mutations in *KAT6A* cause differential chromatin accessibility, histone acetylation, and transcriptomic changes over posterior *HOXC* genes (*HOXC10*, *HOXC11*, *HOTAIR*, *HOXC-AS3*, and *HOXC-AS2*). Furthermore, a mass spectrometry-based assay on histone extracts identified novel histone PTMs on the Histone 2A and Histone 3 that are disrupted by *KAT6A* mutations. Our study provides valuable insights into the gene regulatory mechanisms that are perturbed in primary samples of Arboleda-Tham syndrome.

Results

To study the direct effect of *KAT6A* mutations on gene expression in ARTHS patients, we established fibroblast cell lines from ARTHS patients ($n = 8$) and unaffected individuals ($n = 14$). The pathogenic mutations in the patients disrupted the KAT6A protein between amino acid 379 and 1551 with seven of the eight mutations being de novo *truncating* mutations and one being a de novo missense change in a highly conserved region of *KAT6A* (Fig. 1A, Figure S1). The information regarding the mutation, age, sex, and type of dataset for each of the cell line used in this study is reported in Table S1. The clinical phenotypes of these eight patients were previously reported (Kennedy et al. 2019; Arboleda et al. 2015; Tham et al. 2015).

To understand how these *KAT6A* mutations disrupt the epigenome and RNA expression, we performed epigenomic and transcriptomic assays (Fig. 1B). Since KAT6A functions as part of a histone acetyltransferase, we sought to quantify the disruption of histone acetylation marks, H3K9 and H3K23 acetylation, using chromatin immunoprecipitation sequencing and chromatin accessibility using ATAC-seq. We also performed DNA methylation using Epic 850K array and RNA sequencing. Finally, we also performed a histone-targeted mass spectrometry assay to identify novel histone PTMs that might be regulated by KAT6A (Fig. 1B, Table S1). Representative samples were assessed for each assay.

We first assessed the RNA and protein levels of KAT6A in our dermal fibroblast lines. We did not find a statistically significant difference in *KAT6A* RNA (Fig. 1C) from our transcriptomic data and this was validated by qRT-PCR (Figure S2A). At the protein level, we did not see an appreciable difference in two of the three mutations. We did notice that our early truncating mutation patient, MDG1135, that has a nonsense mutation at amino acid position 379, showed a 80% decrease in protein levels and is the earliest truncating mutation line we tested (Fig. 1D, E, Figure S2B-C).
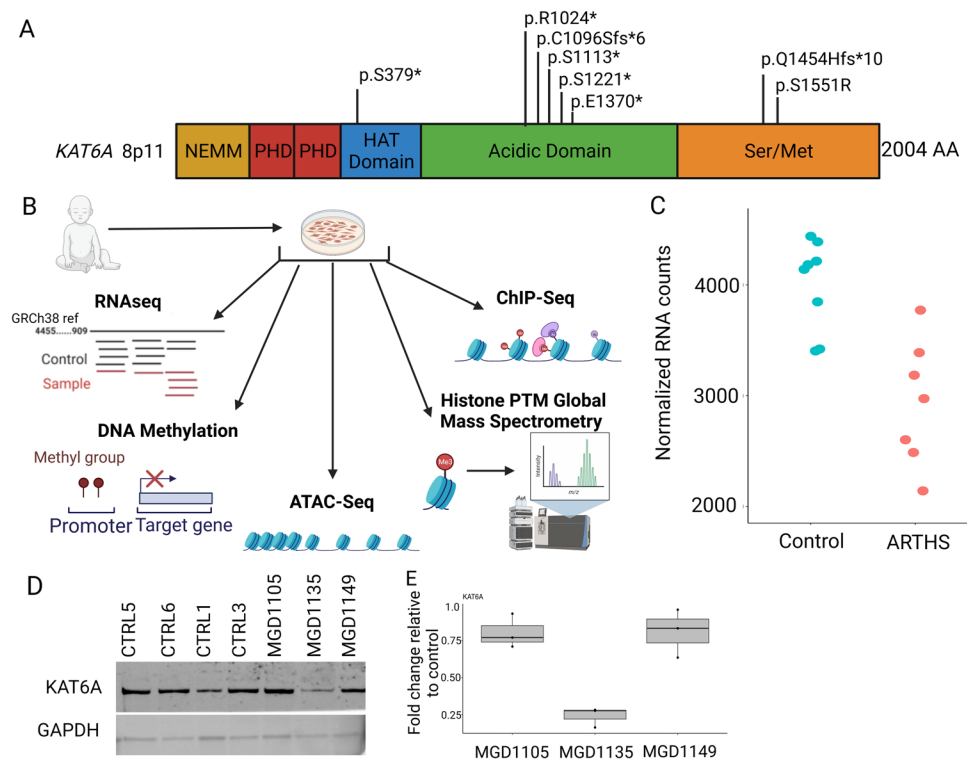
Fig. 1 *KAT6* mutations in patient-derived dermal fibroblasts undergo multi-omic profiling. **A** KAT6A protein structure highlighting important domains and location of the eight pathogenic mutations included in our study. **B** In this study, ARTHS fibroblasts were subject to transcriptomics (RNAseq), histone post-translational modification profiling by mass spectrometry, DNA methylation, ATAC-seq, and ChIP-Seq for histone PTMs. **C** Assessment of *KAT6A* mRNA expression from transcriptome profiling show slightly decreased levels of *KAT6A* expression in ARTHS patients compared to controls. **D** Western blot of KAT6A protein shows decreased protein only in MGD1135 which harbors an early truncating nonsense mutation at p.S379*. **E** Quantification of western blots shows an 80% decrease in protein levels in MGD1135

KAT6A mutations lead to modest changes in the transcriptional landscape

We next performed RNAseq analysis from eight controls and seven patients and assessed differential gene expression between the cases and controls. As expected, the ARTHS cases and controls separated out in unsupervised clustering (Fig. 2A) across the 60 genes that were identified as significantly differentially expressed ($p_{adj} < 0.05$) between control and ARTHS fibroblasts. The significantly differentially expressed genes were predominantly protein-coding genes (46/60) but a smaller subset was long non-coding RNA (lncRNA) (9/60) (Fig. 2B, Table S2). An equal number of genes were upregulated and downregulated in patients with ARTHS (50%, 30/60).

We noted that 4 of the 5 genes with the highest fold-change in ARTHS fibroblasts were a set of developmentally regulated genes that belong to the HOX gene cluster. Specifically, this subset of genes include: *HOXC11* ($p_{adj} = 2.8e-3$, log2FC = 7.9), *HOXC10* ($p_{adj} = 8.9e-4$, log2FC = 5.2), *HOXC-AS3* ($p_{adj} = 2.0 \times 1-3$, log2FC = 7.2), *HOTAIR* ($p_{adj} = 3.0e-3$, log2FC = 9.4), and *HOXC-AS2* ($p_{adj} = 1.8e-2$,

log2FC = 3.3) (Fig. 2C, Table S2). The HOXC gene cluster is located on chromosome 12 and, like all the other three *HOX* gene clusters, oriented in a specific order that corresponds to their spatial–temporal expression during early embryonic patterning (Wellik 2007).

Although fibroblasts are not necessarily representative of the most affected cell type in ARTHS, we are able to identify differentially expressed genes highlighting a potential role for KAT6A in differentiated cells. In the ARTHS fibroblasts RNAseq, we identified transcriptionally dysregulated genes with established neural-related function. Here, we highlight 5 neural-related genes (i.e., *CDH18*, *PITX1*, *PCP4*, *PABPC4L*, and *PKNOX2*) that we discovered are significantly upregulated in ARTHS fibroblast compared to unaffected controls. *Cadherin 18 (CDH18)* ($p_{adj} = 4.8e-8$, log2FC = 3.8) is expressed in neurons and implicated in neuropsychiatric disorders (Redies et al. 2012). *Paired like homeodomain 1 (PITX1)* ($p_{adj} = 1.0e-4$, log2FC = 3.5) deletions in humans lead to lower limb deformities and other skeletal anomalies (Klopocki et al. 2012), but the effect of overexpression of *PITX1* remains unknown. *Purkinje cell protein 4*
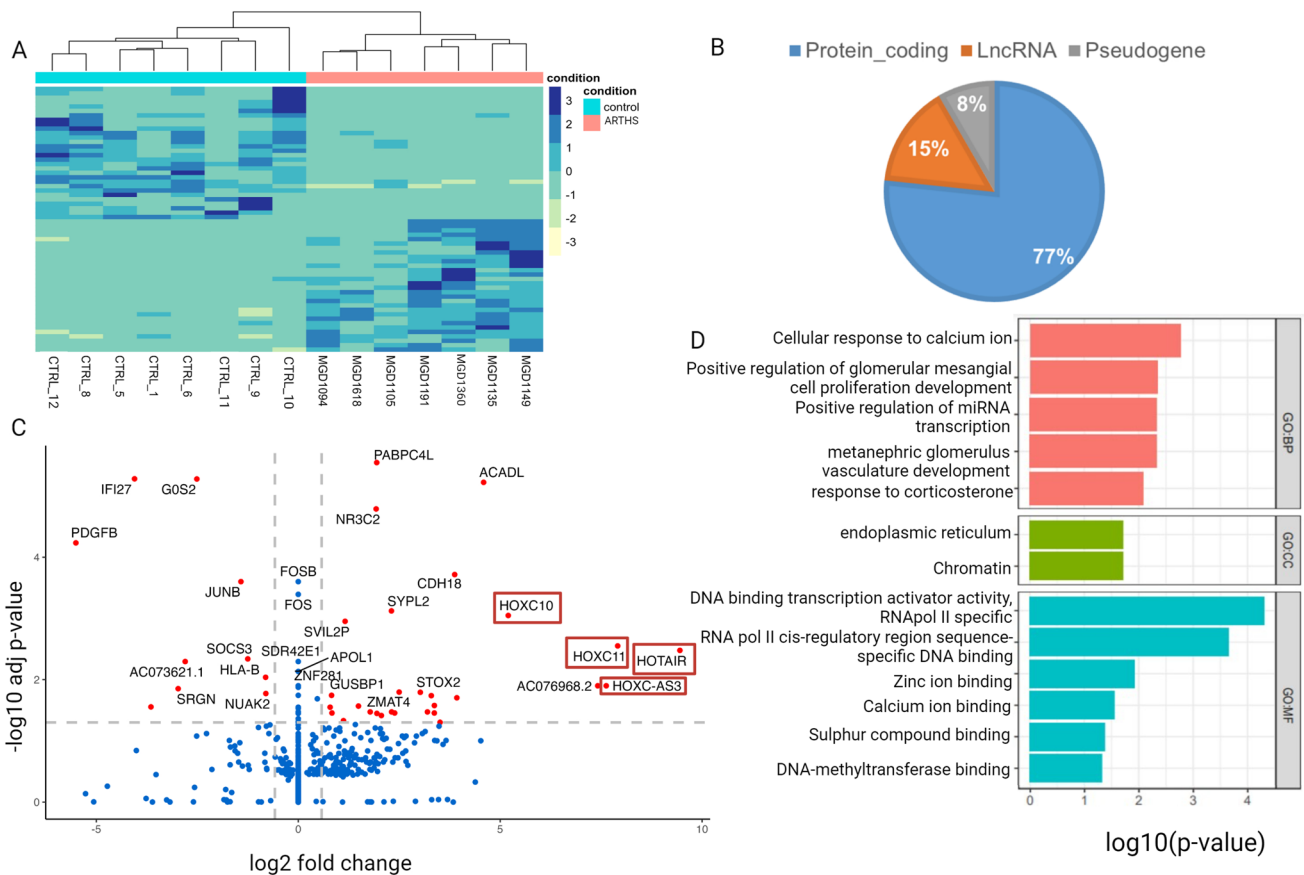
Fig. 2 *KAT6A* mutations cause upregulation of the posterior *HOXC* cluster in ARTHS fibroblasts. **A** Heatmap of 60 significantly differentially expressed genes (DEGs) in ARTHS fibroblast compared with controls. **B** The gene type distribution of the DEGs shows that most of the RNAs identified represent protein-coding genes (blue). **C** Volcano plot of significant DEGs in ARTHS fibroblast compared with control samples highlight 40 genes that have an absolute fold-change of 1.5 and p-adjusted of less than 0.05 (red). All other genes are represented in blue. **D** Gene Ontology bar graph for the differentially expressed genes for biological processes (BP), cellular components (CC), and molecular function (MF)

($PCP4$) ($p_{adj} = 3.3e{-}5$, log2FC = 3.4) is another upregulated gene that is known to modulate calmodulin-binding activity in neuronal cells (Mouton-Liger et al. 2011). Poly(A)-binding protein, cytoplasmic 4-like (PABPC4L) ($p_{adj} = 8.9e{-}4$, log2FC = 5.2) is an RNA-binding protein that has recently been shown to be involved in familial atypical Parkinsonism (Aslam et al. 2019). *PBX/knotted 1 homeobox 2* (*PKNOX2*) ($p_{adj} = 3.67e{-}8$, log2FC = 3.6) is also upregulated. PKNOX2 targets the PBX proteins, and is hence thought to regulate tissue-specific protein expression (Imoto et al. 2001). *PKNOX2* has also been identified as a candidate gene for schizophrenia in two large studies (Wang et al. 2012).

Of the genes that are transcriptionally downregulated, we show that *KAT6A* mutations are associated with significant downregulation of *platelet derived growth factor subunit B* (*PDGFB*) ($p_{adj} = 1.25e{-}8$, log2FC = −5.5) in ARTHS fibroblasts. Endothelial cell-specific deletion of *Pdgfb* leads to vascular defects across highly vascular organ systems such

heart and kidney (Bjarnegård et al. 2004; Levéen et al. 1994).

Gene Ontology enrichment of the differentially expressed genes identified critical pathways and cellular components. The differentially expressed genes were enriched for roles in chromatin, which confirmed the quality of our data given the known role of *KAT6A* as a histone acetyltransferase. We also identified significant enrichment for DNA-binding activity, and calcium and zinc ion binding (Fig. 2D) that are related to transcriptional regulation. The enriched GO terms were driven by the inclusions of *JUNB*, *FOS*, *FOSB*, *HOXC10*, *HOXC11*, and *EGR1* genes (Table S3).
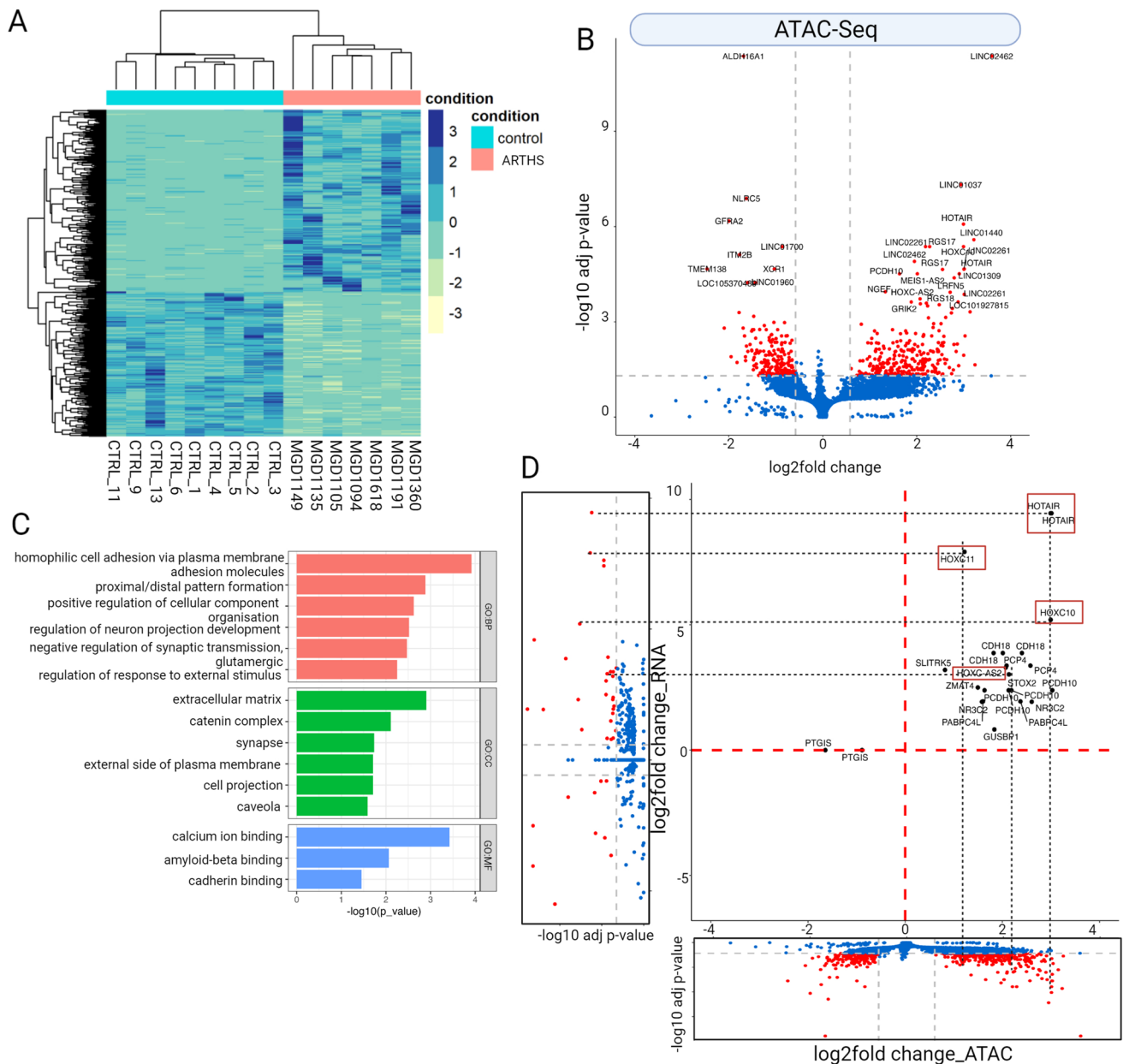
KAT6A mutations alter chromatin accessibility and are associated with differential gene expression

Since KAT6A protein belongs to the class of chromatin modifiers, we next asked whether the differences in gene expression might be associated with changes in chromatin

accessibility. We performed ATACseq on ARTHS and control fibroblasts to see if *KAT6A* mutations are associated with local disruption of chromatin state. Our analysis identified a total of 514 significantly differentially accessible regions, of which 287 peaks (55.8%, 287/514) correspond to more open regions in ARTHS fibroblasts and 227 peaks (44.2%, 227/514) that were more closed in ARTHS fibroblasts (Fig. 3A). Of all the peaks identified, 56.9% (292/514) of the peaks correspond to intergenic regions,

189 peaks correspond to gene-body regions (5′UTR = 5/514, 3′UTR = 3/514, Intron = 181/514), and 3.8% (20/514) peaks correspond to promoters of known genes. These 514 differentially accessible peaks correspond to 414 unique genes (Fig. 3B). In this dataset, of all the closest genes identified, 59% (242/414) are protein-coding, and 38% (156/414) are ncRNAs (non-coding RNAs) (Table S4).

We next wanted to investigate whether differentially accessible regions and the genes associated with these open

Fig. 3 *KAT6A* mutations show that differential chromatin accessibility drives 23% of genes that are differentially expressed: **A** Heatmap depicting 514 differentially accessible ATAC peaks between patient and control samples. **B** Volcano plot highlighting 395 genes that have an absolute fold-change of 1.5 and p-adjusted of less than 0.05 (red). All other genes are marked in blue. **C** Significant GO term distribution of the SigDEGS highlighting biological processes of neuron projection and synaptic transmission. **D** Correlation between ATAC and RNAseq data. It highlights the expression change and chromatin accessibility being changed on *HOXC* genes

chromatin regions had common features or harbored genes that drove specific cellular signaling pathways. We used Gene Ontology (GO) (Raudvere et al. 2019) enrichment analysis on the list of 414 genes located closest to differentially accessible regions, and we identified enrichment for cellular component organization, neuron projection, synaptic regulation, RNA, and protein binding (Fig. 3C, Table S5).

Finally, we asked whether there was a correlation between changes in the chromatin accessible regions with gene expression changes. To address this, we asked whether genes that display increased chromatin accessibility also have higher RNA expression, and vice versa, in ARTHS fibroblasts. We performed correlation between all the genes that were differentially accessible and differentially expressed in ARTHS as determined by the ATACseq and RNAseq analysis, respectively. We showed that several genes in the *HOXC* cluster are differentially expressed (Fig. 2C) and we also find that these same genes have differential chromatin accessibility in our ATACseq data (Fig. 3D). There is a broad accessibility to the chromatin around the *HOXC* gene cluster and is correlated with increased gene expression in ARTHS samples ($HOTAIR$ ($p_{adj}=2.16e-5$, log2FC = 3.01), $HOXC10$ ($p_{adj}=4.27e-6$, log2FC = 2.9), $HOXC$-$AS2$ ($p_{adj}=1.8e-4$, log2FC = 2.07), $HOXC9$ ($p_{adj}=2.7e-4$, log2FC = 2.07), and $HOXC11$ ($p_{adj}=1.2e-2$, log2FC = 1.22) (Table S4). Overall, we find that *KAT6A* mutations lead to localized specific changes to chromatin accessibility.

KAT6A mutations lead to a differential H3K9 and H3K23 acetylation at specific genomic loci.

Previous studies have identified several histone 3 lysine acetylations that were disrupted in the presence of *KAT6A* mutations, including H3K9 (Arboleda et al. 2015; Voss et al. 2009) and H3K23 (Yan et al. 2020; Lv et al. 2017; Huang et al. 2014). We first confirmed that global histone levels were not changed by *KAT6A* mutations by performing western blots using antibodies specific to H3K9ac and H323ac and did not identify any significant global differences between ARTHS and control samples (Fig. 4A–C, Figure S3A and S3B).

Given this, we asked whether there were specific genomic loci where H3K9 and H3K23 were disrupted. We performed Chromatin Immunoprecipitation sequencing (ChIPseq) to identify genomic loci that have disrupted H3K9Ac (Fig. 4D, Figure S3C, TableS6) and H3K23Ac (Fig. 4E, Figure S3D, Table S8) binding in ARTHS samples. The direct number of peaks identified in H3K9Ac was limited to 27 peaks in 24 unique genes. Of these 41% (10/24) were hypoacetylated and 58% (14/24) were hyperacetylated. However, H3K23ac ChIPseq identified 603 significantly differential accessible peaks which correspond to differentially acetylated regions at the H3K23 residue (Fig. 4E, Figure S3D, Table S8). These

632 peaks correspond to 403 unique genes. A gene ontology enrichment for the genes closest to the peak (Figure S4E and S4F, Table S7 and S9) again showed enrichment for terms like RNA and protein binding in the H3K23ac dataset. Two loci that were significantly bound to H3K23 acetylation in ARTHS were directly located in the posterior *HOXC* gene cluster that were differentially expressed and showed differential chromatin accessibility. Together, our data suggest that H3K23 acetylation drives the increased chromatin accessibility around the *HOXC* cluster and increased gene expression that is observed in ARTHS fibroblasts.

Histone-mass spectrometry identifies differential histone post-translational modifications in ARTHS samples

Given that *KAT6A* functions as a histone acetyltransferase, we wanted to investigate the repertoire of histone residues that might be acetylated by the *KAT6A* mutation and those that might be indirectly affected by changes in KAT6A's acetylation pattern. Previous studies have focused on established marks for which there are well-curated antibodies that can quantitatively measure acetylation levels (Yan et al. 2020, 2017; Arboleda et al. 2015) known to be disrupted by the KAT6A/B histone acetylation complex. However, there are many more acetylation marks across H2, H3, and H4 that, to date, have not been comprehensively queried in ARTHS samples. Furthermore, assessment of not just acetylation marks but also changes to alternative post-translational modifiers that might represent indirect events shed light on the combinatorial role of different histone post-translational modifications.

To address this question, we performed mass spectrometry (MS) on histone extracts from 3 ARTHS and 3 control fibroblast samples to identify differential histone modifications in ARTHS relative to unaffected controls (Table S10). The MS data did not show any global changes to the H3K9ac ($p=0.288$, log2FC = 0.362) or H3K23ac ($p=0.356$, log2FC = 0.253) marks in ARTHS. However, as shown in Fig. 4F, we did observe a significant depletion of several novel histone marks that have not previously been associated with KAT6A: H3K56ac ($p=0.0086$, log2FC = −4.95), H3K9me3 ($p=0.0046$, log2FC = −0.196), H2A.1K13ac ($p=0.0043$, log2FC = −0.649), and H2A.3K13me1 ($p=0.0212$, log2FC = −0.261). We also discovered significantly increased levels of H3K56me2 ($p=0.0273$, log2FC = 1.23) (Fig. 4F). In yeast, H3K56ac impacts global transcriptional activation (Topal et al. 2019; Rufiange et al. 2007), while in mammals, H3K56ac plays a role in cell division and chromosomal organization by depositing this mark in newly formed chromatin (Fang et al. 2022). Histone-mass spectrometry is a powerful tool for comprehensively
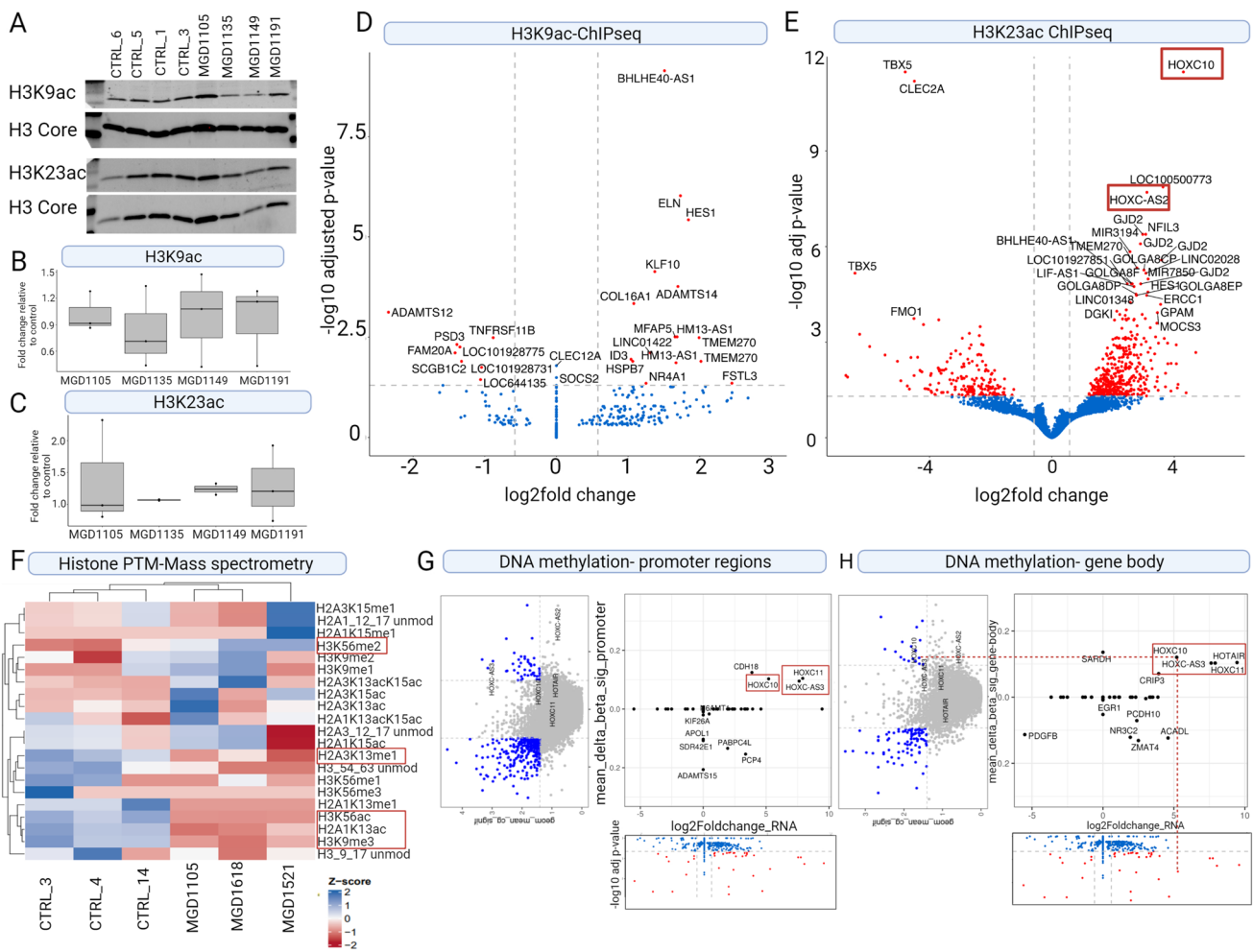
Fig. 4 Histone post-translational modifications and DNA methylation are disrupted in ARTHS fibroblasts around posterior *HOXC* genes. Western blot of global histone acetylation for **A** H3K9ac and H3K23ac is performed compared to H3 core. Relative quantification of the westerns does not show significant differences in global PTM levels of **B** H3K9ac; and **C** H3K23ac. Volcano plot showing the genes corresponding to significantly differentially accessible peaks associated with **D** H3K9 acetylation and **E** H3K23 acetylation. Genes showing significant differences (padj < 0.05 and abs(log2FC) > 0.58 are highlighted in red. All other genes are highlighted in blue. **F**

Heatmap of the normalized abundance of single histone PTMs for peptides that were detected and identified as having at least one histone PTM that is significantly different between ARTHS fibroblast and control samples as determined by mass spectrometry. Histone PTMs that were found to be significantly different ($p < 0.05$) between case and control are boxed in red. Correlation between DNA methylation on **G** promoter regions and RNAseq data and **H** gene-body regions and RNAseq data. We observe differential methylation across CpG islands being changed on *HOXC* genes

ARTHS samples have hypomethylated CpGs as compared to controls

Finally, we obtained previously published DNA-methylation data from ARTHS fibroblasts and controls (Bondhus et al. 2022). Our data indicate that *KAT6A* mutations lead to altered epigenome to regulate the expression of specific gene sets. To gain a more in-depth understanding of this regulation, we used the DNA methylation array on bisulfite

converted DNA from five ARTHS and four control cell lines to identify differentially methylated CpGs and regions. DNA methylation array data show that the CpG sites are more hypomethylated in ARTHS samples as compared to controls (Fig. 4G and H, Figure S4A, S4B and S4C). With this dataset, we observed a modest hypermethylation in the genebody region of the *HOXC* genes which was correlated with higher transcription (Fig. 4H). The role of hypermethylated CpGs in the promoter regions has been well established and are associated with gene silencing; however, there have been few studies that have shown that the methylated CpGs in the gene body have an opposite effect and are associated with

active transcription (Hellman and Chess 2007; Laurent et al. 2010; Ball et al. 2009).

Multi-omic data integration shows that *KAT6A* mutations cause overexpression of the posterior *HOXC* cluster genes

Our multi-omics data analysis identified differential chromatin accessibility and acetylation around posterior *HOXC* cluster genes (Figure S5A and S5B). Since *HOX* genes play spatiotemporal roles during development and maintenance of specific cell type, and they manifest a collinear expression pattern, we wanted to see how the chromatin accessibility, epigenome, and RNA expression is maintained around the entire *HOXC* clusters. We created coverage plots mapped around the entire *HOXC* gene locus, we observed no changes to the anterior parts of the *HOXC* cluster, across ATACseq,

ChIPseq for H3K23ac and gene expression (Fig. 5A). We see increased accessibility, acetylation, and expression of the posterior *HOXC* genes (*HOXC11*, *HOXC10*, *HOTAIR*, *HOXC-AS2,* and *HOXC-AS3*) (Fig. 5B) and DNA hypermethylation at CpG sites in the posterior HOX genes (Fig. 5C). Our study indicates that the *KAT6A* mutations have an important impact on transcriptional regulation of the posterior *HOXC* genes.

The knowledge gained from this multi-omics data integration coupled with the current literature around *KAT6A* and *HOX* genes indicate that different types of *KAT6A* genetic mutations in model organisms have differential effects on HOX gene expression. Overexpression of HOX genes have been associated with cancer (Kim et al. 2019; Fang et al. 2021) and motor neuron defects (Misra et al. 2009). We show that in our model, which is unique in that it leverages human fibroblasts from ARTHS patients, truncating *KAT6A*
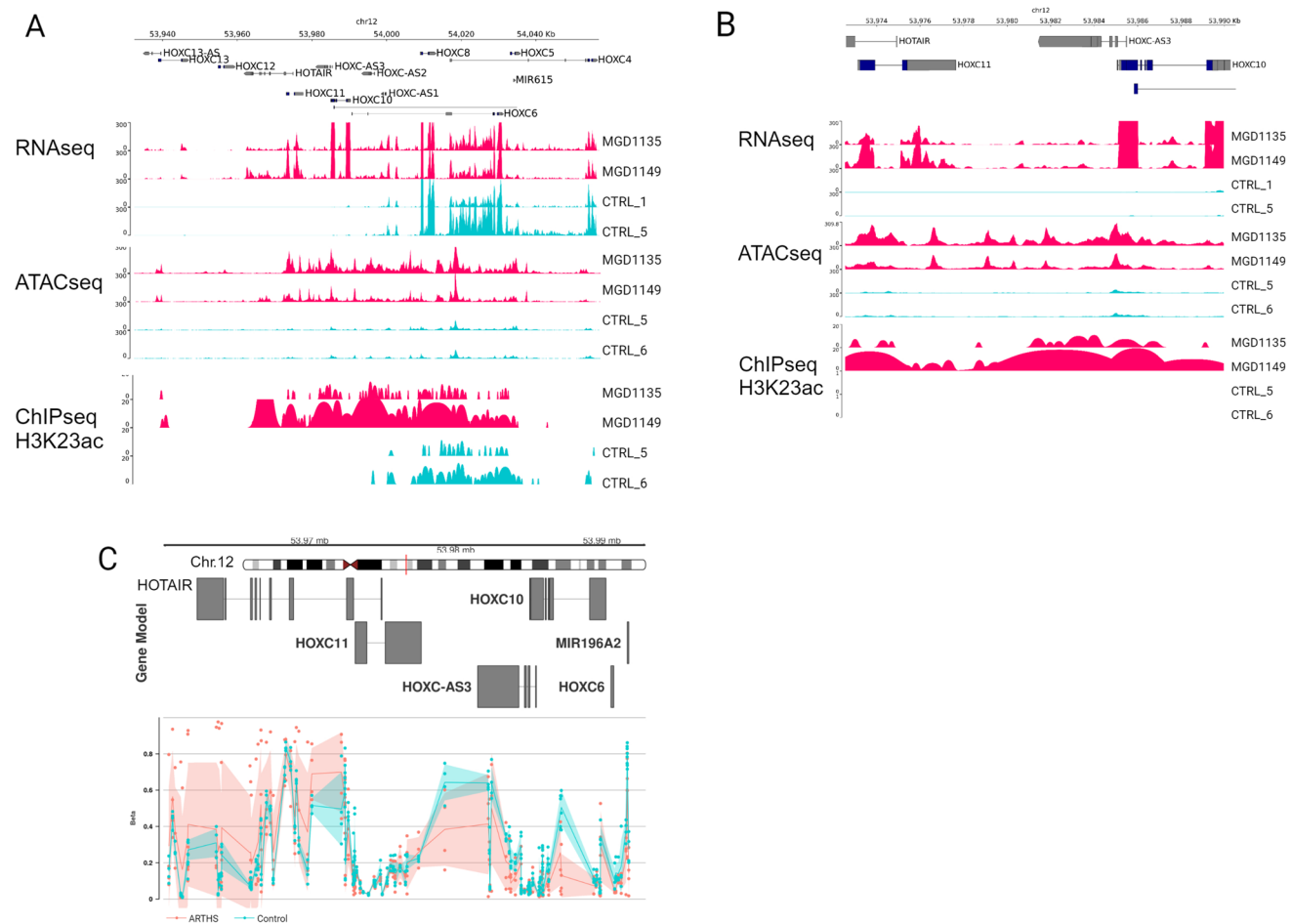
Fig. 5 ARTHS mutations regulate the expression of posterior *HOXC* genes through H3K23 acetylation. Multi-omic epigenomic and transcriptomics data from ARTHS samples compared to controls show clear regulation of posterior *HOXC* cluster genes. **A** Coverage tracks for RNAseq, ATACseq, and ChIPseq across the entire *HOXC* cluster. **B** Zoomed-in coverage of RNAseq, ATACseq, and ChIPseq across the posterior *HOXC* genes (*HOXC10*, *HOXC11*, *HOTAIR*, *HOXC-AS2*, and *HOXC-AS3*). **C** Differential CpG methylation pattern across the posterior *HOXC* cluster

mutations epigenetically upregulates *HOXC* gene cluster expression. Our study suggests that *KAT6A* mutations in ARTHS fibroblasts cause HOX overexpression which likely contributes to gross motor phenotypes in ARTHS individuals, via mechanisms similar to transgenic overexpression of Hox gene in model organism studies.

Discussion

Our multi-omic study of primary ARTHS fibroblast samples quantitatively assayed the epigenomic-driven changes in gene expression across different pathogenic KAT6A mutations. For the first time, we precisely identify the genomic loci and epigenetic marks that are transcriptionally dysregulated in primary ARTHS samples as a result of pathogenic KAT6A mutation through the systematic integration of our RNAseq, ATACseq, ChIPseq, DNA methylation array, and Histone-PTM-specific mass spec data. Most notably, we identified that differential chromatin accessibility and H3K23 acetylation within the *HOXC* gene cluster drove extremely high expression of *HOXC10*, *HOXC11*, *HOTAIR*, *HOX-AS3*, and *HOXCAS2*. Moreover, our unbiased Histone PTM proteomics approach allowed us to identify depletion of acetylated histone marks at H3K56 and H2A.1K13 in ARTHS fibroblasts compared to controls, which suggests that KAT6A may regulate the acetylation status of these previously undescribed histone marks in the context of ARTHS. Taken together, our multi-omic profiling of primary ARTHS fibroblasts allowed us to glean novel insights into the gene regulatory biology controlled by KAT6A.

The majority of truncating mutations tested here are located in the C-terminal region of the gene, but clinical studies have shown a correlation of late-truncating mutations with more severe clinical presentation. Therefore, it is possible that early- and late-truncating mutations may represent different genetic mechanisms (loss- versus gain-of-function). Additional studies with a larger number of early truncating mutations would be required to answer this question.

The *KAT6A* gene is part of a broader group of genes that, when mutated, cause chromatinopathy syndromes due to disruption of epigenome regulation (Nava and Arboleda 2023). Chromatinopathies affect multiple tissues and organ systems, making the study of patient primary cells (e.g., fibroblast and blood) an invaluable resource for scientific discovery (Nava and Arboleda 2023). Since *KAT6A* functions as an epigene, we can measure the downstream molecular effects through multi-omic measurements of downstream markers, such as: histone PTMs, localization of histone PTMS, chromatin accessibility, and gene expression. To date, this represents the first multi-omic study of ARTHS in which we leverage patient-derived cells to directly interrogate the role of KAT6A on the epigenome and gene expression.

KAT6A mutations disrupt the HOXC gene cluster that is associated with lumbar motor neuron development

The *HOX* genes are an evolutionarily conserved class of genes, initially described in drosophila, that play a significant role in differentiation and anterior–posterior body patterning (McGinnis and Krumlauf 1992; Khan et al. 2023). In vertebrates, the *hox* gene cluster underwent gene-level duplications resulting in four clusters of *HOX* genes located on different chromosomes (Sánchez-Herrero 2013; Nolte et al. 2015). The *HOX* cluster genes are expressed in a co-linear fashion and coordinate temporal and spatial patterns during development. The clusters allow for some redundancy within essential processes and fine tuning of the process by which the three germ layers differentiate into cell, tissues, and organ systems in three-dimensional space (Kmita and Duboule 2003; Dasen and Jessell 2009).

We and others have identified disrupted expression of HOX genes in various model systems and tissues where the *KAT6A* gene is mutated. In normal development, mouse studies have shown that at E9.5, *Hoxa,* and *Hoxb* are localized to the hindbrain, while *Hoxc* and *Hoxd* are localized to the spinal column at E12.5. *Hoxc10* and *Hoxc11* are restricted to the lumbar region of the spinal column and are essential to hind-limb motor neuron development (Hostikka and Capecchi 1998; Wellik and Capecchi 2003; Wu et al. 2008; Dasen and Jessell 2009). The upregulation of posterior *HOXC* cluster genes in our data may have an impact on the early development of motor neurons, potentially contributing to the hypotonia and gross motor delay observed in ARTHS patients. In our dataset, we identified the upregulation of five genes from the *HOXC* cluster: *HOXC10*, HOXC11, *HOXC-AS3*, *HOXC-AS2*, and *HOTAIR*. Interestingly, the previous studies have primarily associated defects in motor neuron development with the absence of Kat6a/Moz/Myst3 and, consequently, the *Hox* cluster genes.

While much attention has been focused on *HOX* gene deletions, the effect of *HOX* overexpression in normal development remains unclear. Notably, a study in a chicken embryo model demonstrated that overexpression of *Hoxd10* leads to a reduction in the total motor neuron population and shifted the population toward increased lumbar motor neurons (Misra et al. 2009). Similarly, in mice, the duplication of *Hoxd11* results in altered development of axial skeletal tissue (Boulet and Capecchi 2002). Overexpression of the *HoxC* cluster genes has not been independently studied but would be required to link developmental *HoxC* overexpression to a specific aspect of the ARTHS phenotype in patients.

While our findings in ARTHS patients did not reveal changes to *KAT6A* expression at the transcript or protein level, we believe that the presence of the truncated protein

product causes mistargeting of the KAT6 complex and disruption of *HOXC10 and HOXC11* gene expression in ARTHS fibroblasts. Further studies in complex model systems that directly link the *KAT6A* mutations to these loci and histone marks would be required to prove the direct effects of *KAT6A* mutations on these critical regions.

KAT6A's effect on *HOXC10* gene expression is conserved across different cell types derived from ARTHS patients

The diversity enabled by epigenome organization is a major mechanism that allows a single genome to give rise to thousands of diverse cell types in the human body (Nava and Arboleda 2023). While the genome remains the same, the epigenome coordinates the structure and conformation of specific genomic loci to direct differentiation of each cell type. *KAT6A* is an epigene, which is defined as a protein-coding gene that regulates or affects the epigenome structure and function (Nava and Arboleda 2023). Therefore, disruption of *KAT6A* function is known to affect acetylation of specific histone marks in the KAT6 complex (Zu et al. 2022). But how the KAT6 complex is directed toward specific genomic loci for acetylation during growth and differentiation is not completely understood, especially in the context of human biology (Nava and Arboleda 2023). Strikingly, upon comparing the genes that were differentially expressed in our ARTHS fibroblast data to genes that were differentially expressed in ARTHS cerebral organoids, we uncovered that *HOXC10* was significantly upregulated in ARTHS across both independent RNAseq datasets in comparison to unaffected individuals (Nava et al. 2023). This suggests that some of the *HOXC* cluster genomic targets of KAT6A are conserved across different cell types in ARTHS, which might be due to the KAT6A mutation or an imprint from the iPSC derivation from the fibroblast cells. Studies in mice, zebrafish, and medaka fish have shown that *KAT6A* deletion decreases expression of specific *HOX* gene cluster expression (Hibiya et al. 2009; Voss et al. 2009; Miller et al. 2004); our study suggests that *KAT6A* truncating mutations may have some abnormal, potentially gain-of-function effects causing upregulation of specific *HOX* genes and clusters. However, this molecular role requires additional study to demonstrate definitive proof of the gain-of-function effects in model systems.

Global effects on specific histone acetylation modifications due to *KAT6A* mutations

KAT6A is a lysine acetyltransferase enzyme that is involved in lysine acetylation of Histone 3 N-terminal tail (Dreveny et al. 2014; Lv et al. 2017). However, most studies are done on a subset of possible acetylation markers, based on the availability of histone-mark specific antibodies. In ARTHS fibroblasts, global histone acetylation of H3K9 or H3K23 was not decreased and ChIPseq studies identified a moderate number of loci that showed differential H3K9 or H3K23 acetylation (Fig. 4D and E) several of which correlated with the differential gene expression. We performed the first unbiased histone mass spectrometry assays to comprehensively explore all histone PTMs that are directly or indirectly altered by *KAT6A* mutations. We believe that mass spectrometry assay has broad potential to explore global changes across histone modifications. These, coupled with ChIP-seq to assess localization of these histone marks, studies show that the commonly assayed H3K9ac or H3K23ac marks did not show significant changes to global acetylation by mass spectrometry or significant changes on the localization of these marks (Fig. 4A-E). Our study has identified two novel marks that show global depletion upon KAT6A mutations in fibroblasts: H3K56ac and H2AK13ac.

The role of H3K56ac has been well established in yeast, where it plays crucial roles in chromatin remodeling and DNA repair (Xu et al. 2005; Shuttleworth 1991). Recent studies in mammals also support the role of H3K56ac in chromatin stability and DNA damage and repair (Yuan et al. 2009). These are consistent with earlier studies demonstrating KAT6A's role in DNA damage and repair (Rokudai et al. 2013). H3K56ac has been shown to facilitate the binding of certain pluripotency markers or transcription factors in mouse embryonic stem cells (Tan et al. 2013). Our mass spectrometry data analysis also shows changes in H2A13Kac in ARTHS samples as compared to control. H2AK13ac has not been extensively studied and more in-depth investigations on the gene regulatory effect of H2AK13ac are required; however, it is believed that both H2A and H2B play essential roles in chromatin organization.

Limitations of the study

While this is the first large-scale study of KAT6A epigenomic and transcriptomic regulation, this study has several limitations. Previously published datasets have shown that age and sex have a significant effect on gene expression and DNA methylation (Lee and Lee 2017; Jones et al. 2015). All of our data sets are matched based on sex; however, given the rareness of these disorders, collecting fully age-matched fibroblasts is challenging. A limitation in this study of KAT6A mechanisms is that our control cohort has a higher average age compared to the ARTHS-patient cohort. While this might have some influence on our study, we do not believe it to be the driving force in the gene expression differences. Our multi-omic approach identified five genes that were significantly dysregulated across all datatypes; these genes were also tested as a part of larger study on

effects of age on breast cancer (Lee and Lee 2017; Jones et al. 2015) and none of these genes were affected by age.

A second source of variation is the source of the fibroblast cells (Rinn et al. 2006; Forte et al. 2022) which retain an imprint of the location from which they are derived (Kirk et al. 2021) and potential drift of a given clone after multiple cell culture passages. To alleviate the effect of biopsy source, all of our fibroblasts were derived from skin punch biopsies taken from the forearm or thigh region. All cells were cultured under the same conditions and similar passage number at the time of data acquisition. The best control for age, sex, and genetic background would be to leverage the CRISPR–Cas9 system to correct the mutation in the patient-derived cell lines and using them as control. While this has the risk of off-target effects and is cost-prohibitive, it would remove other sources of variation allowing for only effects of the mutation to be evident. An alternative would be to include multiple independent clones for each line; however, given the fact that we have eight patient samples, this might be less relevant compared to conventional cell lines. Despite these limitations, we believe that the effect size of the highly pathogenic *KAT6A* mutations supercedes the effects of sex, origin or age (Figure S2D and E).

Our findings expand on previous studies showing that *KAT6A* mutations disrupt *HOX* gene expression but that the gene regulation by KAT6A is likely to be cell-type and tissue-specific. Furthermore, our newly described connection to H3K56 acetylation and H2AK13 acetylation highlights the critical role of unbiased approaches in identifying associated histone marks that are globally disrupted by epigenes. To gain a comprehensive understanding of the role of *KAT6A* mutations in human development and disease, further in-depth studies using relevant human tissue types or model systems are warranted. Such investigations will contribute to unraveling the precise mechanisms underlying how *KAT6A* mutations disrupt normal development and disease.

Methods

Cell culture

Skin punch biopsies were performed on patients with a confirmed pathogenic KAT6A mutation and on control individuals under IRB#11-001087 approved by the UCLA Institutional Review Board. Skin punch biopsies were performed and processed to create dermal fibroblast lines in the UCLA Pathology Research Portal. Control neonatal, GM01651 and GM00323 fibroblast lines were obtained from the Coriell Institute for Medical Research in Camden, NJ. All control lines ($n = 14$) are unrelated to the patient lines. Fibroblast cell lines were grown in DMEM (Gibco™ #11995073), 10% FBS(Gibco™ #10–082-147), 1% Non-essential Amino

Acid (100X, Gibco™, 11140-050) and 1% PenStrep (100X, Gibco™ #15,140,122) at 37℃ in 5% CO_2 incubators. Cell lines were tested for mycoplasma on a monthly basis.

RNAseq library preparation and analysis

RNA was extracted from cells grown to 80–90% confluence using the PureLink RNA mini kit (Invitrogen #12183018A). Samples were processed for stranded total RNAseq libraries. The rRNA was depleted using the Qiagen QIAseq Fastselect-rRNA/Globin Kit (Qiagen #335377) followed by illumina TruSeq® Stranded Total RNA Library Prep Gold (illumina #20020599). Samples were multiplexed, and RNA sequencing was performed at the UCLA Technology Center for Genomics and Bioinformatics Core Facility at UCLA and sequenced on an Illumina NovaSeq for an average of 30 million reads per sample. Raw read quality, adaptor content, and duplication rates were assessed with FastQC. Raw reads were then aligned against the Gencode human genome version hg38 (GRCh38) version 31 using STAR 2.7.0e with default parameters (Dobin et al. 2013). Gene counts from raw reads were generated using featureCounts 1.6.5 from the Subread package. For each gene, we counted reads that uniquely mapped to the gene's exons.

Differential expression was quantified using DESeq2 v1.34.0 (Love et al. 2014). Genes with an adjusted p value (Benjamini–Hochberg correction) less than 0.05 were considered as significantly differentially expressed (Wald's test). LFC shrinkage of the DESeq object was done with apeglm (Zhu et al. 2019). A heatmap was created from normalized counts from the DESeq using R-package pheatmap (Kolde 2015), which shows the top differentially expressed (DE) genes based on adjusted p values less than 0.05. Color on legend represents normalized counts. The top DE genes are listed in Table S2.

ATACseq library preparation and analysis

ATACseq was performed with 50,000 cells in both control and patient-derived dermal fibroblast cell lines. ATACseq library generation was performed as described in (Corces et al. 2017). Samples were run on the Agilent DNA TapeStation to confirm tagmentation pattern, and the genomically barcoded libraries were multiplexed and run on the Hiseq3000 with paired end, 150 bp libraries for a minimum of 30 million reads per sample.

ATACseq data were analyzed using an in-house bioinformatics pipeline (Lin et al. 2023). Briefly, the quality of reads was assessed using FastQC. Raw reads were then aligned to GENCODE Human genome version hg38 (GR38) version 31 using BWA-MEM (Li 2013). BAM files were then sorted, indexed, and filtered against chrX, chrY, and MT reads using *SAMtools*. *Picard* tools were then used to generate insert

size histograms and remove duplicates from BAM files. Narrow peaks from each sample were called using MACS2 callpeak (Gaspar 2018); any peak that overlapped by at least one base was then merged using BEDtools (Quinlan and Hall 2010) *merge*. Reads overlapping merged peaks were counted using featureCounts (Liao et al. 2014). *DESeq2* (Love et al. 2014) was used to identify differentially open peaks between disease and control samples. Peaks with p-adj value (Benjamini–Hochberg) less than 0.05 were classified as significantly differentially open (Wald's test), and fold changes were shrunk using approximate posterior estimation for GLM coefficients. LFC shrinkage of the DESeq object was done with apeglm (Zhu et al. 2019). Significant peaks were identified as promoter peaks if their distance from their respective closest gene was less than 1 kb upstream or 2 kb downstream relative to the gene transcription start site.

ChIP-Seq sample processing and analysis

The dermal fibroblast cells were grown to 80–90% confluency and then fixed using fresh media containing 1% formaldehyde. Cells were incubated in formaldehyde media for 10 min, and then, the reaction was neutralized with 1.25 M glycine. After this, the media was removed and cells were washed twice with ice-cold PBS containing protease inhibitors. Post PBS wash, the cells were scraped and collected in a 1.5 ml tube. The cell pellets were resuspended in ChipLysis buffer and sonication was performed using Diagenode bioruptor at high for 45 cycles (30 s on, 30 s off). 5 µl of sonicated product was de-crosslinked and purified to check optimal lysis (200–600 bp).

The sonicated chromatin was pre-cleared with Antigen G dynabeads for 2 h on a nutator at 4 °C. The pre-cleared chromatin was then incubated with 5 µl of H3K9ac (Cell signaling #9649), 2 µl of H3K23ac (AbCam #ab177275), and 4ul of IGg (Cell signaling #2729) at 4 °C overnight. The next day, the antibody-chromatin complex was incubated with antigen G dynabeads for an additional 2 h on a nutator, and then, the beads were washed by a series of buffers and then de-crosslinked at 65 °C overnight. Post de-crosslinking, the DNA was purified and processed for library preparation using illumina Chip library prep (Illumina #15034288). Chip libraries were sequenced on illumina Hiseq3000 using single read 75 bp read-length.

ChIPseq data were analyzed using an in-house bioinformatics pipeline similar to the ATACseq pipeline used above (Lin et al. 2023). In short, the quality of reads were assessed using FastQC. Raw reads were then aligned to GENCODE Human genome version hg38 (GR38) version 31 using BWA-MEM (Li 2013). BAM files were then sorted, indexed, and filtered against chrX, chrY, and MT reads using *SAMtools*. *Picard* tools was used to remove duplicates from BAM files. Narrow peaks from each sample were called

using MACS2 callpeak (Gaspar 2018) using the control sample BAM to account for nonspecific antibody parameter *-c*; based on resulting peaks called, any peak that overlapped by at least one base was then merged using BEDtools (Quinlan and Hall 2010) *merge*. Reads overlapping merged peaks were counted using featureCounts (Liao et al. 2014). Samples with over 15 million reads were kept.

DESeq2 (Love et al. 2014) was used to identify differentially acetylated peaks between disease and control samples. Peaks with *p*-adj value (Benjamini-Hochberg) less than 0.05 were classified as significantly differentially captured (Wald's test), and fold changes were shrunk using approximate posterior estimation for GLM coefficients. LFC shrinkage of the DESeq object was done with apeglm (Zhu et al. 2019).

DNA methylation library preparation and analysis

DNA methylation generated from ARTHS and control fibroblasts were processed as described in (Bondhus et al. 2022; Awamleh et al. 2022). Briefly, DNA was extracted from both control and patient-derived dermal fibroblast cell lines. The extracted DNA was bisulfite converted and run on the Illumina MethylationEPIC Array as previously described (Mansell et al. 2019) at the UCLA Neuroscience Genomics Core. The MINFI package (Aryee et al. 2014) was used to perform QC on the resulting idat files. Probes overlapping SNPs and those on the sex chromosomes were filtered out. After QC and filtering, 832,159 measured CpGs remained of the 865,919 measured CpGs included on the MethylationEpic Array. Background correction (Triche et al. 2013) and functional normalization (Fortin et al. 2014) were used for preprocessing and normalization of individual probes.

Gene ontology enrichment

Gene ontology over-enrichment tests were completed using g:Profiler (Raudvere et al. 2019) by submitting the either the differentially expressed genes or the closest genes to significantly differentially open peaks against all genes from the Gencode hg38 annotation, version 31. Gene ontologies were classified as significantly enriched when p-adj (Benjamini–Hochberg) was less than 0.05 (hypergeometric test).

Cytoplasmic and nuclear protein extraction

The cells were harvested using trypsin–EDTA and centrifuged at $500 \times g$ for 5 min, and the cell pellet was washed with PBS. The dry cell pellet was then processed for nuclear and cytoplasmic protein extraction as per the manufacturer's protocol (Thermo Scientific™NE-PER™ Nuclear and Cytoplasmic Extraction Reagents, #78833). For western blotting assay, 10 µg of the nuclear extract protein were loaded on the

4–15% Criterion™ TGX Stain-Free™ Protein Gel (Bio-Rad #5678083). The protein was transferred to the nitrocellulose (Bio-Rad #1704271) using a turbo-transfer system, blocked with 1X TBST with 5% non-fat milk for an hour at 4 °C, and then incubated with Anti-KAT6A/MOZ antibody (Santa cruz #sc-293283, at 1:500) overnight. The blots were then washed and detected with IRDye® 800CW secondary antibodies (LI-COR #926–32210 and #926–32213). For control antibody, the blots were incubated in HDAC2 (cell signaling #5113, at 1:1000) and hFAB™ Rhodamine Anti-Actin (Bio-Rad #12,004,163 at 1:2000 dilution). All experiments were performed in triplicate.

Histone extraction

For histone extraction, the cells were washed twice with ice-cold dPBS (Gibco #14190144) containing 0.5 nM sodium butyrate. 750 µl of prepared lysis buffer (10 mM HEPES, 1.5 mM $MgCl_2$, 10 mM KCl, 0.34 M Sucrose, and 10% glucose) with fresh 1× protease (Thermofisher, 78442) and 1 mM of DTT (Thermofisher, A39255) were added to each plate. Cells were scraped and transferred to an eppendorf tube, and 1% Triton-X was added to a final concentration of 0.1%. Tubes were incubated on ice for 8 min followed by centrifugation for 5 min at 4 °C and 4500×g. After aspirating the supernatant, the nuclei were lysed by adding 750 µl of lysis buffer and then spun again at 4 °C at 4500×g for 2 min. The supernatant was aspirated, 50 µl of extraction solution (50 µl 10% glycerol, 6 µl concentrated H_2SO_4, 1 µl BME, and 443 µl water) was added to the pellet, and the tubes were incubated on ice for 10 min and then centrifuged at top speed for 10 min at 4 °C. 50 µl of the supernatant was transferred to a new tube, and 12.5 µl of 100% TCA was added. After being vortexed briefly, the tubes were centrifuged at 4 °C for 10 min at top speed. The supernatant was aspirated, and 1 ml of cold ethanol was added. The tubes were vortexed and then incubated at −80 °C for 10 min before being centrifuged again at 4 °C for 10 min at 16,000×g. The ethanol was aspirated and the pellets were allowed to air dry for at least 20 min. Histone pellets were dissolved in 100 µl water. For quantification of histone acetylation levels, 250 ng to 1 µg of histones were loaded on the gel and were detected for H3K9ac (Cell Signaling #9649, 1:1000 dilution) H3K23ac (Cell Signaling #8848, 1:1000 dilution) and H3core (Cell Signaling #3638, 1:1000 dilution).

Histone extraction for mass spectrometry

Fibroblasts isolated from unaffected controls ($n = 3$) and ARTHS patients ($n = 3$) were grown as previously described until cells were approximately 80–90% confluent. Cells were then harvested, washed once with DPBS, pelleted, snap-frozen, and stored at −80 °C for subsequent histone extraction and mass spec analysis. Histones were extracted from snap-frozen pellets using a previously published protocol (Sidoli et al. 2016). Briefly, the cells were thawed and lysed on ice using a nuclear isolation buffer (NIB) containing NP-40. After lysis, the cells were centrifuged and washed in NIB without NP-40 to obtain pure nuclear extracts. After this, the nuclei were resuspended in 0.2 M H_2SO_4 and incubated on constant rotation a 4 °C for 2–4 h. Then, the samples were centrifuged at 3400×g for 5 min and supernatants were collected in fresh tubes. The histones were precipitated using chilled 100% TCA and incubated for either 1 h/overnight. The sample tubes were centrifuged and the nuclei pellets were washed with ice-cold acetone and dried and dissolved in nuclease free water. Once the histones were quantified, about 20 µg of histones were dissolved in 50 mM NH_4HCO_3 pH 8.0 to a final concentration of 1 µg/µl. Then, fresh propionylation reagent was added to the sample in 1:2 v/v ratio, and NH_4OH was quickly added in 1:3 v/v ratio to keep maintaining the pH 8.0. The samples were vortexed and incubated for 15 min at room temperature. The propionylation procedure was performed again for a total of two rounds. The samples were then completely dried down using a vacuum evaporator and resuspended in 50 mM NH_4HCO_3 pH 8.0 to achieve a concentration of 1 µg/µl of protein. The samples were digested using trypsin (1:50 ratio wt/wt) and incubated overnight at 37 °C. The Digestion was stopped by storing the samples at −80 °C. The samples were dried down using vacuum and then resuspended in 50 mM NH_4HCO_3 to make up the volume to 20 µl. After this, the samples undergo another two rounds of propionylation as mentioned above. The samples were desalted by C18 stage tips, dried down, and resuspended in 0.1% formic acid to a final concentration of 0.5 µg/µl. Once the samples are ready, they are analyzed using LC–MS.

Author contributions MS and VAA conceptualized and coordinated this study. MS performed experiments and designed the various sequencing approaches. AAN, FV, and BG performed and analyzed proteomics experiments. JL, GE, and DG performed cell culture and western experiments. SS and AW analyzed RNAseq, ATAC-seq, and ChIPseq studies. LB and SA analyzed DNA methylation data. All authors contributed to the writing and editing of the paper.

Data availability DNA methylation study is available under GEO record GSE210484. RNAseq, ATACseq, and ChIPseq data from this study are uploaded into GEO record GSE237023.

Declarations

Conflict of interest None declared.

References

Arboleda VA, Lee H, Dorrani N, Zadeh N, Willis M, Macmurdo CF, Manning MA 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(3):498–506

Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA (2014) Minfi: a flexible and comprehensive bioconductor package for the analysis of infinium DNA methylation microarrays. Bioinformatics 30(10):1363–1369

Aslam M, Ullah A, Paramasivam N, Kandasamy N, Naureen S, Badshah M, Khan K et al (2019) Segregation and potential functional impact of a rare stop-gain PABPC4L variant in familial atypical parkinsonism. Sci Rep 9(1):13576

Awamleh Z, Chater-Diehl E, Choufani S, Wei E, Kianmahd RR, Yu A, Chad L et al (2022) DNA methylation signature associated with bohring-opitz syndrome: a new tool for functional classification of variants in ASXL genes. Eur J Hum Genet EJHG. https://doi.org/10.1038/s41431-022-01083-0

Ball MP, Li JB, Gao Y, Lee J-H, LeProust EM, Park I-H, Xie B, Daley GQ, Church GM (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat Biotechnol 27(4):361–368

Bjarnegård M, Enge M, Norlin J, Gustafsdottir S, Fredriksson S, Abramsson A, Takemoto M, Gustafsson E, Fässler R, Betsholtz C (2004) Endothelium-specific ablation of PDGFB leads to pericyte loss and glomerular, cardiac and placental abnormalities. Development 131(8):1847–1857

Bondhus L, Wei A, Arboleda VA (2022) DMRscaler: a scale-aware method to identify regions of differential DNA methylation spanning basepair to multi-megabase features. BMC Bioinformatics 23(1):364

Boulet AM, Capecchi MR (2002) Duplication of the Hoxd11 gene causes alterations in the axial and appendicular skeleton of the mouse. Dev Biol 249(1):96–107

Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT et al (2017) An improved ATAC-Seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods 14(10):959–962

Dasen JS, Jessell TM (2009) Hox networks and the origins of motor neuron diversity. Curr Top Dev Biol 88:169–200

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-Seq aligner. Bioinformatics 29(1):15–21

Dreveny I, Deeves SE, Fulton J, Yue B, Messmer M, Bhattacharya A, Collins HM, Heery DM (2014) The double PHD finger domain of MOZ/MYST3 induces α-helical structure of the histone H3 tail to facilitate acetylation and methylation sampling and modification. Nucleic Acids Res 42(2):822–835

Fang J, Wang J, Liangliang Yu, Wenxia Xu (2021) Role of HOXC10 in cancer. Front Oncol 11(May):684021

Fang L, Chen D, Zhang J, Li H, Bradford B, Jin C (2022) Potential functions of histone H3.3 lysine 56 acetylation in mammals. Epigenetics 17(5):498–517

Forte E, Ramialison M, Nim HT, Mara M, Li JY, Cohn R, Daigle SL et al (2022) Adult mouse fibroblasts retain organ-specific transcriptomic identity. Elife. https://doi.org/10.7554/eLife.71008

Fortin J-P, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, Greenwood CM, Hansen KD (2014) Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol 15(12):503

Gaspar JM (2018) Improved peak-calling with MACS2. bioRxiv. https://doi.org/10.1101/496521

Genais T, Gigan D, Augé B, Moussalem D, Waltzer L, Haenlin M, Gobert V (2020) The drosophila MOZ homolog enok controls notch-dependent induction of the RUNX gene lozenge independently of its histone-acetyl transferase activity. bioRxiv. https://doi.org/10.1101/2020.07.27.222620

Hellman A, Chess A (2007) Gene body-specific methylation on the active X chromosome. Science 315(5815):1141–1143

Hibiya K, Katsumoto T, Kondo T, Kitabayashi I, Kudo A (2009) Brpf1, a subunit of the MOZ histone acetyl transferase complex, maintains expression of anterior and posterior hox genes for proper patterning of craniofacial and caudal skeletons. Dev Biol 329(2):176–190

Hostikka SL, Capecchi MR (1998) The mouse Hoxc11 gene: genomic structure and expression pattern. Mech Dev 70(1–2):133–145

Huang Fu, Paulson A, Dutta A, Venkatesh S, Smolle M, Abmayr SM, Workman JL (2014) Histone acetyltransferase enok regulates oocyte polarization by promoting expression of the actin nucleation factor spire. Genes Dev 28(24):2750–2763

Hubert KA, Wellik DM (2023) Hox genes in development and beyond. Development. https://doi.org/10.1242/dev.192476

Imoto I, Sonoda I, Yuki Y, Inazawa J (2001) Identification and characterization of human PKNOX2, a novel homeobox-containing gene. Biochem Biophys Res Commun 287(1):270–276

Jones MJ, Goodman SJ, Kobor MS (2015) DNA methylation and healthy human aging. Aging Cell 14(6):924–932

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 Dev 20(10):1321–1330

Kennedy J, Goudie D, Blair E, Chandler K, Joss S, McKay V, Green A et al (2019) KAT6A syndrome: genotype-phenotype correlation in 76 patients with pathogenic KAT6A variants. Genet Med 21(4):850–860

Khan S, Mishra RK, Srivastava S (2023) Chapter 3–Epigenetic regulation of *cis*-regulatory elements and transcription factors during development. In: Singh G (ed) Perinatal and developmental epigenetics. Translational epigenetics, vol 32. Academic Press, Cambridge, pp 71–113. https://doi.org/10.1016/B978-0-12-821785-6.00004-9

Kim J, Bae D-H, Kim JH, Song K-S, Kim YS, Kim S-Y (2019) HOXC10 overexpression promotes cell proliferation and migration in gastric cancer. Oncol Rep 42(1):202–212

Kirk T, Ahmed A, Rognoni E (2021) Fibroblast memory in development, homeostasis and disease. Cells 10(11):2840. https://doi.org/10.3390/cells10112840

Klopocki E, Kähler C, Foulds N, Shah H, Joseph B, Vogel H, Lüttgen S et al (2012) Deletions in PITX1 cause a spectrum of lower-limb malformations including mirror-image polydactyly. Eur J Hum Genet EJHG 20(6):705–708

Kmita M, Duboule D (2003) Organizing axes in time and space; 25 years of colinear tinkering. Science 301(5631):331–333

Kolde R (2015) Pheatmap: pretty heatmaps. R package version 1.0. 8

Kong Y, Grimaldi M, Curtin E, Dougherty M, Kaufman C, White RM, Zon LI, Liao EC (2014) Neural crest development and craniofacial morphogenesis is coordinated by nitric oxide and histone acetylation. Chem Biol 21(4):488–501

Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, Low HM et al (2010) Dynamic changes in the human methylome during differentiation. Genome Res 20(3):320–331

Lee G, Lee M (2017) Classification of genes based on age-related differential expression in breast cancer. Genom Inform 15(4):156–161

Levéen P, Pekny M, Gebre-Medhin S, Swolin B, Larsson E, Betsholtz C (1994) Mice deficient for PDGF B show renal, cardiovascular, and hematological abnormalities. Genes Dev 8(16):1875–1887

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv [q-bio.GN]. arXiv. http://arxiv.org/abs/1303.3997

Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30(7):923–930

Lin I, Wei A, Awamleh Z, Singh M, Ning A, Herrera A, Russell BE, Weksberg R, Arboleda VA (2023) Multiomics of bohring-opitz syndrome truncating ASXL1 mutations identify canonical and noncanonical Wnt signaling dysregulation. JCI Insight. https://doi.org/10.1172/jci.insight.167744

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. Genome Biol 15(12):550

Lv D, Jia F, Hou Y, Sang Y, Alvarez AA, Zhang W, Gao W-Q et al (2017) Histone acetyltransferase KAT6A upregulates PI3K/AKT signaling through TRIM24 binding. Can Res 77(22):6190–6201

Mallo M, Wellik DM, Deschamps J (2010) Hox genes and regional patterning of the vertebrate body plan. Dev Biol 344(1):7–15

Mansell G, Gorrie-Stone TJ, Bao Y, Kumari M, Schalkwyk LS, Mill J, Hannon E (2019) Guidance for DNA methylation studies: statistical insights from the illumina EPIC array. BMC Genomics 20(1):366

McGinnis W, Krumlauf R (1992) Homeobox genes and axial patterning. Cell 68(2):283–302

Miller CT, Maves L, Kimmel CB (2004) Moz regulates Hox expression and pharyngeal segmental identity in zebrafish. Development 131(10):2443–2461

Misra M, Shah V, Carpenter E, McCaffery P, Lance-Jones C (2009) Restricted patterns of Hoxd10 and Hoxd11 set segmental differences in motoneuron subtype complement in the lumbosacral spinal cord. Dev Biol 330(1):54–72

Mouton-Liger F, Thomas S, Rattenbach R, Magnol L, Larigaldie V, Ledru A, Herault Y, Verney C, Créau N (2011) PCP4 (PEP19) overexpression induces premature neuronal differentiation associated with Ca(2+)/calmodulin-dependent kinase II-δ activation in mouse models of down syndrome. J Comp Neurol 519(14):2779–2802

Nava AA, Arboleda VA (2023) The omics era: a nexus of untapped potential for mendelian chromatinopathies. Hum Genet. https://doi.org/10.1007/s00439-023-02560-2

Nava AA, Jops CT, Vuong CK, Niles-Jensen SL, Bondhus L, Ong CJ, de la Torre-Ubieta L, Gandal MJ, Arboleda VA (2023) KAT6A mutations drive transcriptional dysregulation of cell cycle and autism risk genes in an arboleda-tham syndrome cerebral organoid model. bioRxiv. https://doi.org/10.1101/2023.06.17.545322

Nolte C, Alexander TB, Krumlauf R (2015) Mammalian embryo: hox genes. eLS. Wiley. https://doi.org/10.1002/9780470015902.a0000740.pub3

Nolte C, Krumlauf R (2013) Expression of Hox genes in the nervous system of vertebrates. Landes Bioscience, Austin

Perez-Campo FM, Borrow J, Kouskoff V, Lacaud G (2009) The histone acetyl transferase activity of monocytic leukemia zinc finger is critical for the proliferation of hematopoietic precursors. Blood 113(20):4866–4874

Philippidou P, Dasen JS (2013) Hox genes: choreographers in neural development, architects of circuit organization. Neuron 80(1):12–34

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26(6):841–842

Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, Vilo J (2019) g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res 47(W1):W191–W198

Redies C, Hertel N, Hübner CA (2012) Cadherins and neuropsychiatric disorders. Brain Res 1470(August):130–144

Reifsnyder C, Lowell J, Clarke A, Pillus L (1996) Yeast SAS silencing genes and human genes associated with AML and HIV–1 Tat interactions are homologous with acetyltransferases. Nat Genet 14(1):42–49

Rinn JL, Bondre C, Gladstone HB, Brown PO, Chang HY (2006) Anatomic demarcation by positional variation in fibroblast gene expression programs. PLoS Genet 2(7):e119

Rokudai S, Laptenko O, Arnal SM, Taya Y, Kitabayashi I, Prives C (2013) MOZ increases p53 acetylation and premature senescence through its complex formation with PML. Proc Natl Acad Sci 110(10):3895–3900

Rufiange A, Jacques P-E, Bhat W, Robert F, Nourani A (2007) Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1. Mol Cell 27(3):393–405

Sánchez-Herrero E (2013) Hox targets and cellular functions. Scientifica 2013(December):738257

Scott EK, Lee T, Luo L (2001) Enok encodes a drosophila putative histone acetyltransferase required for mushroom body neuroblast proliferation. Current Biology: CB 11(2):99–104

Sheikh BN, Yang Y, Schreuder J, Nilsson SK, Bilardi R, Carotta S, McRae HM, Metcalf D, Voss AK, Thomas T (2016) MOZ (KAT6A) is essential for the maintenance of classically defined adult hematopoietic stem cells. Blood 128(19):2307–2318

Shuttleworth A (1991) Public disquiet on health reforms should be addressed. Prof Nurse 6(10):560

Sidoli S, Bhanu NV, Karch KR, Wang X, Garcia BA (2016) Complete workflow for analysis of histone post-translational modifications using bottom-up mass spectrometry: from histone extraction to data analysis. J vis Exp JoVE. https://doi.org/10.3791/54112

Tan Y, Xue Y, Song C, Grunstein M (2013) Acetylated histone H3K56 interacts with oct4 to promote mouse embryonic stem cell pluripotency. Proc Natl Acad Sci USA 110(28):11493–11498

Tham E, Lindstrand A, Santani A, Malmgren H, Nesbitt A, Dubbs HA, Zackai EH et al (2015) Dominant mutations in KAT6A cause intellectual disability with recognizable syndromic features. Am J Hum Genet 96(3):507–513

Thomas T, Voss AK (2007) The diverse biological roles of MYST histone acetyltransferase family proteins. Cell Cycle 6(6):696–704

Topal S, Vasseur P, Radman-Livaja M, Peterson CL (2019) Distinct transcriptional roles for histone H3–K56 acetylation during the cell cycle in yeast. Nat Commun 10(1):4372

Triche TJ Jr, Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD (2013) Low-level processing of illumina infinium DNA methylation bead arrays. Nucleic Acids Res 41(7):e90

Vanyai HK, Garnham A, May RE, McRae HM, Collin C, Wilcox S, Smyth GK, Thomas T, Voss AK (2019) MOZ directs the distal-less homeobox gene expression program during craniofacial development. Development 146(14):dev.175042. https://doi.org/10.1242/dev.175042

Voss AK, Collin C, Dixon MP, Thomas T (2009) Moz and retinoic acid coordinately regulate H3K9 acetylation, Hox gene expression, and segment identity. Dev Cell 17(5):674–686

Wang K-S, Zhang Q, Liu X, Longyang Wu, Zeng M (2012) PKNOX2 is associated with formal thought disorder in schizophrenia: a meta-analysis of two genome-wide association studies. J Mol Neur MN 48(1):265–272

Wellik DM (2007) Hox patterning of the vertebrate axial skeleton. Dev Dyn off Publ Am Assoc Anatom 236(9):2454–2463

Wellik DM, Capecchi MR (2003) Hox10 and Hox11 genes are required to globally pattern the mammalian skeleton. Science 301(5631):363–367

Wu Y, Wang G, Scott SA, Capecchi MR (2008) Hoxc10 and Hoxd10 regulate mouse columnar, divisional and motor pool identity of lumbar motoneurons. Development 135(1):171–182

Xu F, Zhang K, Grunstein M (2005) Acetylation in histone H3 globular domain regulates gene expression in yeast. Cell 121(3):375–385

Yan K, Rousseau J, Littlejohn RO, Kiss C, Lehman A, Rosenfeld JA, Stumpel CT et al (2017) Mutations in the chromatin regulator gene BRPF1 cause syndromic intellectual disability and deficient histone acetylation. Am J Hum Genet 100(1):91–104

Yan K, Rousseau J, Machol K, Cross LA, Agre KE, Gibson CF, Goverde A et al (2020) Deficient histone H3 propionylation by BRPF1-KAT6 complexes in neurodevelopmental disorders and cancer. Sci Adv 6(4):eaax0021

Yuan J, Mintie Pu, Zhang Z, Lou Z (2009) Histone H3–K56 acetylation is important for genomic stability in mammals. Cell Cycle 8(11):1747–1753

Zhu A, Ibrahim JG, Love MI (2019) Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics 35(12):2084–2092

Zu G, Liu Y, Cao J, Zhao B, Zhang H, You L (2022) BRPF1-KAT6A/KAT6B complex: molecular structure, biological function and human disease. Cancers 14(17):4068. https://doi.org/10.3390/cancers14174068

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