OXFORD

Human Molecular Genetics, 2022, Vol. 31, 21, 3715–3728
https://doi.org/10.1093/hmg/ddac121
Advance access publication date 28 May 2022
Original Article

Kabuki syndrome stem cell models reveal locus specificity of histone methyltransferase 2D (KMT2D/MLL4)

Malvin Jefri[1,2], Xin Zhang[1,2], Patrick S. Stumpf[3], Li Zhang[4], Huashan Peng[1,2], Nuwan Hettige[1,2], Jean-Francois Theroux[1,2], Zahia Aouabed[1,2], Khadija Wilson[5], Shriya Deshmukh[6], Lilit Antonyan[1,2], Anjie Ni[1,2], Shaima Alsuwaidi[1,2], Ying Zhang[1,2], Nada Jabado[6,7,8], Benjamin A. Garcia[5], Andreas Schuppert[3], Hans T. Bjornsson[4,9,10] and Carl Ernst[1,2,7,*]

[1]Psychiatric Genetics Group, McGill University, 6875 LaSalle Boulevard, Frank Common Building, Room 2101.2, Verdun, Montreal, QC H4H 1R3, Canada
[2]Department of Psychiatry, McGill University and Douglas Hospital Research Institute, Montreal, QC H4H 1R3, Canada
[3]Institute for Computational Biomedicine, RWTH Aachen University, Aachen 52056, Germany
[4]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
[5]Department of Biochemistry and Molecular, Washington University School of Medicine in St. Louis, St. Louis, MO 63110, USA
[6]Division of Experimental Medicine, Department of Medicine, McGill University, Montreal, QC H4A 3J1, Canada
[7]Department of Human Genetics, McGill University, Montreal, QC H3A 0C7, Canada
[8]Department of Pediatrics, McGill University and The Research Institute of the McGill University Health Centre, Montreal, QC H4A 3J1, Canada
[9]Faculty of Medicine, University of Iceland, Reykjavik, Iceland
[10]Department of Genetics and Molecular Medicine, Landspitali University Hospital, 101 Reykjavik, Iceland
*To whom correspondence should be addressed at: Department of Psychiatry, McGill University and Douglas Hospital Research Institute, 6875 LaSalle boulevard, Frank Common building, Room 2101.2 Verdun, QC H4H 1R3, Canada. Tel: +1 514-761-6131 ext 3382; Fax: +1 514-762-3023; Email: carl.ernst@mcgill.ca

Abstract

Kabuki syndrome is frequently caused by loss-of-function mutations in one allele of histone 3 lysine 4 (H3K4) methyltransferase *KMT2D* and is associated with problems in neurological, immunological and skeletal system development. We generated heterozygous *KMT2D* knockout and Kabuki patient-derived cell models to investigate the role of reduced dosage of KMT2D in stem cells. We discovered chromosomal locus-specific alterations in gene expression, specifically a 110 Kb region containing Synaptotagmin 3 (*SYT3*), C-Type Lectin Domain Containing 11A (*CLEC11A*), Chromosome 19 Open Reading Frame 81 (*C19ORF81*) and SH3 And Multiple Ankyrin Repeat Domains 1 (*SHANK1*), suggesting locus-specific targeting of KMT2D. Using whole genome histone methylation mapping, we confirmed locus-specific changes in H3K4 methylation patterning coincident with regional decreases in gene expression in Kabuki cell models. Significantly reduced H3K4 peaks aligned with regions of stem cell maps of H3K27 and H3K4 methylation suggesting KMT2D haploinsufficiency impact bivalent enhancers in stem cells. Preparing the genome for subsequent differentiation cues may be of significant importance for Kabuki-related genes. This work provides a new insight into the mechanism of action of an important gene in bone and brain development and may increase our understanding of a specific function of a human disease-relevant H3K4 methyltransferase family member.

Introduction

Kabuki syndrome (KS) (*OMIM# 300867* and *#147920*) is characterized by post-natal growth deficiency, skeletal/-dermatological anomalies and mild to moderate intellectual deficiency (1). It affects approximately 1 in 32 000 live births, and most cases are caused by *de novo* heterozygous mutations in lysine methyltransferase 2D *KMT2D* (2) or lysine demethylase 6A *KDM6A* (aka *UTX*) (3), where *KMT2D* is mutated in greater than 60% of cases.

KMT2D (aka MLL2/MLL4) functions to transfer methyl groups to the fourth lysine residue of histone 3 lysine 4 (H3K4) (4). It is a large protein (~5.5 K amino acids) that contains a SET domain which is required for enzymatic transfer of methyl groups from an S-adenosylmethionine donor to H3K4. KMT2D is one of seven KMT2 family members (KMT2A-G), also known as the mixed lineage leukemia (MLL) or trithorax family and is partially redundant with KMT2C (5). KMT2 family members have two core commonalities: (1) they all transfer a methyl group to H3K4 and (2) they associate with a common group of required binding proteins for their stabilization, called complex of proteins associated with set1 (COMPASS), named for the yeast homologue of the KMT2 family, Set1 (6). These essential partner proteins are WDR5, RBBP5, DPY30 and ASH2L (7) and are needed for an efficient methylation reaction (8).

The precise functional differences between KMT2 family members remain largely unknown. Although the family arose from gene duplications (9), there are significant differences across members with respect to specific peptide domains. It is these domain differences that will likely affect their specific genomic targets, their differing set of unique binding proteins (10,11) and their individual methyl transferase preferences, such

Received: February 14, 2022. **Revised:** May 7, 2022. **Accepted:** May 18, 2022

as whether to add one, two or three methyl groups to H3K4 (10). For example, KMT2D/C have preference for specific binding partners that other families do not show. KMT2C/D can associate with KDM6A, Paired box (PAX) transcription activation domain interacting protein (PTIP) and NCOA6 and preferentially interact with some transcription factors such as the PAX family members, Liver X receptor (LXR) and Peroxisome proliferator-activated receptor (PPAR)-gamma (12). The unique binding to specific proteins and transcription factors likely allows for unique targeting in the genome in tissue-, developmental- and cell type-specific ways. KMT2C/D are predicted to mono-methylate H3K4 at enhancers (13,14). Mechanistically, it is not clear how KMT2D might associate with specific regions, but interactions with transcription factors bound to DNA have been postulated. In blood cells for example, KMT2D is recruited by NFE2 which allows KMT2D to spread across the region and methylate about 40 Kb (15). Alternatively, KMT2D has been postulated to be a component of condensates with KDM6A which segregates the genome via phase separation. Both ideas support a model where specific stretches of chromosome could be methylated by KMT2D (16).

Our understanding of the role of KMT2D comes from several knockout studies, usually involving homozygous knockout of both *Kmt2d* and *Kmt2c* in rodents. But non-human and complete knockout studies are not a model of reduced dosage of *KMT2D* and likely exaggerate or differ from the effects of reduced dosage of *KMT2D*. For this reason, we have made several models of reduced dosage *KMT2D* cells, including heterozygous knockouts but also cell models made from human individuals with KS (heterozygous loss), to understand the genomic preferences of KMT2D in the context of differing dosages. We find that partial loss of *KMT2D* in human stem cells leads to a loss of mono-methylation in at least one, and likely three, specific regions of the human genome, consistent with a spreading model of H3K4 methylation (15), and coincident with the loss of gene expression in these regions.

Results

Stem cell models of Kabuki syndrome

Through simultaneous gene editing and cell reprogramming (17), we generated two heterozygous *KMT2D* Knockout (KO) Induced pluripotent stem cells (iPSC) lines—KMT2D$^{ex2\text{-}del}$ and KMT2D$^{ex2\text{-}ins}$—from a healthy control fibroblast line (Fig. 1B). All control and mutated iPSC lines were validated for pluripotency marker expression and absence of chromosomal anomalies by the cytoscan High-density (HD) array (Fig. 1A). We targeted exon 2 (the first coding exon) of *KMT2D* to create a frameshift mutation. After the selection of successfully edited clones, Sanger sequencing showed two cell lines with different mutations; one had a heterozygous 10 nt deletion in exon 2 (KMT2D$^{ex2\text{-}del}$), and the other had a heterozygous 34 nt insertion in exon 2 of KMT2D$^{ex2\text{-}ins}$ (Supplementary Material, Fig. S1A and B) where 33 of the 34 nucleotides were a duplication of the preceding DNA. Both are predicted to lead to loss of function due to nonsense-mediated decay (NMD). After the purification and expansion of these two cell lines, we performed quantitative polymerase chain reaction (qPCR) to assess mRNA levels of *KMT2D* (Fig. 1C). Surprisingly, we did not detect NMD of mutant alleles given that qPCR showed similar levels of *KMT2D* RNA in mutant models compared with controls, which is inconsistent with the 'rules' of NMD (18). To better understand the allelic effects of gene editing, we wanted to ensure that RNA was transcribed from both alleles along the entire gene body of *KMT2D*. If there is no NMD, then we should detect the mutation in RNA. RNAseq in mutant and isogenic control IPSCs (Fig. 1D, Supplementary Material, Fig. S1C) shows the creation of a novel splice junction between exons 1 and 3 in KMT2D$^{ex2\text{-}del}$ in approximately 50% of reads, consistent with heterozygous gene editing and lack of NMD (Fig. 1E). The presence of the exon1–exon3 splicing event in ∼50% of the reads is consistent with biallelic expression. For RNAseq analysis from the KMT2D$^{ex2\text{-}ins}$ cell line, we expected to detect a novel mutant insertion in RNA which would not have mapped to the reference genome. To this end, we queried RNAseq reads that were filtered out of the primary RNAseq analysis for any reads that might contain this new insertion adjacent to wild-type sequences on either side of the insertion. We found several reads that were consistent with this including one that fully spanned the insertion (Supplementary Material, Fig. S1C), providing unambiguous evidence for the mutant allele in the transcribed RNA. Both frameshift mutations detected in mRNA are predicted to lead to protein loss by putting the coding strand out of frame. If this were true, we would expect a decrease of protein levels of KMT2D in mutant cells lines which we did observe at the predicted KMT2D molecular weight of ∼600 KDa (Fig. 1F).

Locus-specific gene expression changes and loss of *KMT2D* in stem cells

KMT2D is an H3K4 methylase as well as a protein hub for interaction proteins to guide protein/chromosomal interactions (4). H3K4me is associated with transcriptional activation of the marked locus, in which H3K4me1/2 are located predominantly at enhancers and H3K4me3 is located mostly at promoters. We analyzed RNAseq data for case/control expression differences using the two *KMT2D* mutant lines and isogenic controls. We predicted that the loss of *KMT2D* might reveal genes where KMT2D acts to directly add methyl groups to increase gene expression. Unsurprisingly, we found significantly more genes were downregulated than upregulated in the *KMT2D* KO lines (Supplementary Material, Fig. S1D and E). Organizing gene expression data by genes with the most significant decreased expression differences by $\log_2$ fold-change revealed a clustering of chromosomal band locations of genes whose expression was the most significantly affected (Fig. 2A).

The top ranked hit by $\log_2$ fold change is *CLEC11A* in the 19q13.33 band (hg38) which is located 74 Kb from

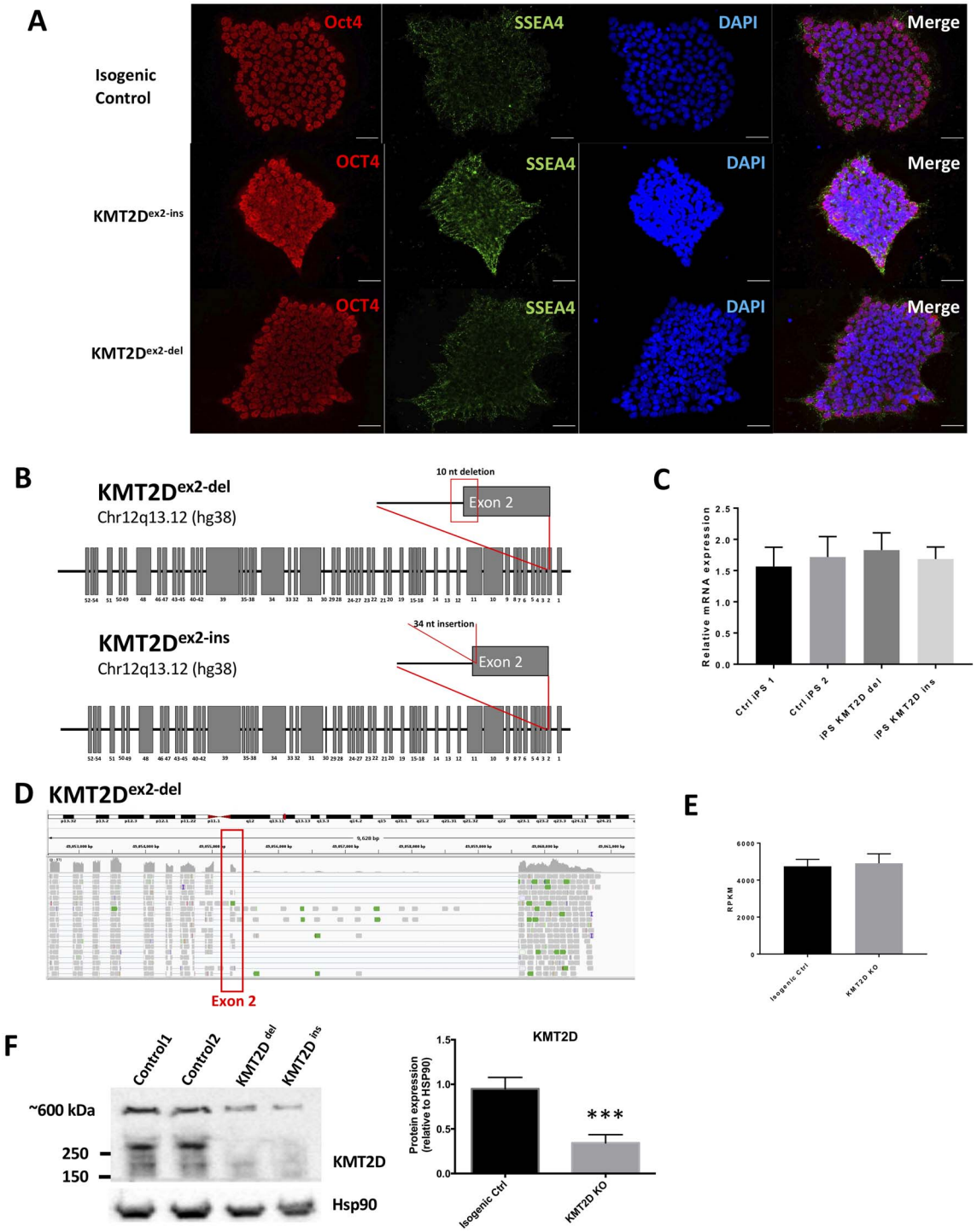

Figure 1. Generation of CRISPR-Cas9 engineered *KMT2D* mutation iPSC lines. (**A**) Immunostaining of pluripotency markers of the control and engineered KMT2D KO iPSCs used in the study. Scale bars indicate 50 μm. (**B**) Illustration of the editing location and size in exon 2 of KMT2D in the two mutation lines generated. (**C**) qPCR analysis of *KMT2D* mRNA expression showing absence of nonsense RNA-mediated decay in the lines carrying mutations. There were four independent samples in the analysis (Control1, Control2, KMT2D$^{ex2\text{-}del}$ and KMT2D$^{ex2\text{-}ins}$). The two independent isogenic control lines were derived from the same fibroblast line used for the generation of the two KMT2D KO lines, reprogrammed and isolated from two different iPSC clones. Each KMT2D$^{ex2\text{-}del}$ and KMT2D$^{ex2\text{-}ins}$ KO line were derived from different gene-edited iPSC colonies expanded from single cells. Each iPSC sample underwent independent RNA extraction, cDNA synthesis and qPCR analysis. (**D**) The alignment and identification of RNAseq reads of exon2 of *KMT2D* of KMT2D$^{ex2\text{-}del}$ iPSC lines. (**E**) The RPKM of *KMT2D* expression of the KMT2D KO cells compared to control cells. (**F**) Western blot analysis of KMT2D expression in mutation cells in comparison to control cells. ***$P \leq 0.001$.

C19ORF81, the second-ranked hit (Fig. 2B). There is no evidence for cotranscription of these genes (i.e. they are not transcribed on the same RNA strand), so it appears that the top two differential expression differences are extremely close to each other in 19q13.33. This could suggest that both are regulated by KMT2D, and in the Kabuki models may fail to acquire a full contingent of H3K4 methylation to fully activate their expression.

A

Gene	log2FC	Adj p	Chr Band (hg38)
CLEC11A	-3.59	4.4E-12	19q13.33
C19orf81	-3.31	1.2E-10	19q13.33
ANKRD1	-3.30	2.1E-05	10q23.31
PCDHA6	-2.68	1.5E-16	5q31.3
TNNI3	-2.14	4.3E-08	19q13.42
C14orf80	-1.92	3.4E-23	14q32.33
LINC01139	-1.88	1.2E-05	1q43
GNA14	-1.79	1.2E-31	9q21.2
RP11-496I9.1	-1.68	1.7E-02	11p15.5
SSPO	-1.64	7.3E-03	7q36.1
CRIP2	-1.63	5.0E-12	14q32.33
MIR205HG	-1.61	2.2E-07	1q32.2
SYT5	-1.60	2.1E-02	19q13.42
GBX2	-1.58	1.7E-05	2q37.2
MMP9	-1.53	6.1E-15	20q13.12
ADAM33	-1.52	2.5E-11	20p13
AHSP	-1.52	9.2E-03	16p11.2

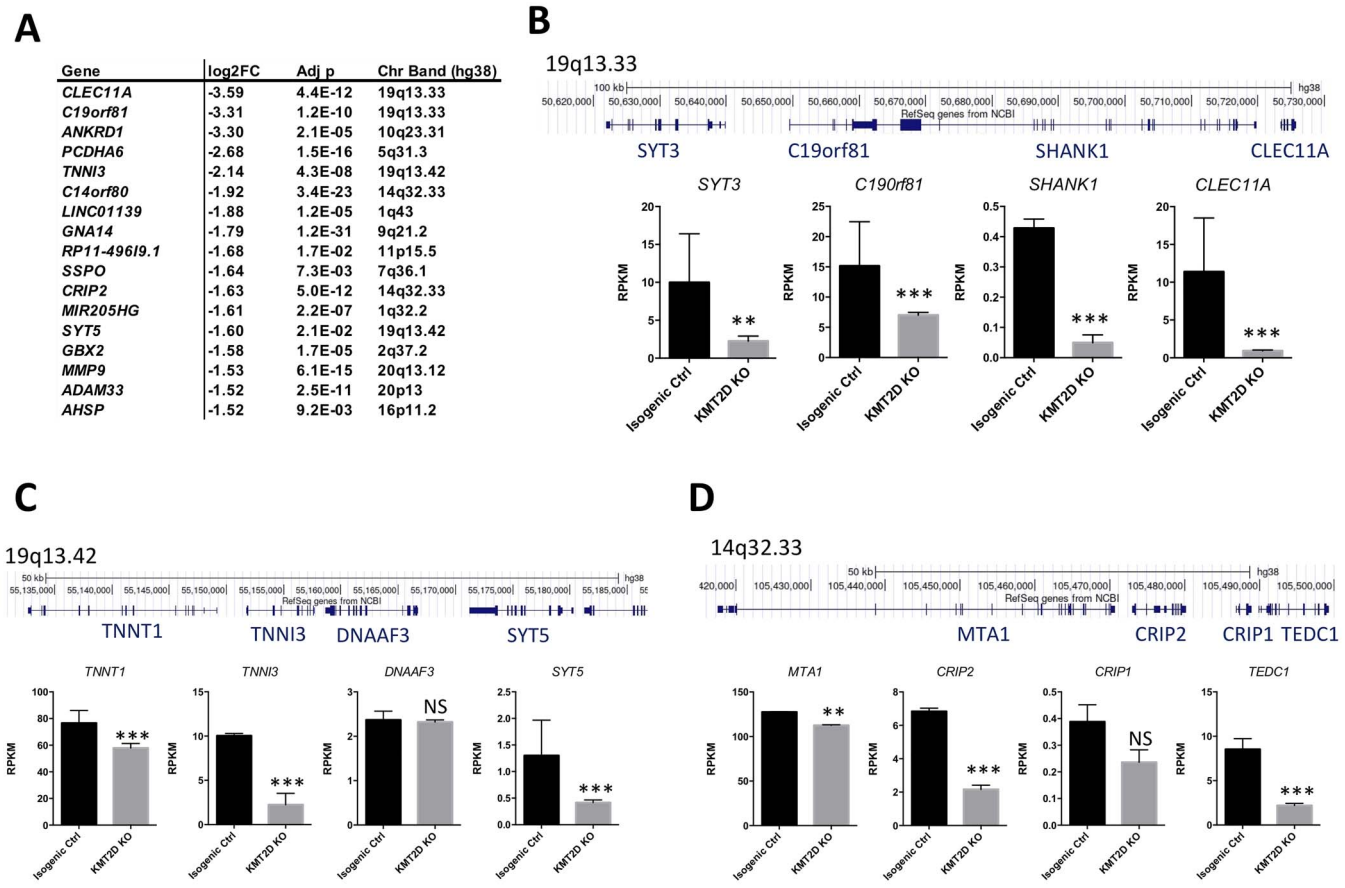

Figure 2. Identification of locus-specific gene downregulation in *KMT2D* deficient cells. (**A**) List of most significant downregulated genes in the RNAseq analysis. There were four independent samples in the analysis (Control1, Control2, KMT2D$^{ex2\text{-}del}$ and KMT2D$^{ex2\text{-}ins}$). The two independent isogenic control lines were derived from the same fibroblast line used for the generation of the two KMT2D KO lines, reprogrammed and isolated from two different iPSC clones. Each KMT2D$^{ex2\text{-}del}$ and KMT2D$^{ex2\text{-}ins}$ KO line were derived from different gene-edited iPSC colonies expanded from single cells. Each iPSC sample underwent independent RNA extraction, cDNA synthesis and sequencing. (**B**) (top) UCSC database snapshot of the downregulated genes locus located at 19q13.33 (hg38) exhibiting the adjacency of the affected genes; (bottom) Differential expression analysis of the RNAseq result indicating the downregulation of the four genes located within the same locus.$^{**}P \leq 0.01$, $^{***}P \leq 0.001$. (**C**) (top) UCSC database snapshot of the downregulated genes locus located at 19q13.42 (hg38) exhibiting the adjacency of the affected genes; (bottom) Differential expression analysis of the RNAseq result indicating the downregulation of the four genes located within the same locus. $^{**}P \leq 0.01$, $^{***}P \leq 0.001$. (**D**) (top) UCSC database snapshot of the downregulated genes locus located at 14q32.33 (hg38) exhibiting the adjacency of the affected genes; (bottom) Differential expression analysis of the RNAseq result indicating the downregulation of the four genes located within the same locus. $^{**}P \leq 0.01$, $^{***}P \leq 0.001$.

KMT2D may affect an enhancer that controls both genes or affect these genes in a broad, regional way. Whichever is correct, if KMT2D affects expression of this locus it might also affect other immediate neighboring genes in 19q13.33. The two other closest genes are *SYT3* and *SHANK1* (Fig. 2B). *SYT3* was the 1272nd most significant ranked gene ($\log_2$FC = −2.12, $P = 0.0045$) and *SHANK1* was filtered from the RNAseq list because of low expression levels (Reads per kilobase of transcript, per million mapped reads (RPKM) mean = 0.4); however, it was still detected in stem cells and showed a log2FC = −3.0, *P*-value = 0.0001. Thus, it too supports a regional effects model, despite having very low expression. *SHANK1* is neuron specific, but we detect presumably untranslated RNA in the stem cells.

Correlated expression of proximal genes has been described in the literature (19); hence, the observed high correlation in the relative expression of neighboring genes on 19q13.33 is not surprising *per se*. To shed more light on the expected versus observed coexpression levels, we looked at the fold-change of genes on chromosome 19 and their nearest cis neighbors (smallest distance between transcription start sites). Overall, this analysis (Supplementary Material, Fig. S2) reveals a poor correlation in expression differences (Pearson correlation coefficient = 0.095). However, *SYT3*, *C19orf81*, *SHANK1* and *CLEC11A* genes are consistently downregulated (red points) whereas G Protein-Coupled Receptor 32 (*GPR32*), which is located adjacent to *CLEC11A* on Chr19, does not display the same expression difference. Notably, the observed odds ratio for finding *SYT3*, *C19orf81*, *SHANK1* and *CLEC11A* proximal genes coregulated versus all other nearest neighbors on Chr19 is 2e3 (*P*-value = 4e-8; fisher's exact test) and hence highly unusual. Together with the validation experiments, we believe that this strongly supports our interpretation that the effect of KMT2D KO is highly specific to the DNA locus containing *SYT3*, *C19orf81*, *SHANK1* and *CLEC11A*.

Moving out from this core locus of four genes, we were unable to detect other gene expression differences

(Supplementary Material, Fig. S3), suggesting that KMT2D regulates only this specific locus of approximately 110 Kb at this chromosomal region. Other loci that showed this phenomenon were the chromosome 19q13.42 and 14q32.33 bands (Fig. 2C and D) though not as consistently across genes as the 19q13.33 region. In genes achieving the $\log_2FC < -1.5$, Padj. < 0.005 criteria, we found a total of 16 chromosome bands (Supplementary Material, Table S2) that had at least two genes differentially expressed in the same band, though which are not necessarily adjacent to each other. We consider these loci potentially suggestive as KMT2D target regions.

Stem cells from individuals with Kabuki syndrome recapitulate engineered control cells

Engineered genetic mutations are an excellent way to probe gene function because they allow for a direct comparison with an isogenic cell line, reducing experimental noise; however, they are not the exact mutations found in Kabuki cases. To determine whether gene expression changes observed in engineered models are similar in KS cases, we made iPSCs from renal epithelial cells collected from a KS case and her biological mother. The use of a healthy first-degree sex-matched control reduces experimental noise since these subjects share 50% of the genome. KS1 underwent clinical ExomeSeq at GeneDx which identified a 4 nt deletion in exon 53 of *KMT2D* consistent with a diagnosis of KS, and which we confirmed by Sanger sequencing in iPSCs (Fig. 3A). A complete clinical description and photo of KS1 can be found in the Supplemental Material. Pluripotency of stem cells and chromosomal integrity were both assessed and were normal for both lines (Fig. 3B) suggesting that heterozygous loss of *KMT2D* is dispensable for iPSC induction and maintenance. We checked whether there was NMD given that a frameshift occurred, but like engineered models, we found no NMD but did find depleted protein (Fig. 3C and D). To test whether gene expression differences from 19q13.33 were also affected in KS1, we made qPCR primers for *SYT3*, *SHANK1*, *C19ORF71* and *CLEC11A*, then validated expression changes first in heterozygous *KMT2D* KO lines (Fig. 3E). Next, we assessed RNA levels of these genes in iPSCs from KS1 and her mother and confirmed depletion in cells derived from a diagnosed KS case (Fig. 3F). Despite the concern that epigenetic memory in iPSCs originating from different somatic origins could potentially cause variation in gene expression (20), we found highly consistent results between the KO (skin-derived) iPSCs and the KS case (urine-derived) iPSCs (Fig. 3E and F) highlighting the robustness of our discovery. Moreover, we ran an off-target effect of CRISPR-Cas9 analysis to identify potential off-targets in our KO cells and found no evidence to suggest that the reduced expression previously found could be due to this reason (Supplementary Material, Table S3).

To rule out that this finding is limited to a particular mutation type, we used iPSCs from another KS case with nonsense mutation in exon 32 of *KMT2D* (21). We find almost identical effect sizes in qPCR done from genes at the 19q13.33 locus in cells made in an independent laboratory and performed by an independent operator (Supplementary Material, Fig. S4). These data support the finding that loss of *KMT2D* leads to the decreased expression of *SYT3*, *C19orf81*, *SHANK1* and *CLEC11A* in stem cells; however, these results could be due to indirect, rather than direct, effects of *KMT2D* loss, an idea we assess in the following section.

Reduction of H3K4 methylation in KS cell models

Given clustering of genes whose gene expression is decreased in *KMT2D* depletion models, we hypothesized that the mechanism of action for this effect was the loss of association of KMT2D with this region and subsequent loss of H3K4me (Fig. 4A). The clustering of gene expression differences could either be due to H3K4me deposition at a shared enhancer or KMT2D targeting the region broadly. Our first question was thus to determine whether KMT2D associates with this region. Using two control lines from our study (the isogenic control line and the mother of the KS case), we found significantly increased binding of KMT2D at these sites compared to IgG control (Fig. 4B). This suggests that KMT2D is able to bind to 19q13.33 locus in human stem cells.

H3K4 methylation comes in three types, all of which can be deposited by KMT2D, and it was not immediately obvious which mark might be the best to test our hypothesis, as H3K4me1/2 mark active enhancers whereas H3K4me3 marks active promoters. We performed histone mass spectrometry analysis (22) in control iPSCs and found that H3K4me1 is the most abundant methylation mark genome wide in the iPSC state (Fig. 4C), reflecting about 12% of total H3K4me, while H3K4me2 was about 2% and H3K4me3 was $\sim$0.5–1% in stem cells. Notably $>$ 85% of H3K4 tails identified in mass spectrometry show no methylation. We opted to assay H3K4me1 because of its relative abundance and because of its relationship to gene activation outside of promoters, which we thought an unlikely model given the regional gene expression effects.

To directly test differential methylation at H3K4, we performed H3K4me1 ChIPseq analysis to assess the genome-wide differential level of H3K4me1 in the KS1 iPSCs versus iPSCs from her mother. We detected a lower number of total peaks in KS1 cells (146 099; 3.98 peaks per 1000 reads) compared with control cells (156 770 peaks; 4.4 peaks per 1000 reads), a decrease of about 10.5%. 19 950 peaks were lost or showed decreased levels in KS cells whereas 9279 peaks were gained. In other words, most H3K4me1 peaks in both KS and control iPSCs are not different, except at 29 229 peaks, where about two-thirds of these changes are in a direction expected by the loss of *KMT2D*. These data do not support a general decrease in H3K4me1 across the KS genome, which is what might be expected from the continued presence of KMT2C and some KMT2D,

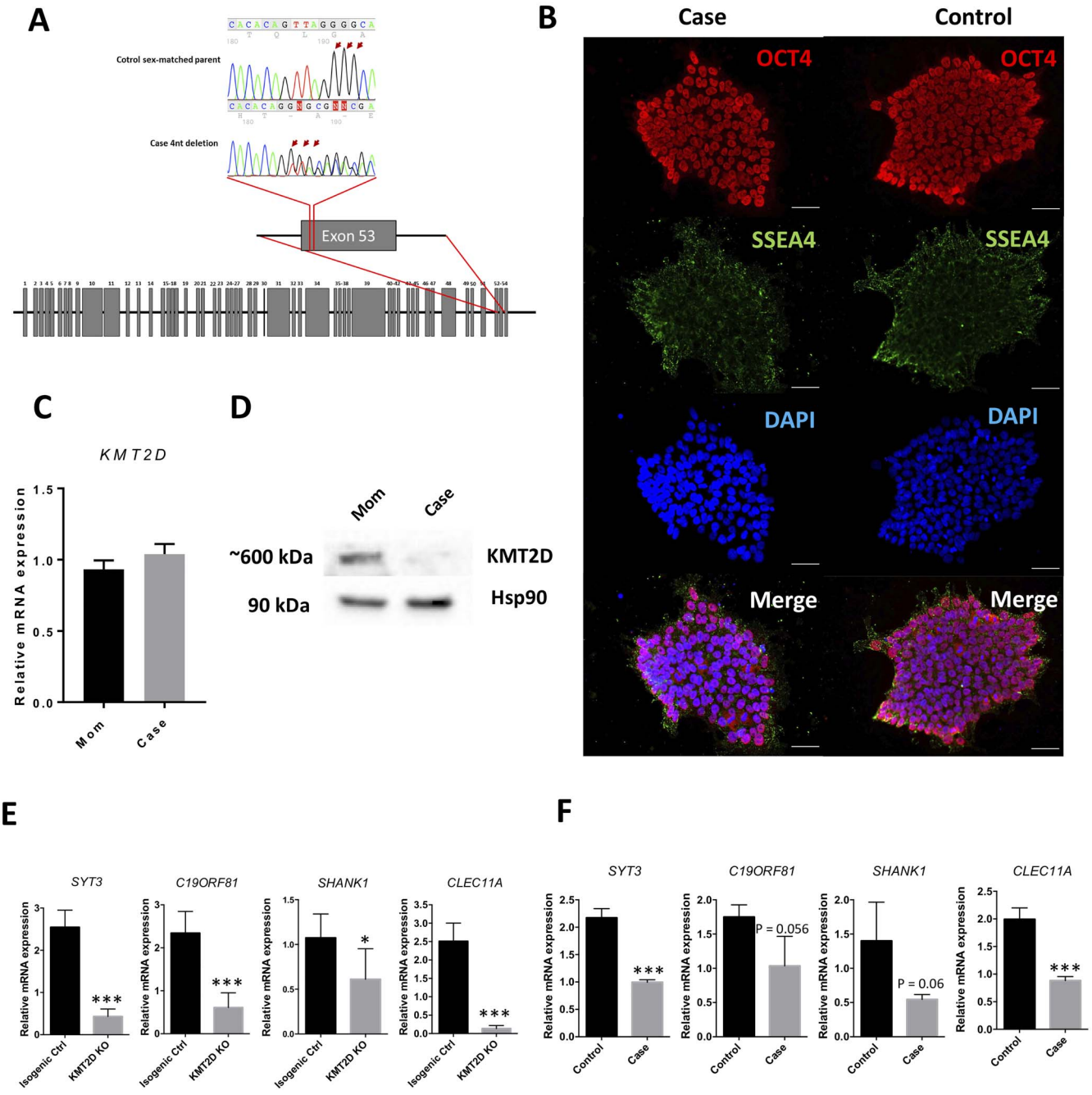

Figure 3. Generation of KS case-derived *KMT2D*-deficient iPSCs. (**A**) Illustration of the location of 4 bp deletion in exon 53 of KMT2D in the KS case cell. (**B**) Representative images of immunostaining of pluripotency markers expression in generated using the KS case and sex-matched family control cells. Scale bars indicate 50 μm. (**C**) qPCR analysis and validation of absence of non-sense mediated decay in *KMT2D* mRNA expression in KS case cells. There were two independent samples in the analysis (Mom and KS case). Each sample was extracted with three replicates for RNA samples, where replicates were the same iPSC line grown in separate wells and underwent independent RNA extraction, cDNA synthesis and qPCR analysis. (**D**) Western blotting result showing reduced expression of KMT2D expression in KS case cells. (**E**) qPCR analysis of *SYT3*, *SHANK1*, *CLEC11A* and *C19orf81* in engineered KMT2D KO iPSCs. There were four independent samples in the analysis (Control1, Control2, KMT2D$^{ex2\text{-}del}$ and KMT2D$^{ex2\text{-}ins}$). The two independent isogenic control lines in the control group were derived from the same fibroblast line used for the generation of the two KMT2D KO lines, reprogrammed and isolated from two different iPSC clones. The two KMT2D KO lines consist of KMT2D$^{ex2\text{-}del}$ and KMT2D$^{ex2\text{-}ins}$; each KO line were derived from different gene-edited iPSC colonies expanded from single cells. Each sample was extracted with 3 replicates for RNA samples, where replicates were the same iPSC line grown in separate wells and underwent independent RNA extraction, cDNA synthesis and qPCR analysis.***$P \leq 0.001$. (**F**) qPCR analysis of *SYT3*, *SHANK1*, *CLEC11A* and *C19orf81* in KS case iPSCs compared to the healthy sex-matched control family iPSCs. There were two independent samples in the analysis (Mom and KS case). Each sample was extracted with three replicates for RNA samples, where replicates were the same iPSC line grown in separate wells and underwent independent RNA extraction, cDNA synthesis and qPCR analysis. ***$P \leq 0.001$.

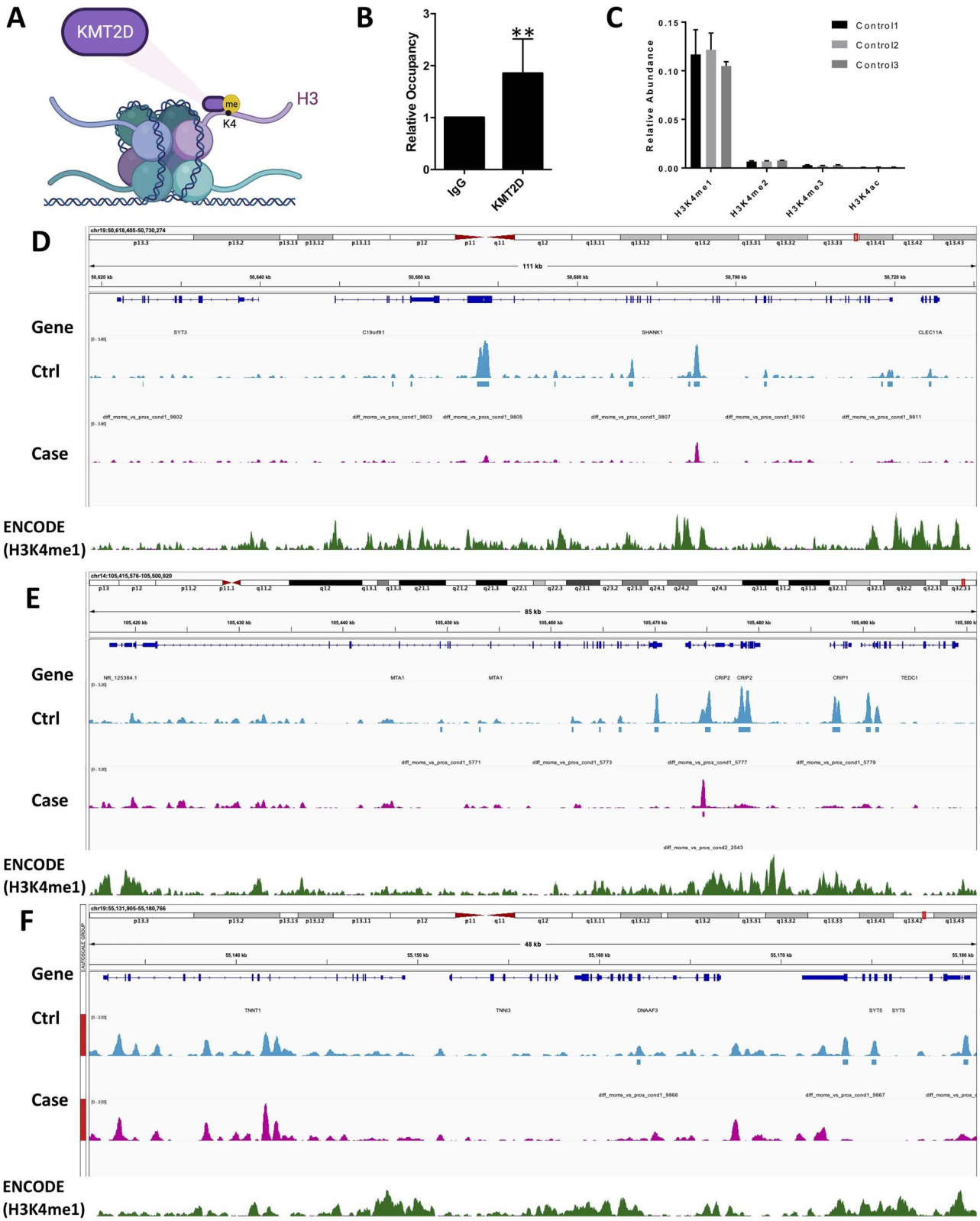

Figure 4. KMT2D deficiency leads to locus-specific reduction of H3K4me1 levels. (**A**) Illustration of the proposed mechanism of action on how KMT2D regulates the affected locus. (**B**) ChIP-PCR analysis of KMT2D binding in comparison to IgG control antibody with two different control iPSCs. There were two independent samples in the analysis (Mom and KS1). Each sample was extracted with three replicates for DNA samples, where replicates were the same iPSC line grown in separate dishes with different passage number, and underwent independent chromatin immunoprecipitation and PCR analysis. Each replicate signal was normalized to its own input samples. $^{*}P \leq 0.05$, $^{***}P \leq 0.001$. (**C**) Mass spectrometry analysis result showing different H3K4me1/me2/me3 level in control cells. There were three independent iPSC samples in the analysis (Control1, Control2 and Control3). Control1 and Control2 were reprogrammed from two different healthy control renal epithelial cells, and Control3 were reprogrammed from a healthy control fibroblast line. Each sample was extracted with three replicates, where replicates were the same iPSC line grown in separate dishes with different passage number, and underwent independent histone extraction and mass spectrometry analysis. (**D**) Integrative Genomics Viewer (IGV) map snapshot of ChIPseq peaks identified at the 19q13.33 locus

consistent with the redundancy of H3K4me deposition across the KMT2 family. Still, the data suggest that some genomic locations where H3K4me1 is found may require KMT2D for me1 deposition, and these could be the locations where we observed reduced mRNA expression in reduced dosage KMT2D models.

We conditioned analyses to the three regions that showed clear RNA differentials of adjacent genes at the same locus. We focused on the 19q13.33 locus because of the large RNA expression differences observed in that region. We found significant reduction in H3K4me1 level as compared with the control iPSCs (Fig. 4D) and these data appeared robust with decreases in H3K4me1 observed across the region. Expanding out across the region from the core four genes, we do not observe differential H3K4me1 in neighboring sites (Supplementary Material, Fig. S3). This suggests a degree of specificity to the ∼110 Kb region and that loss of *KMT2D* may lead to lower H3K4me1, possibly leading to lower expression of *SYT3, C19orf81, SHANK1* and *CLEC11A*. To ensure these data are not specific to any single cell line, we performed targeted H3K4me1 ChIP-PCR (Supplementary Material, Fig. S5) over the 19q13.33 locus in all KS models ($KMT2D^{ex2\text{-}del}$, $KMT2D^{ex2\text{-}ins}$, KS case iPSCs, $n=3$ in triplicate) compared with all controls (isogenic control to engineered lines and mother of KS case, $n=2$ in triplicate). We found significantly lower H3K4me1 in KS models compared with the control models (Supplementary Material, Fig. S5B and C). Finally, in Figure 4E and F we show peaks from the 19q13.42 and 14q32.33 regions which show a similar pattern as the 19q13.33 region: a broad pattern of peaks across the regions with depletion in KS iPSCs. H3K4me1 and H3K27me3 establish a bivalent state in stem cells on genes important in development (23). Focusing on the 19q13.33 locus, we mapped four marks from Embryonic stem cells (ES) in ENCODE—H3K4me1, H3K4me3, H3K27ac and H3K27me3—and compared these to our own iPSC H3K4me1 ChIPseq peaks to determine if significantly depleted peaks due potentially to KMT2D haploinsufficiency overlap with loci of known bivalency (i.e. H4K4me1 and H3K27me; Supplementary Material, Fig. S6). We found excellent overlap of our peak regions with both H3K4me1 and H3K27me3 in ES cells. This suggests that the losses of H3K4me1 in KS stem cells are in locations of bivalent enhancers, at least at this particular locus. This likely has important implications for disease mechanism given that mutations in the H3K27me demethylase, KDM6A, also cause KS.

Discussion

The role of individual KMT2D family members remains an open question, and most work to date has been done in non-human species or with complete knockout of more than one family member. KS, caused in most cases by loss-of-function of a single copy of *KMT2D*, is associated with a significant impairment of immune, bone and brain function, thus there is clearly an important role of more subtle effects of *KMT2D* loss-of-dosage in humans. The current work has explored this idea in human stem cells, in which *KMT2D* is highly expressed and where it might be needed to prime or prepare the genome for future cell type induction or differentiation (24). Our data suggest that KMT2D may establish H3K4me1 patterns in at least one specific chromosomal region in stem cells and affect transcription (25). This step could be important before stem cells are induced to different fates.

In our study, we generated $KMT2D^{+/-}$ lines from a KS patient and through CRISPR gene-editing. Despite the existence of a frameshifting mutation in *KMT2D*, there is no NMD event detected in any of the cell lines through qPCR and RNAseq analyses. Interestingly, we still observe reduced KMT2D protein expression without any detectable truncated protein being expressed despite the potential for NMD escape (26). NMD occurs when a premature termination signal happens upstream of an exon–junction complex. Premature stops attract proteins that then interact with both 5′ cap proteins and intragenic exon–junction proteins to facilitate mRNA degradation during pioneering transcriptional rounds (27). The mutations investigated in this report should theoretically fit this model, however, something specific to the KMT2D locus may allow for escape, whether the large number of other transcripts produced from the locus (ENSEMBL: ENSG00000167548), or a competitive change in NMD protein machinery specifically at this locus, as has been observed in some situations (28). We cannot rule out that mutant alleles are translating mutant protein and/or in-frame products from alternative start sites. If it is using an alternative start codon, there are two potential in-frame products of 572 or 593 KDa, which would likely be indistinguishable from the full length predicted product of 600 KDa since there is a lower gel resolution for larger product sizes in protein gel electrophoresis. The antibody epitope targets amino acids 2916–3785 (total length is 5537), so smaller out-of-frame fragments (which could be generated from the induced mutations in exon 2 and terminate within 100 bps) will not be detected. We expect that these are likely made but then degraded as they will be very short peptides. We conclude that the current models have reduced dosage of KMT2D protein and that the most likely effects on the cell will be caused by decreased KMT2D protein levels. Finally, KMT2D haploinsufficiency in KS cases has been previously thought to occur through an NMD mechanism

(hg38) in control and KS case iPSCs. There were two independent samples in the analysis (Mom and KS1). Each sample was extracted with three replicates for DNA samples, where replicates were the same iPSC line grown in separate dishes with different passage number, and underwent independent chromatin immunoprecipitation and PCR analysis. Each replicate signal was normalized to its own input samples. (**E**) IGV map snapshot of ChIPseq peaks identified at the 19q13.42 locus (hg38) in control and KS case iPSCs. (**F**) IGV map snapshot of ChIPseq peaks identified at the 14q32.33 locus (hg38) in control and KS case iPSCs.

(29), so this idea needs to be further explored in future studies.

KS genotype phenotype relationships are complex (30,31) and the results from this study should be interpreted within the context of heterozygous *KMT2D* loss-of-function mutations. To date, hundreds of KS causing mutations have been identified but the large size, multi-domain and potential transcriptional complexity (32) of the gene mean caution is warranted in interpretation of rare genetic variation and disease-causing effects, particularly in light of reports of missense mutations and the phenotypic spectrum of KS (33).

We performed whole genome transcriptome analysis to identify genes that may be affected by decreased dosage of *KMT2D*. In an attempt to identify direct targets of KMT2D, we focused on those genes whose expression had diminished, as it would allow us to test a model whereby these genes may require H3K4me to become active in stem cells. Our ranking of fold-change of these genes showed a surprising clustering of neighboring genes by chromosomal location and we focused on the most differentially expressed region identified which included *CLEC11A, SYT3, SHANK1* and *C19ORF81*; none of these have been previously defined as KMT2D-target genes. The specific change in this 110 kb region area may suggest specific effects in this area potentially due to it being a chromosomal domain (34). These domains are stretches of chromosome that can undergo region specific regulation due to boundary effects by CCCTC-binding factor (CTCF), or phase separation by DNA binding factors. The segregation of this region might allow for H3K4me1 spreading model (35) that could facilitate the expression of genes in the region.

We observed only a 10.5% decrease in H3K4me1 peaks per 1000 reads mapped in Kabuki patient iPSCs, arguing against a general model of genome-wide depletion of H3K4me1. Rather, these data are consistent with region specific targeting. We focused this work on one region where we found large and consistent effect sizes in diminished gene expression of neighboring genes in KMT2D reduced dosage models. Carefully done H3K4me1 ChIPseq following strict ENCODE analysis standards [as described in the Materials and Methods section (36)] suggested significant loss of peaks in a 19q13.33 region in Kabuki stem cells. Similar to RNAseq data where genes outside this region did not differ in their expression levels, there was no change in H3K4me1 as we extended the analysis outside of these areas, implying a relatively restricted deposition of H3K4me1 in healthy cells but which may require a full dose of KMT2D.

How might loss of KMT2D dosage lead to the gene expression effects observed? We propose that KMT2D is specifically targeted to these domains through transcription factor interaction and association with COHESIN and MEDIATOR complexes, as has been described (37). These effects could be specific to KMT2D and are non-redundant with other KMT2 family members. We suggest that the association of KMT2D with these domains allows for the spreading of H3K4me1 potentially through phase separation with KDM6A (16). Perhaps these specific regions require significant amounts of KMT2D to allow for the spreading of the H3K4me1 signal—smaller regions <50 Kb may not require such high levels and thus are spared from effects of *KMT2D* dosage loss. The heterozygous loss of *KMT2D* in stem cells at this developmental timepoint may affect all downstream developmental stages as this genomic region may need to be primed for future histone methylation deposition and gene activation and stage-specific actions (24). The genes at least in 19q13.33 region are potentially highly relevant to the Kabuki phenotype as haploinsufficiency of two of the genes—*CLEC11A* (or stem cell growth factor/osteolectin) and *SHANK1*—can by themselves cause significant problems in bone and brain development, respectively (38,39).

This work may provide a significant and testable model for why mutations in either *KDM6A* and *KMT2D* cause KS. KMT2D is an H3K4 methyltransferase, adding me1 to H3K4 presumably at many genomic sites, but at the least in the 19q13.33 region; we conjecture that KDM6A may remove H3K27me3 from the same poised enhancers targeted by KMT2D. The removal of the methyl group from H3K27 allows for the addition of the mutually exclusive acetylation mark, which finalizes the switch between a poised/bivalent state to an active enhancer. The haploinsufficiency of either *KMT2D* or *KDM6A* may lead to a loss of active enhancers in the genome. In the case of KMT2D, there may be a failure to fully establish the bivalent state, while loss of KDM6A may lead to the failure of fully poised enhancers to transition to an active state. This is directly testable in that we might predict that stem cells with *KDM6A* loss (which is X linked), will result in a specific gain of H3K27me3 at the chr 19q13.33 locus. This suggests that the significant loss of either enzyme could tip the delicate balance of the permissive and restrictive epigenetic regulation hence repressing the baseline expression of the affected genes in stem cells.

Materials and Methods

Somatic cell reprogramming

There are two kinds of somatic cells used in this study—fibroblast and renal epithelial cells. KMT2D$^{ex2\text{-}del}$, KMT2D$^{ex2\text{-}ins}$ and the isogenic control iPSCs were generated from the same fibroblast line whereas KS case and control iPSCs were generated from their respective renal epithelial cells, extracted from urine. All somatic cells were reprogrammed through the transfection of episomal reprogramming vectors encoding for *OCT4, SOX2, MYC3/4, KLF4* and shRNA *P53* (ALSTEM). Transfection was conducted with the neon electroporation system (Invitrogen, Burlington) using 2–3 $\times$ 10^5 cells and 3 μg of episomal vectors for every reaction. The transfected cells were plated on to Matrigel (Corning) coated tissue culture dish with 10% Fetal Bovine Serum (FBS) Dubelcco's Modified Eagle Medium (DMEM) to allow cell reattachment.

The culture media was then replaced with TesR-E7 (STEMCELL Technologies, Vancouver) after 48 h with daily media change afterwards. After 2 weeks in culture in TesR-E7, iPS colonies started to form into an observable size. Colonies with desirable size (500–1000 μm in diameter) were manually picked after a brief treatment with ReLeSR media and replated into new Matrigel-coated dish and cultured in TesR-E8 media. Throughout this study, no comparisons were done that mixed cases or control iPSCs from different tissues of origin.

Quality control of induced pluripotent stem cells

Pluripotency marker expression

iPSC colonies were re-plated onto matrigel-coated glass cover slips in a petri dish or suspension culture dish. Cells were cultured for at least 24 h to allow reattachment. iPSC colonies were fixed with 4% paraformaldehyde (Sigma-Aldrich) for 10–15 min, followed with downstream procedure described in immunofluorescence section.

Endogenous marker expression

Total RNA extraction and cDNA synthesis were conducted as it is described in quantitative polymerase chain reaction section. Endogenous pluripotent markers expression was amplified using primers listed in the Supplementary Material, Table S1.

Quantitative polymerase chain reaction

Total RNA extraction was conducted by the RNeasy mini kit (Qiagen, Hilden). Afterwards, reverse transcription was conducted using 1 μg of extracted RNA samples, 5 μM random primers, 0.5 mM dNTPs, 0.01 M Dithiothreitol (DTT) and 400 U Moloney Murine Leukemia (M-MLV) Reverse Transcription (RT) (Carlsbad, California). The qPCR reaction was conducted in 384 well plates using a Quant Studio 6 Flex Real time PCR machine (Life Technology). A reference pool of cDNA samples was used to generate a reference sample for relative mRNA expression quantification. Each well included 5 μl of 2X gene expression master mix (Luna Universal qPCR Master Mix, New England BioLabs), 1 μl of 5 μM primer mix, 1 ng of cDNA and nuclease free water to make up to 10 μl of total mix volume. Beta-actin was used as an internal control for normalization unless specified otherwise.

Immunoblotting

After reaching the desired confluency, cells were lysed using Radioimmunoprecipitation assay (RIPA) buffer (Sigma-Aldrich) supplemented with cOmplete, Ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor cocktail tablets (Sigma-Aldrich). Cell lysates were well mixed for 10 min and centrifuged with 15 000 rpm in low temperature. Protein samples were then isolated from the supernatant. Protein concentrations were measured using Pierce BCA protein assay kit (ThermoFisher). Protein samples were then run in Mini-PROTEAN TrisGlycine eXtended (PROTEAN TGX) stain-free precast gels (Bio-Rad) with parameters: 15 μg protein/well, 150 V for 30–35 min. Resulting gel were then transferred to nitrocellulose membrane using Trans-Blot Turbo Transfer system (Bio-Rad). Afterwards, the membranes were immersed in blocking buffer (4% non-fat milk dissolved in Tris Buffered Saline with Tween 20 (TBST) buffer) for 30 min followed by overnight incubation with gentle shaking in primary antibody solution (specified primary antibody dilution in blocking buffer). Blots were then washed three times with TBST buffer before the following incubation in secondary antibody buffer for 60 min at room temperature. Blots were washed three times before visualizing the band signal using Clarity western Enhanced Chemiluminescence (ECL) blotting substrates (Bio-Rad). The blot imaging and analysis were done using Bio-Rad ChemiDoc imaging system and ImageLab software (Bio-Rad). All signals were normalized to β-actin or HSP90. All the antibodies used are listed in the Supplementary Material, Table S1.

Immunostaining

Cells were plated onto glass coverslips in a Petri dish or suspension dish. After reaching optimal imaging density, cells were washed with phosphate-buffered saline (PBS), then fixed with 4% paraformaldehyde (Sigma-Aldrich) for 10–15 min. Cells were then permeabilized using 0.25% TritonX-100 (Sigma-Aldrich) in 0.5% PBS-Bovine Serum Albumin (BSA) for 10–15 min followed by blocking process in 10% PBS-BSA for 60 min. Afterwards, samples were incubated in primary antibody solution (appropriate antibody concentration in 5% PBS-BSA) for 60 min in room temperature. Samples were then washed with 1×PBS for three times followed by secondary antibody dilution for 60 min in the dark. Finally, samples were washed twice with 1×PBS and visualized on an Apotome fluorescent microscope (Zeiss). Images were analyzed using ImageJ. All the antibodies used are listed in the Supplementary Material, Table S1.

RNA sequencing

RNA samples with RNA Integrity Number (RIN) values > 9 were submitted to Genome Quebec for RNA sequencing. Each iPSC line (two replicates for each of two controls and two experimental groups for a total of eight sequenced samples) were cultured in TesR-E8 media and actively proliferating at the time of RNA extraction. PolyA enrichment library preparation for coding transcripts was done by experts at the Genome Quebec Innovation Center in McGill University. Total RNA was quantified using a NanoDrop Spectrophotometer ND-1000 (NanoDrop Technologies, Inc.) and its integrity was assessed using a 2100 Bioanalyzer (Agilent Technologies). Libraries were generated from 250 ng of total RNA as following: mRNA enrichment was performed using the NEBNext Poly(A) Magnetic Isolation Module (New England BioLabs). cDNA synthesis was achieved with the NEBNext RNA First Strand Synthesis and NEBNext Ultra Directional RNA Second Strand Synthesis Modules (New

England BioLabs). The remaining steps of library preparation were completed by the NEBNext Ultra II DNA library prep kit for Illumina (New England BioLabs). Adapters and PCR primers were purchased from New England BioLabs. Libraries were quantified using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies) and the Kapa Illumina GA with Revised Primers-SYBR Fast Universal kit (Kapa Biosystems). Average size fragment was determined using a LabChip GX (PerkinElmer) instrument. The libraries were normalized and pooled at 3 nM and then denatured in 0.05 N NaOH and neutralized using HT1 buffer. Exclusion-Amplification (ExAMP) was added to the mix following the manufacturer's instructions. The pool now at 200 pM was loaded on a Illumina cBot and the flowcell was ran on a HiSeq 4000 for 2 × 100 cycles (paired-end mode). A phiX library was used as a control and mixed with libraries at 1.5% level. The Illumina control software was HiSeq Control Software (HCS) HD 3.4.0.38, the real-time analysis program was Real Time Analysis (RTA) v. 2.7.7. Program bcl2fastq2 v2.20 was then used to demultiplex samples and generate fastq reads. Libraries were run using Illumina HiSeq4000 PE100, which produced 40 million reads per library. Our downstream bioinformatics processing used FASTX-Toolkit, TopHat, Bowtie2 and Cufflinks2 with default parameters to preprocess, align and assemble reads into transcripts, estimate abundance and test differential expression (40,41).

Chromatin immunoprecipitation assays

Cells (three replicates for each of KS1 and her mother, where replicates are cells from different wells and processed independently) were fixed with crosslinking reagent (11% formaldehyde, 100 mM NaCl, 0.5 mM ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid (EGTA), 50 mM HEPES ph 8) with 1:10 ratio of fixing reagent to culture medium for 15 min at room temperature. 0.125 M Glycine with 1:10 ratio of Glycine to cell medium was then added for 5 min at room temperature. Cells were then washed twice with PBS, then collected in cold PBS and centrifuged. Cells were resuspended in 1 ml Nucleus/Chromatin Preparation (NCP) Buffer I (EDTA 10 mM, EGTA 0.5 mM, Hepes 10 mM, TritonX-100 0.25%) and then centrifuged. Cell pellets were then resuspended in 1 ml NCP Buffer II (EDTA 1 mM, EGTA 0.5 mM, Hepes 10 mM, NaCl 200 mM) and then centrifuged. Then, cell pellets were resuspended in 500–800 μl of RIPA Buffer (150 mM NaCl, 1% (v/v) NP-40, 0.5% (w/v) deoxycholate, 0.1% (w/v) Sodium Dodecyl Sulfate (SDS), 50 mM Tris–HCL ph 8.0, 0.5 mM EDTA). Afterwards, samples were sonicated up to 500–1000 bp fragment size then centrifuged at high speed at 4°C. 1 mg of the resulting protein samples were resuspended in 800 ml of RIPA buffer then used for immunoprecipitation, while the rest were kept as input control samples. The samples were then pre-cleared for 2–3 h at 4°C with 30 μl of protein A/G beads. The non-specific complexes were centrifuged at high speed for 30 s at 4°C. The supernatants were extracted and incubated with 2 μg of specific antibody overnight at 4°C.

After the incubation, 50 μl of protein A/G beads were added followed by 1 h incubation at 4°C. After removing the supernatants, the precipitates were washed 6 times with 500 μl RIPA Buffer. The resulting immunocomplexes were eluted using 1% SDS incubation for 10 min at 65°C. 200 mM of NaCl and 10 mg of RNAseA were then added to the extract (and input samples) and incubated for 5 h or overnight at 65°C. Afterwards, 500 μg/ml proteinase K were added, incubated at 45°C with for 2 h. Chromatin fragments were extracted using 300 μl phenol/chloroform/isoamylalcohol and again with 300 μl chloroform. The fragments were then precipitated with 2.5 volume of Ethyl alcohol (EtOH) overnight incubation at −20°C with 10 μg tRNA or with glycogen carrier. Afterwards, the samples were washed with 70% EtOH and finally resuspended in 50 μl of Tris-EDTA (TE) buffer.

In ChIP-PCR analyses, DNA fragments were analyzed by quantitative polymerase chain reaction with SYBR Green detection. DNA fragments from the IgG (negative control) experiment and KMT2D antibody-bound DNA fragments were normalized to input DNA, then we calculated the fold enrichment of KMT2D fragments to IgG fragments. All primers used are listed in the Supplementary Material, Table S1.

ChIP-sequencing

Raw sequencing data in fastq format were processed using the nf-core (42) ChIPseq (version 1.2.1) processing template (doi: 10.5281/zenodo.3966161) for Quality control (QC), adapter trimming and read alignment using the burrows-wheeler algorithm against the human genome (version GRCh38 included with nf-core). Following read alignment, Binary Alignment Map (BAM) files for controls (and, in a separate file, BAM files for Kabuki patients) were merged, using samtools (43). Peak calling and analysis of differential H3K4me1 binding was conducted using MACS2 (44). For this, the MACS2 callpeak function was used with parameters –f BAMPE, –g 2.7e9, –broad, –broad-cutoff 0.1,–nomodel, –extsize 147, without the use of an input control. Next, the MACS2 bdgdiff function was used with parameters –g 60 and –l 120, to extract differences between control and Kabuki samples. Lastly, annotation of differential H3Kme1 binding was conducted using the annotatePeaks.pl function included in the nf-core ChIPseq (v1.2.1) repository. Further analyses of Kyoto Encyclopedia of Genes and Genomes (KEGG)-pathway and Gene Ontology (GO)-term enrichment were conducted in R (www.r-project.org) using the package clusterProfiler (45).

Histone extraction and derivatization

Nuclei were extracted from cell pellets using nuclei isolation buffer (NIB) as previously described with minor adjustments (46). First, cell pellets lysed using ten cell pellet volumes of NIB +0.3% NP-40 Alternative. The cells were then washed twice with NIB only to remove

any detergent. Next, histones were acid extracted from nuclei with 0.2 M H2SO4, followed by precipitation with 33% trichloroacetic acid overnight at 4°C. Precipitated histones were then washed first with ice-cold acetone-HCl (1%) and then with ice-cold acetone. About 20 μg of histones were then derivatized with propionic anhydride to propionylate unmodified lysines, followed by trypsin digestion overnight (1:20, enzyme: histone). The histone peptides were then propionylated at the N-termini by another round of derivatization. Finally, the samples were desalted using in-house prepared C18 stage-tips before nanoLC-MS/MS analysis.

Histone mass spectrometry analysis

Histone samples were analyzed by nanoLC-MS/MS with a Dionex-nanoLC coupled to a Q Exactive-HF mass spectrometer (Thermo Fisher Scientific). The column was packed in-house using reverse-phase 75 μm ID × 17 cm Reprosil-Pur C18-AQ (3 μm; Dr Maisch GmbH). The High Performance Liquid Chromatography (HPLC) gradient was as follows: 5–40% solvent B (A = 0.1% formic acid; B = 80% acetonitrile, 0.1% formic acid) over 47 min, from 40 to 90% solvent B in 5 min, 90% B for 8 min. The flow rate was at 300 nl/min. Data were acquired using a data-independent acquisition method, consisting of a full scan Mass Spectrometry (MS) spectrum (m/z 300–1100) performed in the Orbitrap at 35 000 resolution with an Automatic Gain Control (AGC) target value of 2e5, followed by 16 MS/MS windows of 50 m/z using Higher energy Collisional Dissociation (HCD) fragmentation and detection in the ion trap. The HCD collision energy was set to 28, AGC target at 1e4, and maximum inject time at 50 ms. Histone samples were resuspended in buffer A, and 1 μg of total histones was injected.

An in-house software EpiProfileLite (GitHub at https://github.com/zfyuan/EpiProfileLite, includes the user guide) was used to analyze the raw files to obtain relative ratios of all modified and unmodified forms of H3 and H4 histone peptide. This software efficiently discriminated isobaric peptides using unique fragment ions in the MS/MS scans. In all, we report about 29 peptide sequences with 45 Post-translational modifications (PTMs) (methylations, acetylations and phosphorylations) for a total of 151 histone marks plus 16 unmodified histone peptides. The raw files are deposited in CHORUS (www.chorusproject.org) with project #1758.

Statistical analyses

Error bars in figures represent the standard error of the mean. The t-tests were based on two-tailed student t-tests. Statistical analyses were conducted using Graphpad Prism 6. Statistical output and n are reported at all places where data are reported.

Supplementary Material

Supplementary Material is available at *HMG* online.

Conflict of Interest statement. A.S. declared employment and research funding from Bayer AG for research on computational methods in biomedicine. The other authors declared no potential conflicts of interest.

Funding

This work was funded by a Canadian Institutes of Health Research (CIHR) project grant to Carl Ernst.

Authors' Contributions

M.J., C.E.: conceived the study, wrote the manuscript and were involved in all aspects of the design and execution of the study; X.Z., L.Z., H.P., N.H., S.D., L.A., A.N., S.A., Y.Z., N.J., H.T.B.: generated primary cell-culture data; K.W., B.A.G.: conducted the mass spectrometry experiment for post-translational modification analysis. P.S.S., J.-F.T., Z.A., A.S.: performed bioinformatics analyses. All authors read and approved the final manuscript.

References

1. Niikawa, N., Kuroki, Y., Kajii, T., Matsuura, N., Ishikiriyama, S., Tonoki, H., Ishikawa, N., Yamada, Y., Fujita, M., Umemoto, H. *et al.* (1988) Kabuki make-up (Niikawa-Kuroki) syndrome: a study of 62 patients. *Am. J. Med. Genet.*, **31**, 565–589.
2. Ng, S.B., Bigham, A.W., Buckingham, K.J., Hannibal, M.C., McMillin, M.J., Gildersleeve, H.I., Beck, A.E., Tabor, H.K., Cooper, G.M., Mefford, H.C. *et al.* (2010) Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat. Genet.*, **42**, 790–793.
3. Miyake, N., Koshimizu, E., Okamoto, N., Mizuno, S., Ogata, T., Nagai, T., Kosho, T., Ohashi, H., Kato, M., Sasaki, G. *et al.* (2013) MLL2 and KDM6A mutations in patients with Kabuki syndrome. *Am. J. Med. Genet. A*, **161A**, 2234–2243.
4. Froimchuk, E., Jang, Y. and Ge, K. (2017) Histone H3 lysine 4 methyltransferase KMT2D. *Gene*, **627**, 337–342.
5. Wang, C., Lee, J.E., Lai, B., Macfarlan, T.S., Xu, S., Zhuang, L., Liu, C., Peng, W. and Ge, K. (2016) Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. *Proc. Natl. Acad. Sci. U. S. A.*, **113**, 11871–11876.
6. Miller, T., Krogan, N.J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J.F. and Shilatifard, A. (2001) COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc. Natl. Acad. Sci. U. S. A.*, **98**, 12902–12907.
7. Shilatifard, A. (2012) The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu. Rev. Biochem.*, **81**, 65–95.
8. Li, Y., Han, J., Zhang, Y., Cao, F., Liu, Z., Li, S., Wu, J., Hu, C., Wang, Y., Shuai, J. *et al.* (2016) Structural basis for activity regulation of MLL family methyltransferases. *Nature*, **530**, 447–452.
9. Schuettengruber, B., Martinez, A.M., Iovino, N. and Cavalli, G. (2011) Trithorax group proteins: switching genes on and keeping them active. *Nat. Rev. Mol. Cell Biol.*, **12**, 799–814.
10. Rao, R.C. and Dou, Y. (2015) Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. *Nat. Rev. Cancer*, **15**, 334–346.

11. Lee, J., Kim, D.H., Lee, S., Yang, Q.H., Lee, D.K., Lee, S.K., Roeder, R.G. and Lee, J.W. (2009) A tumor suppressive coactivator complex of p53 containing ASC-2 and histone H3-lysine-4 methyltransferase MLL3 or its paralogue MLL4. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 8513–8518.
12. Sze, C.C. and Shilatifard, A. (2016) MLL3/MLL4/COMPASS family on epigenetic regulation of enhancer function and cancer. *Cold Spring Harb. Perspect. Med.*, **6**, a026427.
13. Lee, J.E., Wang, C., Xu, S., Cho, Y.W., Wang, L., Feng, X., Baldridge, A., Sartorelli, V., Zhuang, L., Peng, W. and Ge, K. (2013) H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. *elife*, **2**, e01503.
14. Hu, D., Gao, X., Morgan, M.A., Herz, H.M., Smith, E.R. and Shilatifard, A. (2013) The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Mol. Cell. Biol.*, **33**, 4745–4754.
15. Demers, C., Chaturvedi, C.P., Ranish, J.A., Juban, G., Lai, P., Morle, F., Aebersold, R., Dilworth, F.J., Groudine, M. and Brand, M. (2007) Activator-mediated recruitment of the MLL2 methyltransferase complex to the beta-globin locus. *Mol. Cell*, **27**, 573–584.
16. Shi, B., Li, W., Song, Y., Wang, Z., Ju, R., Ulman, A., Hu, J., Palomba, F., Zhao, Y., le, J.P. *et al.* (2021) UTX condensation underlies its tumour-suppressive activity. *Nature*, **597**, 726–731.
17. Bell, S., Peng, H., Crapper, L., Kolobova, I., Maussion, G., Vasuta, C., Yerko, V., Wong, T.P. and Ernst, C. (2017) A rapid pipeline to model rare neurodevelopmental disorders with simultaneous CRISPR/Cas9 gene editing. *Stem Cells Transl. Med.*, **6**, 886–896.
18. Holbrook, J.A., Neu-Yilik, G., Hentze, M.W. and Kulozik, A.E. (2004) Nonsense-mediated decay approaches the clinic. *Nat. Genet.*, **36**, 801–808.
19. Caron, H., van Schaik, B., van der Mee, M., Baas, F., Riggins, G., van Sluis, P., Hermus, M.C., van Asperen, R., Boon, K., Voûte, P.A. *et al.* (2001) The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science*, **291**, 1289–1292.
20. Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L.I. *et al.* (2010) Epigenetic memory in induced pluripotent stem cells. *Nature*, **467**, 285–290.
21. Carosso, G.A., Boukas, L., Augustin, J.J., Nguyen, H.N., Winer, B.L., Cannon, G.H., Robertson, J.D., Zhang, L., Hansen, K.D., Goff, L.A. and Bjornsson, H.T. (2019) Precocious neuronal differentiation and disrupted oxygen responses in Kabuki syndrome. *JCI Insight*, **4**, e129375. https://doi.org/10.1172/jci.insight.129375.
22. Karch, K.R., Sidoli, S. and Garcia, B.A. (2016) Identification and quantification of histone PTMs using high-resolution mass spectrometry. *Methods Enzymol.*, **574**, 3–29.
23. Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K. *et al.* (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, **125**, 315–326.
24. Ernst, C. and Jefri, M. (2021) Epigenetic priming in neurodevelopmental disorders. *Trends Mol. Med.*, **27**, 1106–1114.
25. Baker, C.L., Walker, M., Arat, S., Ananda, G., Petkova, P., Powers, N.R., Tian, H., Spruce, C., Ji, B., Rausch, D. *et al.* (2019) Tissue-specific trans regulation of the mouse epigenome. *Genetics*, **211**, 831–845.
26. Dyle, M.C., Kolakada, D., Cortazar, M.A. and Jagannathan, S. (2020) How to get away with nonsense: mechanisms and consequences of escape from nonsense-mediated RNA decay. *Wiley Interdiscip Rev RNA*, **11**, e1560.
27. Schoenberg, D.R. and Maquat, L.E. (2012) Regulation of cytoplasmic mRNA decay. *Nat. Rev. Genet.*, **13**, 246–259.
28. Huang, L., Lou, C.H., Chan, W., Shum, E.Y., Shao, A., Stone, E., Karam, R., Song, H.W. and Wilkinson, M.F. (2011) RNA homeostasis governed by cell type-specific and branched feedback loops acting on NMD. *Mol. Cell*, **43**, 950–961.
29. Micale, L., Augello, B., Maffeo, C., Selicorni, A., Zucchetti, F., Fusco, C., De Nittis, P., Pellico, M.T., Mandriani, B., Fischetto, R. *et al.* (2014) Molecular analysis, pathogenic mechanisms, and readthrough therapy on a large cohort of Kabuki syndrome patients. *Hum. Mutat.*, **35**, 841–850.
30. Baldridge, D., Spillmann, R.C., Wegner, D.J., Wambach, J.A., White, F.V., Sisco, K., Toler, T.L., Dickson, P.I., Cole, F.S., Shashi, V. and Grange, D.K. (2020) Phenotypic expansion of KMT2D-related disorder: beyond Kabuki syndrome. *Am. J. Med. Genet. A*, **182**, 1053–1065.
31. Daly, T., Roberts, A., Yang, E., Mochida, G.H. and Bodamer, O. (2020) Holoprosencephaly in Kabuki syndrome. *Am. J. Med. Genet. A*, **182**, 441–445.
32. Faundes, V., Malone, G., Newman, W.G. and Banka, S. (2019) A comparative analysis of KMT2D missense variants in Kabuki syndrome, cancers and the general population. *J. Hum. Genet.*, **64**, 161–170.
33. Cuvertino, S., Hartill, V., Colyer, A., Garner, T., Nair, N., Al-Gazali, L., Canham, N., Faundes, V., Flinter, F., Hertecant, J. *et al.* (2020) A restricted spectrum of missense KMT2D variants cause a multiple malformations disorder distinct from Kabuki syndrome. *Genet. Med.*, **22**, 867–877.
34. Sexton, T. and Cavalli, G. (2015) The role of chromosome domains in shaping the functional genome. *Cell*, **160**, 1049–1059.
35. Wang, L.H., Aberin, M.A.E., Wu, S. and Wang, S.P. (2021) The MLL3/4 H3K4 methyltransferase complex in establishing an active enhancer landscape. *Biochem. Soc. Trans.*, **49**, 1041–1054.
36. Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P. *et al.* (2012) ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.*, **22**, 1813–1831.
37. Wang, S.P., Tang, Z., Chen, C.W., Shimada, M., Koche, R.P., Wang, L.H., Nakadai, T., Chramiec, A., Krivtsov, A.V., Armstrong, S.A. and Roeder, R.G. (2017) A UTX-MLL4-p300 transcriptional regulatory network coordinately shapes active enhancer landscapes for eliciting transcription. *Mol. Cell*, **67**, 308–321.e6.
38. Sato, D., Lionel, A.C., Leblond, C.S., Prasad, A., Pinto, D., Walker, S., O'Connor, I., Russell, C., Drmic, I.E., Hamdan, F.F. *et al.* (2012) SHANK1 deletions in males with autism spectrum disorder. *Am. J. Hum. Genet.*, **90**, 879–887.
39. Yue, R., Shen, B. and Morrison, S.J. (2016) Clec11a/osteolectin is an osteogenic growth factor that promotes the maintenance of the adult skeleton. *Elife*, **5**, e18782. https://doi.org/10.7554/eLife.18782.
40. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L. and Pachter, L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.*, **7**, 562–578.
41. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
42. Ewels, P.A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M.U., Di Tommaso, P. and Nahnsen, S. (2020) The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.*, **38**, 276–278.
43. Li H, Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

44. Feng, J., Liu, T., Qin, B., Zhang, Y. and Liu, X.S. (2012) Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.*, **7**, 1728–1740.
45. Yu, G., Wang, L.G., Han, Y. and He, Q.Y. (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*, **16**, 284–287.
46. Sidoli, S., Bhanu, N.V., Karch, K.R., Wang, X. and Garcia, B.A. (2016) Complete workflow for analysis of histone post-translational modifications using bottom-up mass spectrometry: from histone extraction to data analysis. *J. Vis. Exp.*, **111**, 54112. https://doi.org/10.3791/54112.