



OPEN ACCESS

Original research

Novel germline variant in the histone demethylase and transcription regulator KDM4C induces a multi-cancer phenotype

Riku Katainen ,¹ Iikka Donner,¹ Maritta Räsänen,¹ Davide Berta,¹ Anna Kuosmanen,¹ Eevi Kaasinen,¹ Marja Hietala,² Lauri A Aaltonen¹

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/jmedgenet-2021-107747>).

¹Applied Tumor Genomics Research Program and Department of Medical and Clinical Genetics, University of Helsinki Faculty of Medicine, Helsinki, Finland

²Department of Clinical Genetics, TYKS Turku University Hospital and University of Turku Institute of Biomedicine, Turku, Finland

Correspondence to

Dr Lauri A Aaltonen, Applied Tumor Genomics Research Program and Department of Medical and Clinical Genetics, University of Helsinki Faculty of Medicine, Helsinki 63, Finland; lauri.aaltonen@helsinki.fi

RK, ID and MR contributed equally.

Received 23 January 2021
Accepted 25 May 2021
Published Online First 19 July 2021



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Katainen R, Donner I, Räsänen M, *et al.* *J Med Genet* 2022;**59**:644–651.

ABSTRACT

Background Genes involved in epigenetic regulation are central for chromatin structure and gene expression. Specific mutations in these might promote carcinogenesis in several tissue types.

Methods We used exome, whole-genome and Sanger sequencing to detect rare variants shared by seven affected individuals in a striking early-onset multi-cancer family. The only variant that segregated with malignancy resided in a histone demethylase KDM4C. Consequently, we went on to study the epigenetic landscape of the mutation carriers with ATAC, ChIP (chromatin immunoprecipitation) and RNA-sequencing from lymphoblastoid cell lines to identify possible pathogenic effects.

Results A novel variant in *KDM4C*, encoding a H3K9me3 histone demethylase and transcription regulator, was found to segregate with malignancy in the family. Based on Roadmap Epigenomics Project data, differentially accessible chromatin regions between the variant carriers and controls enrich to normally H3K9me3-marked chromatin. We could not detect a difference in global H3K9 trimethylation levels. However, carriers of the variant seemed to have more trimethylated H3K9 at transcription start sites. Pathway analyses of ChIP-seq and differential gene expression data suggested that genes regulated through KDM4C interaction partner EZH2 and its interaction partner PLZF are aberrantly expressed in mutation carriers.

Conclusions The apparent dysregulation of H3K9 trimethylation and KDM4C-associated genes in lymphoblastoid cells supports the hypothesis that the *KDM4C* variant is causative of the multi-cancer susceptibility in the family. As the variant is ultrarare, located in the conserved catalytic JmjC domain and predicted pathogenic by the majority of available *in silico* tools, further studies on the role of *KDM4C* in cancer predisposition are warranted.

INTRODUCTION

Cancers arise from the accumulation of both genetic and epigenetic alterations. Numerous inherited genetic changes have been shown to predispose to certain or multiple cancer types through, for instance, defective DNA repair or cell-cycle checkpoint mechanisms. Examples of this are damaging mutations in the tumour suppressor genes *TP53*, *BRCA1* and *BRCA2* that cause the well-known cancer syndromes Li-Fraumeni and hereditary

breast and ovarian cancer. Cancer predisposing variants in epigenetic modifiers, such as *TET2* and *LSD1/KDMA1*, have also been reported.^{1–3} Mutations in this type of genes can cause epigenetic dysregulation through, for example, aberrant DNA or histone methylation, which can subsequently alter the gene expression profiles of cells and thus promote carcinogenesis in several tissue types.^{4–6} Families with multiple early-onset cancers are instrumental when searching for novel susceptibility genes that can exert pathogenic effects in multiple tissue types.

To this end, we studied a striking early-onset multi-cancer family with six confirmed cases of papillary thyroid carcinoma (PTC), as well as cases of chondrosarcoma and myeloid leukaemia. Some of the PTC affected individuals had also been diagnosed with other cancers, such as melanoma, breast cancer and adenocarcinoma of the cervix. According to patient records of affected relatives, there had been an additional case of thyroid cancer and two cases of leukaemia of unknown type among deceased first-degree relatives.

We found a previously undescribed missense change (NP_055876.2:p.(His217Arg)) in the catalytic Jumonji C (JmjC) domain of histone demethylase KDM4C (hereafter referred to as KDM4C:p.H217R) to segregate with malignancy in the family. KDM4C has numerous roles in genetic and epigenetic regulation. The JmjC domain of KDM4C specifically demethylates trimethylated lysine-9 and lysine-36 residues of histone H3 (H3K9me3 and H3K36me3), thus affecting chromatin conformation and gene expression. In addition, amino acids 1–333, which include the JmjC domain, have been shown to regulate transcriptional repression by interacting with enhancer of zeste homolog 2 (EZH2).⁷ EZH2, a major regulator of gene expression, is frequently overexpressed and underexpressed in a wide variety of cancerous tissue types, including blood, prostate and breast.^{8,9} KDM4C also interacts with lysine-specific histone demethylase 1A (LSD1) and has a role in regulating the androgen receptor (AR).¹⁰ To determine the function of the novel KDM4C variant, differences in chromatin accessibility, H3K9me3 distribution and gene expression in lymphoblastoid cells between KDM4C:p.H217R carriers and healthy controls were evaluated. Our findings suggest that KDM4C:p.H217R is causative of the multi-cancer phenotype of the described

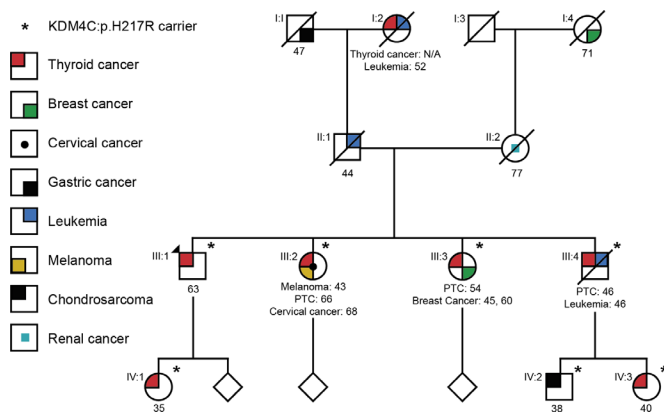


Figure 1 Finnish multi-cancer family. Cancer types and ages at diagnosis are listed beneath the individuals. Confirmed KDM4C mutation carriers are marked with an asterisk. PTC, papillary thyroid carcinoma.

family, and that the effect leads to changes in chromatin accessibility and histone epigenetics. Understanding the impact of demethylase activity is critical when targeting epigenetic modifiers in the development of future cancer therapeutics.

METHODS

Please see online supplemental file 1 (online supplemental methodology) for a complete listing and descriptions of used methods.

Patients and samples

A family with a striking history of cancer, PTC in particular, was forwarded to us by a clinical geneticist. In two generations, there were six cases of PTC, as well as cases of myeloid leukaemia, chondrosarcoma, ductal breast carcinoma, melanoma and adenocarcinoma of the cervix (figure 1). According to patient records, deceased family members had also been affected by at least thyroid cancer and leukaemia of unknown type (figure 1). Peripheral blood samples were collected from individuals III:1, III:2, III:3 and IV:1, whereas for patients III:4, IV:2 and IV:3, formalin-fixed paraffin-embedded (FFPE)-tissue samples were obtained. We also collected blood samples of 10 patients from four unrelated Finnish PTC families.

We used the Finnish Cancer Registry (FCR) to collect a validation set consisting of 59 probable familial cases of PTC. We had earlier performed systematic municipality at birth and family name at birth-based clustering of all cancer cases in the registry, which comprises all cancer diagnoses made in Finland after 1953. For a thorough description of the method, please refer to Kaasinen *et al.*¹¹ For the validation set, we collected FFPE-tissue

samples from high-scoring PTC clusters, which are likely to consist of true relatives.

DNA and cell lines

Genomic DNA for exome, whole-genome and Sanger sequencing was extracted from blood and FFPE-tissue samples with the phenol-chloroform method.

Peripheral blood lymphocytes from patients III:1, III:2, III:3 and IV:1 were infected with Epstein-Barr virus (EBV) to generate lymphoblastoid cell lines. However, we used only three cell lines in this study, since the cell line of patient III:2 failed to grow. As controls, we used in-house lymphoblastoid cell lines from four unrelated cancer-free individuals. All used cell lines had low and similar estimated passage numbers (online supplemental table 1 (online supplemental data 1)).

Variant calling and analysis

WES data were processed with GATK best practices variant calling pipeline (GRCh37). WGS raw data were processed (primary analyses, alignment and variant calling) by Complete Genomics (2011). Variant analyses, including quality filtering, visualisation, sample comparison, gene annotation (Ensembl genes v89) and control data filtering (gnomAD v2.1.1), were performed with BasePlayer.¹²

The following parameters were used in the variant analysis:

- ▶ gnomAD allele frequency: 0.
- ▶ Non-synonymous/indel/splice site variant shared by all three samples.
- ▶ Allelic fraction: $\geq 20\%$.
- ▶ Variant and genotype quality: ≥ 20 .

The analysis resulted in four candidate variants in four genes (table 1).

ATAC-seq

ATAC-seq was performed for three KDM4C:p.H217R-carrier and four control lymphoblastoid cell lines as described in previous work.¹³ See online supplemental data 1, online supplemental table 2 and online supplemental figure 1 for a detailed description of the procedure.

ChIP-seq

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) with an antibody against H3K9me3 was performed for three KDM4C:p.H217R-carrier and four control lymphoblastoid cell lines. See online supplemental data 1 and online supplemental figure 2 for a detailed description of the procedure. Briefly, 10 million cells were crosslinked with formaldehyde and sheared to fragment length of 200–500 bp by sonication.

Table 1 Ultrarare variants shared by the affected exome or genome-sequenced individuals

Genomic location	Effect in transcript 1	Protein effect	Gene/Protein name	Segregation	Protein effect prediction*
NC_000001.10:g.223983870C>G	NM_005426.2:c.1984G>C	NP_005417.1:p.(Val662Leu)	Tumour protein p53 binding protein 2 (TP53BP2)	No	Benign
NC_000005.9:g.170305249dup	NM_022897.4:c.165+2dup	NP_075048.1:p.?	RAN binding protein 17 (RANBP17)	No	Uncertain (splice site insertion)
NC_000009.11:g.6880032A>G	NM_015061.4:c.650A>G	NP_055876.2:p.(His217Arg)	Lysine (K)-specific demethylase 4C (KDM4C)	Yes	Pathogenic
NC_000012.11:g.124915216C>T	NM_006312.5:c.1000G>A	NP_006303.4:p.(Glu334Lys)	Nuclear receptor corepressor 2 (NCOR2)	No	Pathogenic

*Protein effect prediction is based on multiple prediction tools used in VarSome search engine. Pathogenic=the majority of tools give damaging/pathogenic/deleterious prediction.

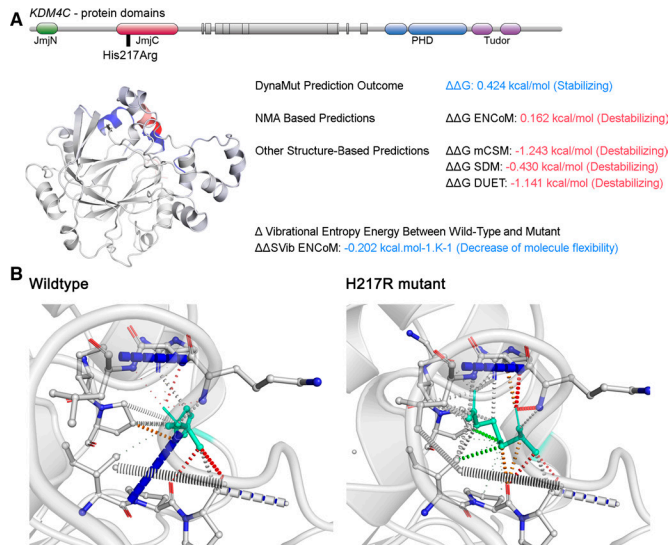


Figure 2 The location and predicted effect of p.H217R in the catalytic JmjC domain of *KDM4C*. (A) Location of the JmjC domain and p.H217R in *KDM4C*. (B) DynaMut prediction results for *KDM4C*:p.H217R.

Chromatin was immunoprecipitated with an antibody against H3K9me3 (Active Motif 39062, lot 09919003). Sequencing libraries were prepared and sequenced at Macrogen using the TruSeq DNA Sample prep Kit (Illumina, San Diego, California, USA), HiSeq 2500 system (Illumina, San Diego, California, USA) and 100 bp single-end sequencing. DiffBind V.2.14.0 in R V.3.6.3 was used to calculate the binding affinity of H3K9me3. Ggplot2 V.3.2.1 in R V.3.6.1 was used in visualisation. Mann-Whitney U test with alternative hypothesis ‘greater’ in R V.3.6.1 was used to compare RNA-seq log₂FCs of TSS peaks with less H3K9me3 and no change in H3K9me3 levels.

Ingenuity pathway analysis

In total, 136 significantly differentially expressed genes (DESeq2 adjusted p value ≤ 0.05) were analysed using Ingenuity pathway analysis (IPA; Qiagen, Redwood City, California, USA) version 52912811 core analysis feature.

We performed the same analysis (with and without expression scores separately) for genes with the top log₂FCs of H3K9me3 binding at their transcription start sites. Five hundred genes with the highest positive and 500 genes with the highest negative score were included.

RESULTS

A novel germline missense variant in *KDM4C* segregates in a multi-cancer family

Variant analysis of three affected siblings in the family revealed a novel variant (NM_015061.4:c.650A>G; Chr9(GRCh37):g.6880032A>G; NP_055876.2:p.(His217Arg)) in *KDM4C*. The heterozygous p.H217R variant is located in the JmjC (Jumonji) domain (figure 2A), which is responsible for the demethylation and protein interaction associated activity of *KDM4C*. The variant is not listed in the gnomAD (n=125 748), FinnGen (n=269 077) or UK Biobank (n=488 000) databases and it is predicted damaging and disease causing by the majority (18/21) of in silico prediction tools available through VarSome search engine.¹⁴ Also, the variant locus is highly conserved (GERP conservation score 5.1599). The variant domain did not harbour any somatic cancer mutation hotspots according to COSMIC and ICGC data. We analysed the putative effect of

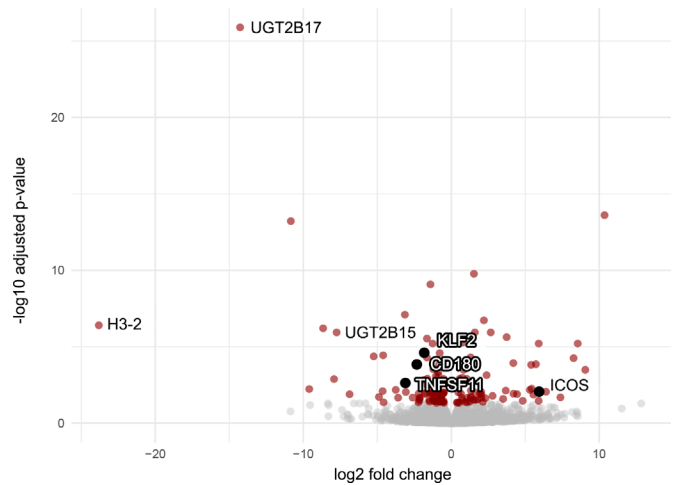


Figure 3 Changes in gene expression due to *KDM4C*:p.H217R. Significantly differentially expressed genes between *KDM4C*:p.H217R carriers and non-carriers (red and black dots). Differentially expressed genes associated with PLZF activation (black dots). PLZF, promyelocytic leukaemia zinc finger protein.

H217R on the structure of the *KDM4C* catalytic domain with DynaMut and MutPred2 tools.^{15 16} The DynaMut Outcome Score was 0.424 kcal/mol (Stabilising), suggesting a gain-of-function effect of the variant (figure 2B). MutPred2 predicted the variant as pathogenic with a score of 0.8. We used Sanger sequencing to analyse the presence of all found candidate variants (table 1) in the rest of the affected family members of whom we had DNA material available. The variant in *KDM4C* was the only one found in all seven patients. We screened the whole JmjC domain of *KDM4C* for additional variants in 63 Finnish families with possible susceptibility to PTC, as this was the dominant cancer type in the family. We found no p.H217R or other non-synonymous variants within this domain through the screening, nor did we detect loss of heterozygosity in available FFPE tumour samples.

Gene expression changes predict activation of PLZF in *KDM4C*:p.H217R carriers

We examined the effect of the discovered *KDM4C*:p.H217R variant on gene expression by RNA-seq. Using DESeq2, we identified 136 differentially expressed genes (adjusted p value ≤ 0.05) (figure 3 and online supplemental table 3 (online supplemental data 2)), of which 76 were downregulated (log₂FC<0) and 60 upregulated (log₂FC>0) in *KDM4C*:p.H217R carriers. No change was detected in the expression level of *KDM4C* (log₂FC=-0.12, adjusted p value=0.93) between *KDM4C*:p.H217R carriers and non-carriers.

We analysed the differentially expressed genes and their fold changes with IPA. The results indicated an overall decrease in the activity of the xenobiotic metabolism CAR and PXR signalling pathways (online supplemental figure 3 (online supplemental data 3)). Changes in gene expression predicted the activation (activation z-score 2.000, p value of overlap 0.00497) of transcription factor, promyelocytic leukaemia zinc finger protein (PLZF) (encoded by *ZBTB16*; online supplemental table 4 (online supplemental data 4)), which is an interaction partner of EZH2.¹⁷ PLZF was the only upstream regulator predicted to be activated with a significant overlap p value.

ATAC-seq reveals differential DNA accessibility at H3K9me3 occupied regions

We examined the effect of KDM4C:p.H217R on chromatin accessibility in lymphoblastoid cell lines of variant carriers ($n=3$) and non-carriers ($n=4$) with ATAC-seq. Altogether, 214 762 reproducible open chromatin regions were identified in the seven samples. First, we examined if open chromatin profiles of KDM4C:p.H217R carriers differ from the non-carrier control lymphoblastoid cell lines more than randomly chosen combinations of these samples differ from each other. We divided six samples (three KDM4C:p.H217R carriers and three non-carriers) randomly into two groups and compared the numbers of differentially accessible regions (DARs, $FDR \leq 0.05$) and the enrichment of DARs at histone modifications from the Roadmap Epigenomics Project. The largest number of DARs (2141) was identified when carriers of KDM4C:p.H217R were compared against non-carriers, for which the enrichment to active chromatin marks and histone 3 lysine 9 trimethylation (H3K9me3) was the most significant (figure 4A). The next largest numbers of DARs were 1560, 1025, 741 and 558 in random groups. The rest had less than 300 DARs.

Next, we analysed differential accessibility between the three KDM4C:p.H217R carriers and four healthy non-carriers (the ones described previously and an additional sample), and identified 2225 more accessible ($\log_2FC > 0$ and $FDR \leq 0.05$) and 957 less accessible ($\log_2FC < 0$ and $FDR \leq 0.05$) regions (online supplemental table 5 (online supplemental data 5)). When compared with all regions of open chromatin, the more accessible regions in KDM4C:p.H217R carriers displayed enrichment at the H3K9me3 histone mark derived from Roadmap Epigenomics Project data on blood cells (figure 4B). The finding is compatible with the specific H3K9me3 demethylase activity of KDM4C.

The less accessible regions in KDM4C:p.H217R carriers also displayed enrichment at the H3K9me3 mark, pointing to a general dysregulation of H3K9 trimethylation (figure 4C). In addition, the less accessible regions showed an enrichment at the active chromatin marks H3K4me1 and H3K27ac, as well as at open chromatin regions defined by DNase hotspots. This suggests that these normally active sites are repressed in KDM4C:p.H217R carriers.

Next, we examined if the changes in chromatin accessibility observed in KDM4C:p.H217R carriers were connected to changes in gene expression. We annotated the closest gene to the more and less accessible regions, and compared the gene expression \log_2FC s of the genes in these groups. We found that genes in the proximity of more accessible regions, especially those with more accessible TSSs, tend to have higher \log_2FC s than genes in regions where no change in chromatin accessibility had been detected, which suggests that increased chromatin accessibility at these sites is enhancing nearby gene expression (figure 4D). The same phenomenon was observed at less accessible sites that showed decreased expression in carriers versus controls. This confirms the quality of the performed ATAC-seq experiments and provides evidence for the functional significance of the DARs discovered in KDM4C:p.H217R carriers.

Changes in H3K9me3 levels at transcription start sites point to aberrant interaction with EZH2

Next, we analysed the distribution of H3K9me3 in KDM4C:p.H217R carrier and non-carrier lymphoblastoid cells with ChIP-seq. We identified 174 611 H3K9me3 binding sites present in at least two samples. Genome-wide significant changes

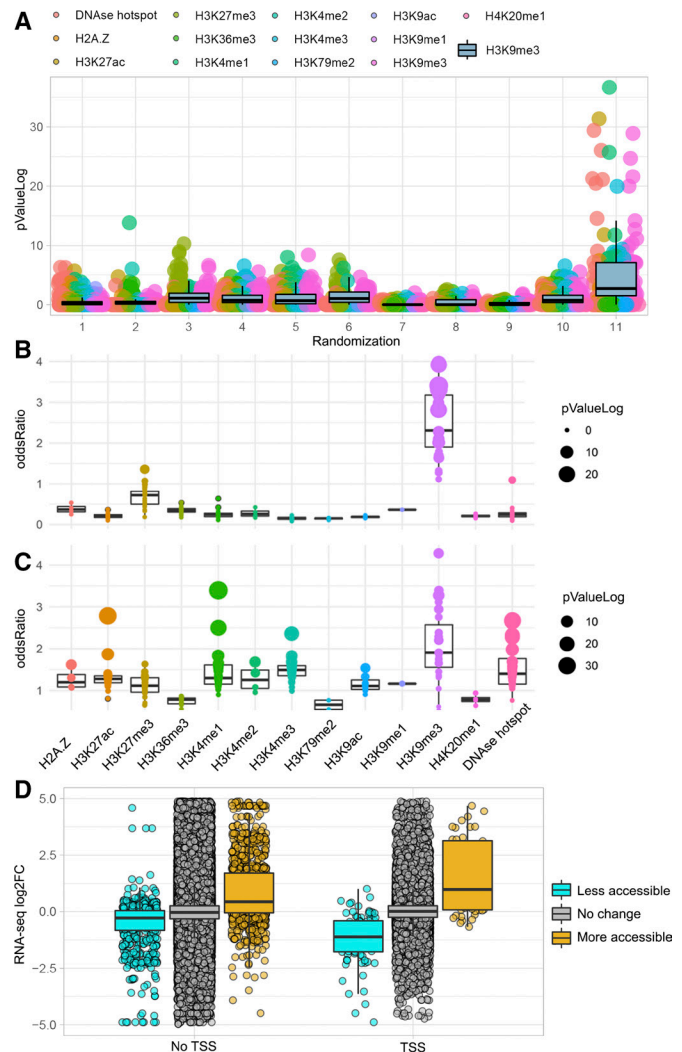


Figure 4 Changes in chromatin accessibility due to KDM4C:p.H217R. (A) Enrichment of DARs of randomly generated sample groups to chromatin marks. Group 11 (three cases vs three controls divided by the carrier status of KDM4C:p.H217R) shows the highest enrichment of DARs to chromatin marks. Dots represent the enrichment of a specific chromatin mark in a specific blood cell type, the boxplot shows the enrichment of H3K9me3 in all blood cell types in the specific randomisation setup. (B) More accessible regions in KDM4C:p.H217R carriers are enriched for the H3K9me3 mark. (C) Less accessible regions in KDM4C:p.H217R carriers enrich in activating chromatin marks (H3K4me1 and H3K27ac) and the H3K9me3 mark. (D) Chromatin accessibility affects expression of proximal genes. More accessible regions in proximity enhance gene expression both at TSS ($p=7.989e-07$) and at non-TSS regions ($p<2.2e-16$). Similarly, less accessible regions in proximity decrease gene expression both at TSS ($p=4.398e-14$) and at non-TSS regions ($p<2.2e-16$). Statistical significance calculated with Welch Two Sample t-test.

($FDR \leq 0.05$) in H3K9me3 levels were not detected between KDM4C:p.H217R carriers and non-carriers. We examined the H3K9me3 levels at the H3K9me3 peaks overlapping with more and less accessible regions defined by ATAC-seq, as the previous analyses had revealed an enrichment of DARs at H3K9me3 occupied loci. At the H3K9me3 peaks overlapping the more accessible sites ($n=118$), H3K9me3 levels were higher in cases (figure 5A). At the less accessible sites ($n=37$), there was no such difference in H3K9me3 levels between KDM4C:p.H217R carriers and non-carriers. Next, we analysed H3K9me3 levels

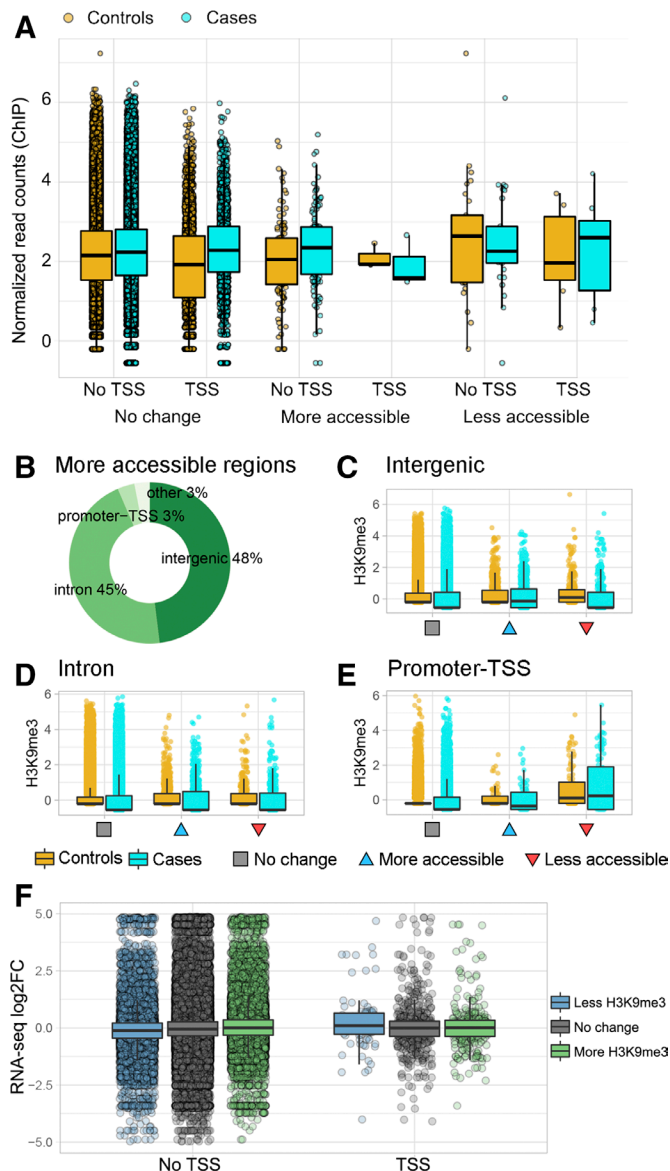


Figure 5 H3K9me3 levels are not equal between KDM4C:p.H217R carriers and non-carriers at transcription start sites and open chromatin regions. (A) H3K9me3 level (normalised read/fragment counts) comparisons between carriers and non-carriers of KDM4C:p.H217R at ChIP-peaks overlapping TSSs (± 1000 bp) and other sites grouped by chromatin accessibility change. (B) More accessible regions map to intergenic regions and introns. (C) H3K9me3 level (normalised read/fragment counts) comparisons between carriers and non-carriers of KDM4C:p.H217R at intergenic (D) at intronic (E) at promoter-TSS open chromatin peaks grouped by chromatin accessibility change. (F) Gene expression fold change between carriers and non-carriers of KDM4C:p.H217R stratified by H3K9me3 changes. Genes are grouped based on less H3K9me3 ($\log_2FC < -1$), more H3K9me3 ($\log_2FC > 1$) or no change in H3K9me3 level ($-1 < \log_2FC < 1$) at TSS (± 1000 bp) or outside TSS. Genes with less H3K9me3 at their TSS have higher expression. FC, fold change; TSS, transcription start site.

at TSSs (± 1000 bp) and outside TSS regions. TSSs displayed increased H3K9me3 levels in KDM4C:p.H217R carriers with no change in chromatin accessibility, and regions outside TSS displayed increased H3K9me3 levels in KDM4C:p.H217R carriers when accessibility was increased (figure 5A).

Due to the low number of obtained data points at more and less accessible sites (118 and 37, respectively), we performed H3K9me3 distribution analyses for all 214762 reproducible open chromatin regions, regardless of overlapping ChIP-seq peaks. We were interested in H3K9me3 levels at more accessible regions in particular, as in the previous analysis described above (figure 5A) we measured—counterintuitively—higher levels of H3K9me3 at the 118 sites included in the analysis. Most more accessible regions were located in intergenic and intronic regions (figure 5B). In this analysis, including all open chromatin regions, we detected reduced H3K9me3 levels in KDM4C:p.H217R carriers at intergenic and intronic regions, as well as more accessible TSSs (figure 5C–E). At less accessible TSSs, KDM4C:p.H217R carriers had more H3K9me3 than controls (figure 5E). These data are consistent with the ATAC-seq findings as well as the gain-of-function prediction of the mutation.

As H3K9me3 was differentially distributed at TSSs between KDM4C:p.H217R carriers and non-carriers, we examined if these differences contribute to changes in gene expression. We divided H3K9me3 peaks ($n=174611$) into groups based on \log_2FC in KDM4C:p.H217R carriers: less H3K9me3 ($\log_2FC < -1$), more H3K9me3 ($\log_2FC > 1$) and no change in H3K9me3 level ($-1 < \log_2FC < 1$). These peaks were further divided into TSS (± 1000 bp) and non-TSS regions, and expression \log_2FC of the nearest gene of each peak was examined. We detected slightly higher gene expression \log_2FC s, although not statistically significant (Mann-Whitney U test, $W=28884$, p value= 0.06752), in genes with less H3K9me3 at their TSS compared with genes with no change in H3K9me3 at TSS (figure 5F). This is compatible with the role of H3K9me3 in gene repression.¹⁸

We performed IPA for genes with H3K9me3 ChIP-seq peaks at TSS showing the highest positive and negative \log_2FC . The analysis without gene expression \log_2FC information identified the transcription factor EZH2 as one of the most significant upstream regulators, with an overlap p value of 0.000624 based on 23 target molecules in the dataset (online supplemental table 6 (online supplemental data 6)). We also performed the analysis for the same gene set with expression \log_2FC scores. PLZF was predicted to be activated (z -score= 2.646) based on changes in gene expression, although p value of overlap was not significant (0.0768) (online supplemental table 7 (online supplemental data 7)). This provides more evidence of the activation of PLZF in KDM4C:p.H217R carriers.

DISCUSSION

Here, we describe a novel germline variant in the highly conserved catalytic JmjC domain of the histone demethylase *KDM4C*. The missense variant was discovered by exome and whole-genome sequencing of selected multi-cancer family members and was confirmed to segregate with malignancy in the family. We did not detect the variant in our validation set consisting of 63 candidate Finnish PTC families or any major publicly available datasets. The consequent amino acid change (NP_055876.2:p.(His217Arg)) is predicted pathogenic by the majority of commonly used in silico variant prediction tools.

The KDM4 subfamily histone demethylases are epigenetic regulators that control chromatin structure and thus gene expression by demethylating histones H3K9, H3K36 and H1.4K26. The subfamily consists of five proteins (KDM4A–E), all of which harbour the catalytic Jumonji C domain (JmjC). The selectivity of the demethylases is determined by multiple interactions within the catalytic domain. KDM4C specifically catalyses

the demethylation of trimethylated lysine-9 and lysine-36 residues of histone H3. In addition, it has been shown to be involved in transcriptional regulation through interaction with EZH2.⁷ Both of these functions are carried out by the JmjC domain, in which KDM4C:p.H217R is located. Recently, rare variants in *KDM4C* were associated with schizophrenia and autism spectrum disorder,¹⁹ and the encoded protein was shown to control tumorigenesis by epigenetically regulating p53 and c-Myc.²⁰

To confirm that the variant indeed has biological consequences, we performed ATAC, ChIP and RNA-sequencing of EBV-transformed lymphoblastoid cell lines derived from three patients carrying the *KDM4C*:p.H217R mutation and four healthy controls. These results have limitations due to the small number of samples and characteristics of EBV-transformed cell lines as transformation is known to cause dysregulation of gene expression and cancer-related pathways.²¹ It should also be noted that all *KDM4C*:p.H217R carrier cell lines shared an additional possibly pathogenic variant in the *Nuclear receptor corepressor 2* (*NCOR2*). *NCOR2* has a role in AR signalling and acts as a transcriptional corepressor by promoting chromatin condensation.²²

RNA-seq analysis revealed 136 differentially expressed genes. The gene with the highest fold change with significant p value was *H3-2*, a member of the H3 family of histones (figure 3). Interestingly, according to the STRING protein-protein interaction network,²³ Histone cluster 2 H3 pseudogene 2 (*H3-2*; *HIST2H3PS2*) and *KDM4C* are interaction partners with an interaction and experiment scores of 0.718 and 0.336, respectively, making *H3-2* one of the highest scoring interaction partners of *KDM4C* in the database. The possible pathogenic significance of *H3-2*, if any, needs further work.

The most significantly differentially expressed gene, *UGT2B17*, has previously been associated with increased risk of prostate cancer.²⁴ Another gene coding a similar enzyme, *UGT2B15*, also had significantly lower expression in the *KDM4C*:p.H217R carriers as compared with controls (figure 3). *UGT2B15* and *UGT2B17* are regulated by the AR and are not expressed in AR negative prostate cancer cell lines.²⁵ We did not detect significant differences in AR expression between the mutation carriers and non-carriers, however, *KDM4C* has been shown to interact and co-localise with AR, and more importantly, to regulate its function.^{10,26} The *NCOR2* variant could also have contributed to this result, as the protein is known to regulate AR signalling.

Upstream regulator analysis revealed that the expression changes of *CD180*, *ICOS*, *KLF2* and *TNFSF11* are consistent with increased activity of the EZH2-associated protein PLZF. PLZF and EZH2 co-associate at the chromatin level, and EZH2 activity was recently shown to regulate PLZF transcriptional output.¹⁷ PLZF is an epigenetic regulator that balances self-renewal and differentiation of haematopoietic cells through binding of chromatin-modifying factors, and gene rearrangements at this locus have been associated with acute promyelocytic leukaemia, a subtype of acute myeloid leukaemia and the very disease the gene was named after.^{27,28} Since one member of the studied family suffered from myeloid leukaemia, and two others of leukaemia of unknown type, and as the measurements were performed with blood samples, this result is particularly interesting.

ATAC-seq revealed 2225 more and 957 less accessible sites between *KDM4C*:p.H217R carriers and controls. These DARs enrich at normally H3K9me3-marked chromatin, pointing to epigenetic plasticity at these sites in *KDM4C*:p.H217R carrier cells. When we examined levels of H3K9me3 at open chromatin regions, we found that intergenic and intronic regions have reduced H3K9me3 in *KDM4C*:p.H217R carriers. As *KDM4C*

specifically demethylates H3K9me3 and the variant is predicted to be gain-of-function (figure 2B), the changes in chromatin accessibility and H3K9me3 levels at these loci support the functional significance of *KDM4C*:p.H217R.

We examined the distribution of H3K9me3 in *KDM4C*:p.H217R carrier cells with ChIP-seq. We did not detect a global reduction or increase in H3K9me3 levels in *KDM4C*:p.H217R carriers compared with controls. However, we detected increased H3K9me3 levels between *KDM4C*:p.H217R carriers and non-carriers at transcription start sites with no change in chromatin accessibility. This result is of particular interest because transcription start sites are commonly occupied by *KDM4C*.²⁹ We examined how the changes in H3K9me3 levels at TSS contribute to gene expression changes in *KDM4C*:p.H217R carriers, and found that reduced H3K9me3 at TSSs weakly associates with increased gene expression. Interestingly, genes with increased H3K9me3 levels at their TSS did not have globally decreased expression compared with genes with no change in H3K9me3 levels. Thus, increased H3K9me3 at TSSs does not seem to globally repress gene expression in *KDM4C*:p.H217R carriers.

KDM4C associates with EZH2 and their binding sites overlap. Full EZH2-associated gene repression requires the co-occupation of *KDM4C* and EZH2 at gene promoters.⁷ When we analysed the genes with the highest positive and negative log₂FCs of H3K9me3 ChIP-seq peaks at their transcription start sites (ie, sites and genes that are likely to be occupied by dysfunctional *KDM4C* and dysregulated through aberrant H3K9 methylation), EZH2 was one of the most significant upstream regulators according to IPA. Thus, *KDM4C*:p.H217R may also affect the interaction between *KDM4C* and EZH2. We further examined expression changes of genes with differential H3K9me3 levels at their TSS with IPA, and identified predicted activation of upstream regulator PLZF. Our findings raise the possibility that TSSs with dysregulated H3K9me3 levels in *KDM4C*:p.H217R carriers are regulated by EZH2 and their expression changes, as predicted through IPA, point to activation of its interaction partner PLZF.

Our data suggest that *KDM4C*:p.H217R drives tumourigenesis in the studied family by perturbing the demethylase and transcriptional regulator functions of *KDM4C* carried out by the JmjC domain. Multiple layers of data support this conclusion, and genetic validation of the finding in other multi-cancer families is an important future goal. *KDM4C*:p.H217R causes chromatin accessibility changes at H3K9me3-marked loci as detected with ATAC-seq and increased H3K9me3 levels at TSSs. EZH2 was identified as an upstream regulator of genes with dysregulated H3K9me3 levels at TSS, and increased PLZF activity was detected by both differentially expressed genes and genes with changes in H3K9me3 levels at their TSS. ATAC, ChIP and RNA-seq were performed using EBV-transformed lymphoblastoid cells derived from normal blood rather than tumour tissue. Even so, the detected difference in chromatin accessibility between *KDM4C*:p.H217R carrier and control cells is striking. The fact that these changes in chromatin conformation are already present in normal lymphoblastoid cells from mutation carriers strongly supports a role for *KDM4C*:H217R in tumour susceptibility. *KDM4C* has been detected in all tissue types and has low tissue specificity. It is thus likely to play a central role in most tissues, which would explain the multi-cancer phenotype observed in the family.

Acknowledgements We would like to thank Iina Vuoristo, Inga-Lill Åberg, Alison London, Justyna Kolakowska Sini Martinen, Marjo Rajalaako, Sirpa Soisalo and

Heikki Metsola for the technical assistance. We acknowledge CSC IT Center for Science Finland for providing computational resources.

Contributors RK and ID designed the study and performed initial variant analyses and screenings. MR and DB performed ATAC and ChIP-seq experiments. MR and EK analysed ATAC and ChIP-seq data. MR and AK analysed RNA-seq data. ID and MR performed pathway analyses. RK performed structural dynamics predictions. MH provided the samples. LAA supervised the study. RK, ID and MR wrote the manuscript. All authors read and approved the final manuscript.

Funding This research was funded by the Academy of Finland (Finnish Center of Excellence Programs 2012–2017, 250345 and 2018–2025, 312041), Juhani Aho Foundation for Medical Research, Orion Research Foundation, Cancer Foundation of Finland, Maud Kuistila Memorial Foundation and Ida Montin Foundation.

Competing interests None declared.

Patient consent for publication Obtained.

Ethics approval This study was approved by the National Supervisory Authority for Welfare and Health (Valvira; 1423/06.01.03.01/2012), National Institute for Health and Welfare (THL; 151/5.05.00/2017), and the ethics committee of the Hospital District of Helsinki and Uusimaa (HUS; 408/13/03/03/09). All living patients gave informed written consent.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. All data relevant to the study are included in the article or uploaded as supplementary information. Data that potentially allow identification of individuals must be protected and thus, sequencing data produced in this study cannot be deposited to public databases. Roadmap epigenomics ChIP-seq and DNase hotspot data provided in LOLA Core and Extended databases was downloaded from <http://cloud.databio.org/regiondb/>.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Riku Katainen <http://orcid.org/0000-0003-4575-4239>

REFERENCES

- Kaasinen E, Kuismin O, Rajamäki K, Ristolainen H, Aavikko M, Kondelin J, Saarinen S, Berta DG, Katainen R, Hirvonen EAM, Karhu A, Taira A, Tanskanen T, Alkods A, Taipale M, Morgunova E, Fransila K, Lehtonen R, Mäkinen M, Aittomäki K, Palotie A, Kurki MI, Pietiläinen O, Hilpert M, Saarentaus E, Niinimäki J, Junntila J, Kaikkonen K, Vahteristo P, Skoda RC, Seppänen MRJ, Eklund KK, Taipale J, Kilpivaara O, Aaltonen LA. Impact of constitutional TET2 haploinsufficiency on molecular and clinical phenotype in humans. *Nat Commun* 2019;10:1252.
- Wei X, Calvo-Vidal MN, Chen S, Wu G, Revuelta MV, Sun J, Zhang J, Walsh MF, Nichols KE, Joseph V, Snyder C, Vachon CM, McKay JD, Wang S-P, Jayabalan DS, Jacobs LM, Becirovic D, Waller RG, Artomov M, Viale A, Patel J, Phillip J, Chen-Kiang S, Curtin K, Salama M, Atanackovic D, Niesvizky R, Landgren O, Slager SL, Godley LA, Churpek J, Garber JE, Anderson KC, Daly MJ, Roeder RG, Dumontet C, Lynch HT, Mullighan CG, Camp NJ, Offit K, Klein RJ, Yu H, Cerchietti L, Lipkin SM. Germline lysine-specific demethylase 1 (*LSD1/KDM1A*) mutations confer susceptibility to multiple myeloma. *Cancer Res* 2018;78:2747–59.
- Solary E, Bernard OA, Tefferi A, Fuks F, Vainchenker W. The ten-eleven translocation-2 (TET2) gene in hematopoiesis and hematopoietic diseases. *Leukemia* 2014;28:485–96.
- Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, Johnson NA, Severson TM, Chiu R, Field M, Jackman S, Krzywinski M, Scott DW, Trinh DL, Tamura-Wells J, Li S, Firme MR, Rogic S, Griffith M, Chan S, Yakovenko O, Meyer IM, Zhao EY, Smailus D, Moks A, Chittaranjan S, Rimsza L, Brooks-Wilson A, Spinelli JJ, Ben-Neriah S, Meissner B, Woolcock B, Boyle M, McDonald H, Tam A, Zhao Y, Delaney A, Zeng T, Tse K, Butterfield Y, Birol I, Holt R, Schein J, Horsman DE, Moore R, Jones SJM, Connors JM, Hirst M, Gascoyne RD, Marra MA. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* 2011;476:298–303.
- Rowbotham SP, Li F, Dost AFM, Louie SM, Marsh BP, Pessina P, Anbarasu CR, Brinson CF, Tuminello SJ, Lieberman A, Ryeom S, Schlaeger TM, Aronow BJ, Watanabe H, Wong KK, Kim CF. H3K9 methyltransferases and demethylases control lung tumor-propagating cells and lung cancer progression. *Nat Commun* 2018;9:4559.
- Ji H, Zhou Y, Zhuang X, Zhu Y, Wu Z, Lu Y, Li S, Zeng Y, Lu QR, Huo Y, Shi Y, Bu H. Correction: HDAC3 deficiency promotes liver cancer through a defect in H3K9ac/H3K9me3 transition. *Cancer Res* 2020;80:923.
- Das PP, Shao Z, Beyaz S, Apostolou E, Pinello L, De Los Angeles A, O'Brien K, Atsma JM, Fujiwara Y, Nguyen M, Ljuboja D, Guo G, Woo A, Yuan G-C, Onder T, Daley G, Hochedlinger K, Kim J, Orkin SH. Distinct and combinatorial functions of Jmjd2b/Kdm4b and Jmjd2c/Kdm4c in mouse embryonic stem cell identity. *Mol Cell* 2014;53:32–48.
- Duan R, Du W, Guo W. Ezh2: a novel target for cancer treatment. *J Hematol Oncol* 2020;13.
- Kim KH, Roberts CWM. Targeting EZH2 in cancer. *Nat Med* 2016;22:128–34.
- Wissmann M, Yin N, Müller JM, Greschik H, Fodor BD, Jenuwein T, Vogler C, Schneider R, Günther T, Buettner R, Metzger E, Schüle R. Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression. *Nat Cell Biol* 2007;9:347–53.
- Kaasinen E, Aavikko M, Vahteristo P, Patama T, Li Y, Saarinen S, Kilpivaara O, Pitkänen E, Knekt P, Laaksonen M, Artama M, Lehtonen R, Aaltonen LA, Pukkala E. Nationwide registry-based analysis of cancer clustering detects strong familial occurrence of Kaposi sarcoma. *PLoS One* 2013;8:e55209.
- Katainen R, Donner I, Cajuso T, Kaasinen E, Palin K, Mäkinen V, Aaltonen LA, Pitkänen E. Discovery of potential causative mutations in human coding and noncoding genome with the interactive software BasePlayer. *Nat Protoc* 2018;13:2580–600.
- Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, Kathiria A, Cho SW, Mumbach MR, Carter AC, Kasowski M, Orloff LA, Risco VI, Kundaje A, Khavari PA, Montine TJ, Greenleaf WJ, Chang HY. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* 2017;14:959–62.
- Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Albarca Aguilera M, Meyer R, Massouras A. VarSome: the human genomic variant search engine. *Bioinformatics* 2019;35:1978–80.
- Rodrigues CH, Pires DE, Ascher DB. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res* 2018;46:W350–5.
- Pejaver V, Uresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam H-J, Mort M, Cooper DN, Sebat J, Iakoucheva LM, Mooney SD, Radivojac P. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat Commun* 2020;11:5918.
- Koubi M, Poplineau M, Vernerey J, N'Guyen L, Tiberi G, Garcia S, El-Kaoutari A, Maqbool MA, Andrau J-C, Guillouf C, Saurin AJ, Duprez E. Regulation of the positive transcriptional effect of PLZF through a non-canonical EZH2 activity. *Nucleic Acids Res* 2018;46:3339–50.
- Kim J, Kim H. Recruitment and biological consequences of histone modification of H3K27me3 and H3K9me3. *ILAR J* 2012;53:232–9.
- Kato H, Kushima I, Mori D, Yoshimi A, Aleksic B, Nawa Y, Toyama M, Furuta S, Yu Y, Ishizuka K, Kimura H, Arioka Y, Tsujimura K, Morikawa M, Okada T, Inada T, Nakatochi M, Shinjo K, Kondo Y, Kaibuchi K, Funabiki Y, Kimura R, Suzuki T, Yamakawa K, Ikeda M, Iwata N, Takahashi T, Suzuki M, Okahisa Y, Takaki M, Egawa J, Someya T, Ozaki N. Rare genetic variants in the gene encoding histone lysine demethylase 4C (KDM4C) and their contributions to susceptibility to schizophrenia and autism spectrum disorder. *Transl Psychiatry* 2020;10:421.
- Lee DH, Kim GW, Yoo J, Lee SW, Jeon YH, Kim SY, Kang HG, Kim D-H, Chun K-H, Choi J, Kwon SH. Histone demethylase KDM4C controls tumorigenesis of glioblastoma by epigenetically regulating p53 and c-myc. *Cell Death Dis* 2021;12.
- Carter KL, Cahir-McFarland E, Kieff E. Epstein-Barr virus-induced changes in B-lymphocyte gene expression. *J Virol* 2002;76:10427–36.
- Cui J, Yang Y, Zhang C, Hu P, Kan W, Bai X, Liu X, Song H. Fbi-1 functions as a novel Ar co-repressor in prostate cancer cells. *Cell Mol Life Sci* 2011;68:1091–103.
- Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV, von MC. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 2019;47:D607–13.
- Karypidis A-H, Olsson M, Andersson S-O, Rane A, Ekström L. Deletion polymorphism of the UGT2B17 gene is associated with increased risk for prostate cancer and correlated to gene expression in the prostate. *Pharmacogenomics J* 2008;8:147–51.
- Bao B-Y, Chuang B-F, Wang Q, Sartor O, Balk SP, Brown M, Kantoff PW, Lee G-SM. Androgen receptor mediates the expression of UDP-glucuronosyltransferase 2 B15 and B17 genes. *Prostate* 2008;68:839–48.
- Lin C-Y, Wang B-J, Chen B-C, Tseng J-C, Jiang SS, Tsai KK, Shen Y-Y, Yuh CH, Shi Z-L, Wang W-C, Kung H-J, Chuu C-P. Histone demethylase KDM4C stimulates the proliferation of prostate cancer cells via activation of Akt and c-myc.

- Cancers* 2019;11. doi:10.3390/cancers11111785. [Epub ahead of print: 13 Nov 2019].
- 27 Chen Z, Brand NJ, Chen A, Chen SJ, Tong JH, Wang ZY, Waxman S, Zelent A. Fusion between a novel Krüppel-like zinc finger gene and the retinoic acid receptor-alpha locus due to a variant t(11;17) translocation associated with acute promyelocytic leukaemia. *EMBO J* 1993;12:1161–7.
- 28 Hussain L, Maimaitiyiming Y, Islam K, Naranmandura H. Acute promyelocytic leukemia and variant fusion proteins: PLZF-RAR α fusion protein at a glance. *Semin Oncol* 2019;46:133–44.
- 29 Pedersen MT, Agger K, Laugesen A, Johansen JV, Cloos PAC, Christensen J, Helin K. The demethylase JMJD2C localizes to H3K4me3-positive transcription start sites and is dispensable for embryonic development. *Mol Cell Biol* 2014;34:1031–45.