Supplemental methodology

ATAC-seq procedure

ATAC-seq was performed with Omni-ATAC protocol for three KDM4C:p.H217R-carrier and four control lymphoblastoid cell lines. All used cell lines had low and similar estimated passage numbers (supplemental table 1). 50 000 nuclei were used in a transposition reaction, and ATAC was simultaneously performed in triplicate for each sample. DNA fragments were pooled together after transposition and purified with Zymo DNA clean and Concentrator-5 kit (Nordic BioSite, Cat. No. D4014). Libraries were amplified by PCR and quantified by qPCR (KAPA library quant. kit, Illumina, Cat. No. KK4824). Quality was measured with TapeStation 2200 (Agilent Technologies). ATAC libraries were sequenced at Macrogen with the HiSeq 4000 system (Illumina, San Diego, CA) and 100 bp paired-end sequencing. 80 million paired-end reads were achieved for each sample.

Supplemental Table 1. Estimated passage numbers of used cell lines.

Cell line	**Estimated passage**
Case III:1 - H217R carrier	8
Case III:3 - H217R carrier	6
Case IV:1 - H217R carrier	5
Control1 - non-carrier	8
Control2 - non-carrier	8
Control3 - non-carrier	10
Control4 - non-carrier	10

Raw sequencing reads were trimmed of sequencing adapters and bad quality bases with cutadapt version 1.16 of Trim Galore version 0.4.5. We used default parameters except for -q,

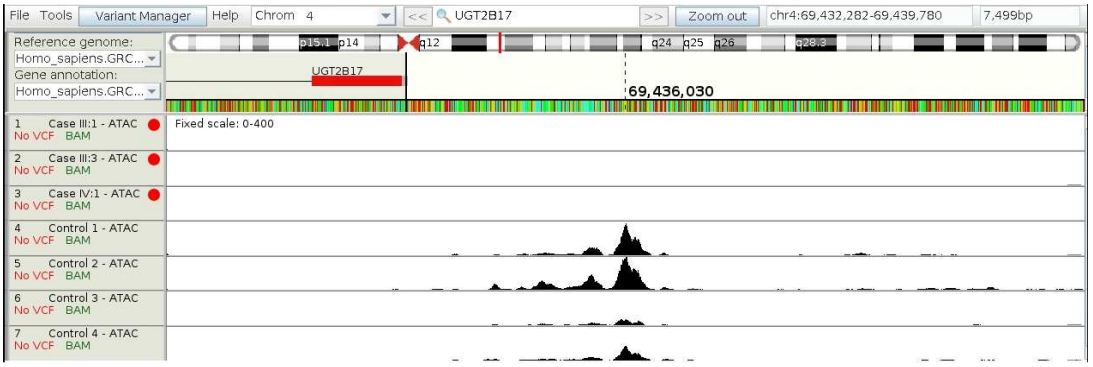
Katainen R, *et al. J Med Genet* 2022; 59:644–651. doi: 10.1136/jmedgenet-2021-107747

which was 30. Bowtie 2 version 2.1.0 was used to map trimmed reads to the reference genome hs37d5 (see supplemental figure 1 below for representative locus of the ATAC-seq data tracks). Samtools version 1.8 was used to remove PCR duplicates and reads with mapping quality below 20, and to count reads mapping to the mitochondrial genome. Fixed-width peaks were called from reads mapping to autosomes and the X chromosome with macs2 version 2.1.2 callpeaks command and parameters "--shift 75 --extsize 150 --nomodel --call-summits --nolambda --keep-dup all -p 0.01". Peak summits were extended 250 bp in both directions, and ENCODE blacklisted genomic regions [1] were removed.

Merged reproducible peak sets for samples were generated as described in [2], resulting in non-overlapping peaks present in at least two samples with peak scores ≥ 5. To calculate Tn5 insertions indicating chromatin accessibility, Tn5 insertion sites in paired-end read alignments in all samples were corrected to Tn5 offset ("+" stranded +4 bp, "-" stranded -5 bp) [3], and insertion sites in merged reproducible peaks were counted.

Differential accessibility between KDM4C:p.H217R carriers and controls was calculated with DESeq2 version 1.26.0 in R version 3.6.1 with default settings and Wald statistics. Differentially accessible regions (DARs) were required to have a false discovery rate (FDR) ≤ 0.05. Annotation of DARs was performed with HOMER annotatePeaks.pl against hg19 v6.4 RefSeq database [4]. The Locus Overlap Analysis (LOLA) R package [5] version 1.16.0 was used to detect enrichment of histone modification loci separately for more and less accessible regions over the merged reproducible peak set that was used as the background set of regions. More specifically, the Roadmap Epigenomics Project [6] histone modification data of cells with anatomy=blood (supplemental table 2) were used as a reference Database,

more or less accessible regions as UserSet, and all merged reproducible peaks as UserUniverse. Results were visualized with ggplot2 version 3.2.1 in R version 3.6.1.



Supplemental Figure 1. Representative locus of ATAC-seq raw read data tracks.

Supplemental Table 2

E029	Primary monocytes from primary blood
E030	Primary neutrophils from primary blood
E031	Primary B cells from cord blood
E032	Primary B cells from primary blood
E033	Primary T cells from cord blood
E034	Primary T cells from primary blood
E035	Primary haematopoietic stem cells (HSCs)
E036	Primary HSCs short term culture
E037	Primary T helper memory cells from primary blood
E038	Primary T helper naive cells from primary blood
E039	Primary T helper naive cells from primary blood
E040	Primary T helper memory cells from primary blood

E041	Primary T helper cells PMA-I stimulated
E042	Primary T helper 17 cells PMA-I stimulated
E043	Primary T helper cells from primary blood
E044	Primary T regulatory cells from primary blood
E045	Primary T cells effector/memory enriched (primary blood)
E046	Primary natural killer cells from primary blood
E047	Primary T CD8+ naive cells from primary blood
E048	Primary T CD8+ memory cells from primary blood
E050	Primary HSCs G-CSF-mobilized female
E051	Primary HSCs G-CSF-mobilized male
E062	Primary mononuclear cells from primary blood
E115	Dnd41 T cell leukaemia
E116	GM12878 lymphoblastoid
E123	K562 leukaemia
E124	Monocytes-CD14+ RO01746

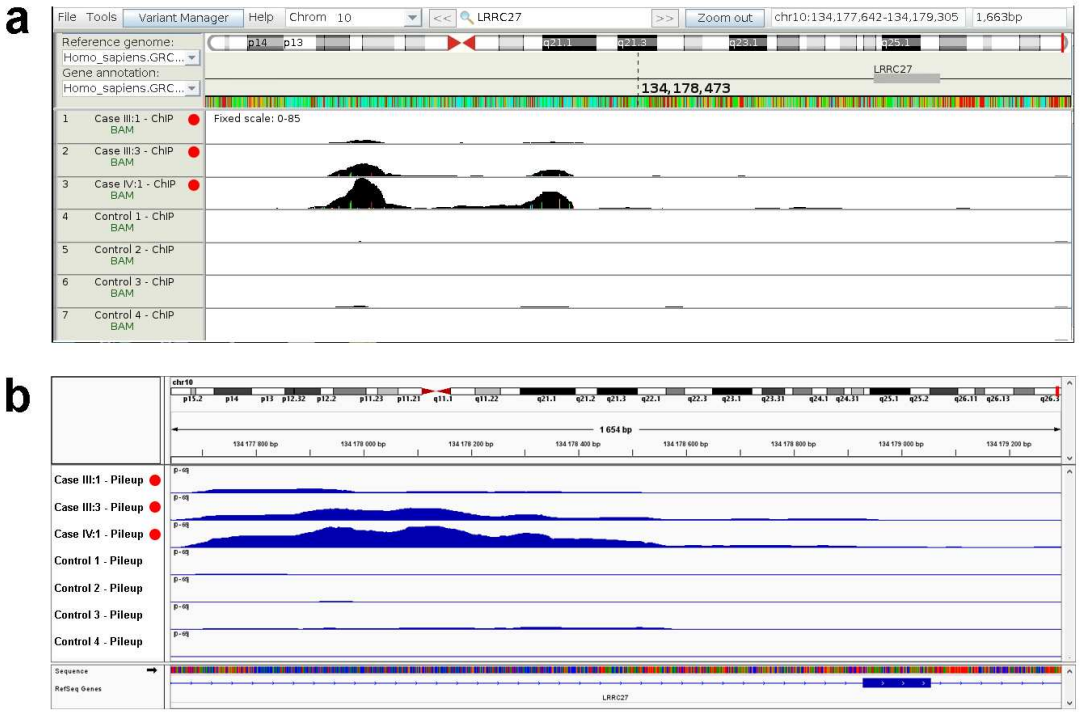
ChIP-seq procedure

10 million cells were crosslinked with 1% formaldehyde for 10 minutes at room temperature, and the reaction stopped by adding 0.125 M glycine. Crosslinked cells were washed three times with ice-cold PBS, and then resuspended in ice-cold lysis buffer (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP-40, 1X protease inhibitor cOmplete). The cells were transferred into Dounce tissue grinder and thoroughly homogenized with pestle B, and then resuspended in 800 µl RIPA buffer (1% NP‑40, 0.5% sodium deoxycholate, 0.1% SDS, 1mM EDTA,

140mM NaCl, 10mM Tris-HCl pH 8, 1X protease inhibitor cOmplete). The DNA was sonicated to an average fragment length of 200-500 bp using Sonicator Misonix S-4000 with the following parameters: amplitude 33%, pulse-on time 30", pulse-off time 60", and total sonication time 10 minutes. The samples were then centrifuged 18 000 g for 20 minutes at 4°C, and the supernatants collected. 50 µl of Dynabeads protein-G (Thermo Fisher Scientific, cat 10003D) were washed with 0.05% Tween-20 in PBS, and 7 µg of antibody against H3K9me3 (Active Motif 39062, lot 09919003) was incubated with the beads for 10 minutes at room temperature with rotation. 50 µl of chromatin was removed for input control, and 50 µl for fragmentation control and quantification. 25 µg of chromatin was incubated with antibody-coated beads overnight at 4°C with rotation. Samples were washed twice with low salt washing buffer (0.1% SDS, 1% Triton X 100, 2 mM EDTA, 20 mM Tris-HCl pH 8.0, 150 mM NaCl), twice with high salt washing buffer (0.1% SDS, 1% Triton X 100, 2 mM EDTA, 20 mM Tris-HCl pH 8.0, 500 mM NaCl), once with LiCl washing buffer (20 mM Tris-HCl pH 8.0, 500 mM LiCl, 1 mM EDTA, 1% NP-40, 1% Sodium deoxycholate) and twice with 1X TE, and eluted from antibody-coated beads by incubating for one hour in 200 µl of IP elution buffer (1x TE (pH 7.5), 1% SDS, 0.1 M NaHCO3) at 65°C with 1300 rpm shaking. Chromatin crosslinks were reversed by incubating overnight at 65°C, and the DNA fragments were purified with the phenol-chloroform purification method. Samples were prepared and sequenced at Macrogen using the TruSeq DNA Sample prep Kit (Illumina, San Diego, CA), HiSeq 2500 system (Illumina, San Diego, CA), and 100 bp single-end sequencing. Approximately 23 million single-end reads were obtained for each sample.

Raw sequencing reads were adapter and quality trimmed with cutadapt version 1.16 in Trim Galore version 0.5.0 using default parameters. Bowtie2 version 2.3.4.1 was used to align trimmed reads to reference genome hs37d5 and reads with a mapping quality below 20 were

filtered out with samtools version 1.7. Macs2 version 2.1.2 was used to call broadPeaks (see supplemental figure 2 below for representative locus of the ChIP-seq data tracks). We used default parameters with the exception of the FDR cutoff, which was set to 0.01 for narrowPeak regions. Peaks in other chromosomes than autosomes, and ENCODE blacklisted genomic regions [1] were removed.



Supplemental Figure 2. Representative locus of ChIP-seq data tracks. a) Raw ChIP-seq reads. **b**) ChIP-seq fragment pileup data.

DiffBind version 2.14.0 in R version 3.6.3 was used to calculate the binding affinity of H3K9me3. Read counts in broadPeaks peaks were calculated with the dba.count -function with parameters minOverlap=2 and summits=250, to include peaks present in at least two samples in the analysis. The concentration of H3K9me3 binding in peaks in both cases and controls was calculated with the dba.analyze -function and extracted with the dba.report -function with parameter th=1. Bedtools intersect with parameter -f 0.2 was used to retrieve

H3K9me3 peaks overlapping more accessible and less accessible regions. The nearest transcription start site (TSS) was annotated with HOMER annotatePeaks.pl against hg19 v6.4 RefSeq database [4]. Ggplot2 version 3.2.1 in R version 3.6.1 was used in visualization. Of the peaks that were located at transcription start sites (TSS +/-1 kbp), 500 peaks with the most positive log2 fold changes (log2FCs) and 500 peaks with the most negative log2FCs were annotated, and genes were submitted to IPA pathway analysis. Mann-Whitney U test with alternative hypothesis "greater" in R version 3.6.1 was used to compare RNA-seq log2FCs of TSS peaks with less H3K9me3 and no change in H3K9me3 levels.

H3K9me3 concentration for all reproducible open chromatin regions (n=214,762) was calculated with the dba.count -function as described above, but without *minOverlap* and *summits* -parameters. Open chromatin regions were annotated with HOMER annotatePeaks.pl against hg19 v6.4 RefSeq database and visualized with ggplot version 3.2.1 (in R version 3.6.1). We defined TSSs based on HOMER annotation +/- 1 kbp.

Exome and whole-genome sequencing

Samples from patients III:1 and III:3 were prepared for exome sequencing with the NEBNext DNA sample Prep Reagent Set 1 (New England Biolabs, Ipswich, MA) and SureSelect Human All Exon Kit V1 (Agilent, Santa Clara, CA) capture kit. We used the SureSelect^{XT} Target Enrichment System for Illumina Paired-End Sequencing Library v1.2 protocol and modified multiplex primers. The ten samples from the separate PTC affected families were prepared for exome-sequencing using the KAPA Hyper Prep Kit (Kapa Biosystems, Wilmington, MA) and SeqCap EZ Exome + UTR (Roche Nimblegen) target enrichment. We used 500-1000 ng of DNA sheared with Covaris S220 Focused‐ultrasonicator (Covaris, Woburn, MA) to an average size of 350 bp. Paired‐end sequencing (size) was carried out

with Illumina HiSeq 2000 or 4000 (Illumina, San Diego, CA) at Karolinska High Throughput Center.

Whole-genome sample preparation and sequencing was performed by Complete Genomics (CG; Mountain View, CA) using the company's proprietary "DNA nanoball" sequencing technology [7].

Sanger sequencing

Segregation of chosen variants was studied by Sanger sequencing. In addition, the full *KDM4C* JmjC domain of all cases in the validation set was Sanger sequenced. Primers were designed using Primer3web version 4.0.0 (http://primer3.ut.ee/), and capillary sequencing was performed at the Institute for Molecular Medicine Finland (FIMM) using the BigDye v.3.1 sequencing reaction and ABI3730xl DNA Analyzer electrophoresis (Applied Biosystems, Foster City, CA). Sequences were analyzed using MutationSurveyor and FinchTV 1.5.

RNA-seq

RNA was extracted from cells using Trizol. Residual DNA was removed with DNase and the samples were cleaned with Qiagen RNeasy MinElute cleanup kit (Qiagen, Hilden, Germany). Sample quality was evaluated with the Agilent 2200 TapeStation system (Agilent Technologies, Santa Clara, CA). Total RNA libraries were prepared and sequenced at Macrogen Europe using TruSeq Stranded Total RNA LT Sample Prep Kit Human/Mouse/Rat (Illumina, San Diego, CA), the Novaseq 6000 platform (Illumina, San Diego, CA), and paired-end sequencing with 100 bp read length. Sequencing was carried out in a single batch. TrimGalore (v0.5.0) was used for quality and adapter trimming of sequenced reads. Reads

were aligned to reference genome (GRCh37) with HISAT2 (v2.1.0). The aligned reads were assembled to transcripts with StringTie (v1.3.4) and reads in each gene were counted with StringTie's prepDE.py [8]. Differential expression was analyzed with DESeq2 [9] version 1.28.0 in R version 4.0.0 for genes that have at least 10 reads in at least 2 samples (n=20413). Visualizations were made with ggplot2 version 3.3.0 in R version 4.0.0.

References

1 Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* 2019;**9**:9354.

2 Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, Silva TC, Groeneveld C, Wong CK, Cho SW, Satpathy AT, Mumbach MR, Hoadley KA, Robertson AG, Sheffield NC, Felau I, Castro MAA, Berman BP, Staudt LM, Zenklusen JC, Laird PW, Curtis C, Cancer Genome Atlas Analysis Network, Greenleaf WJ, Chang HY. The chromatin accessibility landscape of primary human cancers. *Science* 2018;**362**. doi:10.1126/science.aav1898

3 Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 2013;**10**:1213–8.

4 Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010;**38**:576–89.

5 Sheffield NC, Bock C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics* 2016;**32**:587–9.

6 Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu Y-C, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shoresh N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh K-H, Feizi S, Karlic R, Kim A-R, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJM, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai L-H, Wang

W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T, Kellis M. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;**518**:317–30.

7 Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, Carnevali P, Nazarenko I, Nilsen GB, Yeung G, Dahl F, Fernandez A, Staker B, Pant KP, Baccash J, Borcherding AP, Brownley A, Cedeno R, Chen L, Chernikoff D, Cheung A, Chirita R, Curson B, Ebert JC, Hacker CR, Hartlage R, Hauser B, Huang S, Jiang Y, Karpinchyk V, Koenig M, Kong C, Landers T, Le C, Liu J, McBride CE, Morenzoni M, Morey RE, Mutch K, Perazich H, Perry K, Peters BA, Peterson J, Pethiyagoda CL, Pothuraju K, Richter C, Rosenbaum AM, Roy S, Shafto J, Sharanhovich U, Shannon KW, Sheppy CG, Sun M, Thakuria JV, Tran A, Vu D, Zaranek AW, Wu X, Drmanac S, Oliphant AR, Banyai WC, Martin B, Ballinger DG, Church GM, Reid CA. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 2010;**327**:78–81.

8 Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 2015;**33**:290–5.

9 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**:550.