Title

Positional cloning and comprehensive mutation analysis identified a novel *KDM2B* mutation in a Japanese family with minor malformations, intellectual disability, and schizophrenia

Saeko Yokotsuka-Ishida[1], Masayuki Nakamura[1*], Yoko Tomiyasu[1], Mio Nagai[2], Yuko Kato[3], Akiyuki Tomiyasu[1], Hiromi Umehara[1], Takehiro Hayashi[4], Natsuki Sasaki[1], Shu-ichi Ueno[5], Akira Sano[1]

[1] Department of Psychiatry, Kagoshima University Graduate School of Medical and Dental Sciences, Kagoshima, Japan

[2] Division of Psychiatry, Matsuyama Red Cross Hospital, Matsuyama, Japan

[3] Division of Psychiatry, Jiundo Hospital, Tokyo, Japan

[4] Department of Social Welfare, the International University of Kagoshima, Kagoshima, Japan

[5] Department of Neuropsychiatry, Ehime University Graduate School of Medicine Toon, Japan

***Corresponding author:** Masayuki Nakamura

Department of Psychiatry, Kagoshima University Graduate School of Medical and Dental Sciences, 8-35-1 Sakuragaoka, Kagoshima 890-8520, Japan

Email address: nakamu36@m.kufm.kagoshima-u.ac.jp

18 Telephone number: +81-99-2755346

19 Fax number: +81-99-265-7089

20 ORCID ID (Masayuki Nakamura): 0000-0001-5558-0418

21 **Conflict of Interest Statement**

22 The authors have nothing to disclose.

23

Abstract

The importance of epigenetic control in the development of the central nervous system has recently been attracting attention. Methylation patterns of lysine 4 and lysine 36 in histone H3 (H3K4 and H3K36) in the central nervous system are highly conserved among species. Numerous complications of body malformations and neuropsychiatric disorders are due to abnormal histone H3 methylation modifiers. In this study, we analyzed a Japanese family with a dominant inheritance of symptoms including Marfan syndrome-like minor physical anomalies (MPAs), intellectual disability, and schizophrenia (SCZ). We performed positional cloning for this family using a single nucleotide polymorphism (SNP) array and whole-exome sequencing, which revealed a missense coding strand mutation (rs1555289644, NM_032590.4: c.2173G>A, p.A725T) in exon 15 on the plant homeodomain of the *KDM2B* gene as a possible cause of the disease in the family. The exome sequencing revealed that within the coding region, only a point mutation in *KDM2B* was present in the region with the highest logarithm of odds score of 2.41 resulting from whole genome linkage analysis. Haplotype analysis revealed co-segregation with four affected family members (IV-9, III-4, IV-5, and IV-8). Lymphoblastoid cell lines from the proband with this mutation showed approximately halved *KDM2B* expression in comparison with healthy controls. KDM2B acts as an H3K4 and H3K36 histone demethylase. Our findings suggest

that haploinsufficiency of *KDM2B* in the process of development, like other H3K4 and H3K36 methylation modifiers, may have caused MPAs, intellectual disability, and SCZ in this Japanese family.

Keywords

KDM2B, Positional cloning, Haplotype analysis, Whole-exome sequencing, Histone H3 lysine 4 demethylase, Schizophrenia, Marfan syndrome-like minor physical anomalies

1. Introduction

Histone methylation modifications in brain neurons are highly conserved across species in higher mammals, and are thought to play a functionally important role (1). Modulators of lysine 4 in histone H3 (H3K4) methylation are required for memory formation as demonstrated through animal studies, and many of the same modulators are mutated in human dementia (2). Modifiers of H3K4 and lysine 36 in histone H3 (H3K36) have been implicated in neurological or psychiatric cognitive impairment (3). In particular, mutations in H3K4 methylation modifier genes are associated with autosomal dominant diseases (4). Many of these mutations are found in patients with physical abnormalities and intellectual disability, autism, and schizophrenia (SCZ). Modulators of the histone H3 trimethyl lysine 4 (H3K4me3) are associated with the genetic risk architecture of common neurodevelopmental disorders, including SCZ and autism (1) (5) 6). In addition, mutations in *SETD2,* which encodes an H3K36 methyltransferase, are associated with autism spectrum disorders (7, 8).

In the present study, we performed positional cloning and identified a novel mutation in the plant homeodomain (PHD) of *KDM2B*, which encodes an H3K4me3 and H3K36me2 demethylase, in a Japanese family with Marfan syndrome-like minor physical anomalies (MPAs), intellectual disability, and SCZ.

2. Subjects and methods

2.1. Participants

Twelve members of a Japanese family with autosomal dominant inheritance of Marfan syndrome-like MPAs, intellectual disability, and SCZ were participants in this study (Fig. S1 and Fig. 1). All affected individuals were diagnosed independently by at least two experienced psychiatrists according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition criteria. Clinical data were collected by interviews and clinical questionnaires. A hundred unrelated healthy controls and 141 unrelated patients with SCZ were recruited for Sanger sequencing of *KDM2B*. Of these, genomic DNAs from 50 healthy controls and 67 patients with SCZ were used for copy-number variant (CNV) analysis of *KDM2B*. Written informed consent was obtained from each participant. This study was approved according to the guidelines of the Ethics Committee on Life Sciences and Genetic Analysis, Kagoshima University Graduate School of Medical and Dental Sciences and the Ethics Committee on Epidemiological Studies, Kagoshima University.

2.2. Linkage mapping

Genome-wide multipoint linkage analysis was performed for four affected family members and eight unaffected family members (Fig. 1) using the Genome Wide Human SNP Array 6.0 (Affymetrix,

Santa Clara, CA, USA), which features 1.8 million genetic markers, including 908,288 single nucleotide polymorphisms (SNPs) and more than 946,000 comparative genomic hybridization probes. Using the physical location of the single nucleotide variations and the genetic distance correspondence table calculated by the 1000 Genome Project (https://github.com/joepickrell/1000-genomes-genetic-maps), the genetic distances between the SNPs were determined and used at intervals of 0.5 cM. The genotype assignments were determined with the Genome Studio genotyping module software PLINK. Multipoint linkage analysis and reconstruction of the most likely haplotypes were performed using the linkage program MERLIN (9). The allele frequencies of markers, as well as the recombination fractions in males and females, were assumed to be equal. The disease was modeled as an autosomal dominant trait with a disease allele frequency equal to 0.0001. For calculating the logarithm of odds (LOD) scores, the disease penetrance was assumed to be 100%. We extracted 668,795 genotypable SNPs out of the 908,288 SNPs in the Genome-Wide Human SNP Array 6.0 set. For the extracted SNPs, family conflicts were analyzed by PEDSTATS (10), based on the pedigree information. We removed 88 SNPs that were inconsistent with the autosomal dominant inheritance pattern in this family and performed parametric analysis using the remaining 668,707 SNPs for the genome-wide multipoint linkage analysis. On the basis of the "suggestive linkage" proposed by Lander & Kruglyak, we selected linkage regions with LOD scores of 2.1 or higher. (11).

The preliminary haplotype and sequencing analysis revealed a candidate disease mutation: c.2173G>A in *KDM2B* on chromosome 12q, whose proximate region showed the highest LOD score in the genome-wide multipoint linkage analysis. For chromosome 12q, we performed a multipoint linkage analysis using six additional microsatellite markers, D12S321, D12S1721, D12S2073, D12S378 D12S342, and D12S1609, in the following methods.

2.3. Haplotype analysis

Highly polymorphic microsatellite markers were selected from the Polymorphism of Microsatellite Loci in the Japanese Population database (http://www002.upp.so-net.ne.jp/kyama-Q/MS.html). Multipoint linkage analysis and haplotype analysis were performed on chromosome 12q to show genetic evidence of recombination using 31 SNPs from the Genome-Wide Human SNP Array 6.0 (Affymetrix), and the six microsatellite markers above and D12S84 on chromosome 12q24.11 to 12q24.32 in the 12 available family members (Fig. 1). The physical position was verified using the D-Haplo DB-SNP List (Phase III: http://togodb.biosciencedbc.jp). We determined the allele frequencies of the microsatellite markers using 50 unrelated Japanese controls. The microsatellite sequences were amplified with polymerase chain reaction (PCR). Primers were created based on information from the UCSC Genome Browser. The forward primer in each marker set was labeled

with a fluorescent dye at the 5' end. The 12.5 μl PCR mixture contained 1 μl of template DNA (5 ng) and 0.25 μl of each forward and reverse primer (10 μmol), and the DNA region was amplified using TaKaRa Taq DNA polymerase (TaKaRa Bio, Kusatsu, Shiga, Japan) under the following conditions: denaturation at 94 °C for 2 min, 35 cycles at 94 °C for 30 s, 56–62 °C for 30 s and 72 °C for 30 s, followed by a final extension step of 5 min at 72 °C. The amplified products were observed on a 1% GelRed (Biothium Inc., Hayward, CA, USA) pre-stained agarose gel to determine the efficacy of PCR amplification. The PCR products were then denatured and subjected to capillary electrophoresis on an ABI PRISM 3130 Avant Genetic Analyzer (Thermo Fisher Scientific, Waltham, MA, USA). DNA fragment sizes were analyzed relative to Gene Scan 500 RIZ Dye size-standard (Applied Biosystems, Foster City, CA, USA) using the Genemapper software (Applied Biosystems).

2.4. Whole-exome sequencing and Sanger sequencing

Whole-exome sequencing analysis was performed for the proband (IV-9) according to the manufacturers' protocol, which has been previously described (12) (12). In detail, genomic DNA was captured with the SureSelect Target Enrichment System Human All Exon V5 Kit (Agilent Technologies, Santa Clara, CA, USA) and sequenced per lane on an Illumina HiSeq 2000 (Illumina, San Diego, CA, USA) with paired-end 101 bp reads. Image analysis and base calling were

134 performed using sequencing control software with real-time analysis and CASAVA ver.1.8.2
135 software (Illumina). Generated sequence data (FastQ files) were processed using the pipeline with
136 Burrows-Wheeler Aligner (version 0.5.9), Sequence Alignment/Map tools, Picard (version 1.59,
137 http://picard.sourceforge.net/) and GATK32 (version 1.6–5). Variant calls were made using the
138 GATK best practices recommendations. To prioritize variations, we applied several filtering steps.
139 First, we included non-synonymous variations and coding indels. Second, we included variations
140 inherited in an autosomal dominant manner, considering the inheritance pattern of this family. Third,
141 we included variations that existed in the candidate region based on the results of the linkage
142 analysis described above (Fig. 2a). Fourth, we included variations with a depth of alternative alleles
143 greater than 25. Fifth, we included variations with mutant allele frequency less than 0.01 according
144 to the Japanese reference panel project of the Tohoku Medical Megabank (8.3KJPN) and NCBI
145 databases(14).
146 For Sanger sequencing, PCR primers (sequences available upon request) were designed to amplify
147 all coding exons, non-coding RNA exons, 5’ untranslated regions, and intronic regions surrounding
148 each exon. All PCR products were subjected to Sanger sequencing. PCR reactions were performed in
149 a total volume of 12.5 μl containing 10–30 ng of each DNA sample, 0.2 μM each of the forward and
150 reverse oligonucleotide primers, 0.2 mM each of dATP, dGTP, dTTP, and dCTP, and 0.5 units of

TaKaRa Taq. The PCR program was 94 °C for 2 min, 35 cycles of 94 °C for 30 s, 58–64 °C for 30 s and 72 °C for 1 min, followed by 72 °C for 5 min. The amplified PCR products were separated on 1% agarose gels at 100 V for 30 min and were labeled using a BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) with the same primers used in the initial PCR and following the protocol described above. The products were then directly sequenced on an ABI PRISM 3130 Avant Genetic Analyzer (Applied Biosystems).

2.5. CNV analysis using Genome-Wide Human SNP Array

To identify possible pathogenic CNVs, we examined genomic DNA samples from the 12 family members using the Genome Wide Human SNP Array 6.0 (Affymetrix). Labeling and hybridization were performed according to the Affymetrix protocol. Raw data were acquired from the scanner and transferred to a Genotyping Console 3.0.2 (Affymetrix) to determine the genotypes and CNVs from fluorescence intensity data. For all chromosomes, signal intensities for the copy-number predictions of both duplications and deletions were investigated. CNV aberrations of more than 100 kbp were automatically detected by the genotyping console 3.0.2 software. In addition, small CNV aberrations of a few hundred to several thousand bp were visually detected in detail from the raw data.

A detailed CNV analysis was also performed on the genomic region of *KDM2B*. Linkage

disequilibrium maps were created using 1488 SNPs (509 SNPs from the 1000 Genomes Project of the International Genome Sample Resource using 104 Japanese and 979 SNPs from Japanese Multi Omics Reference Panel of the Tohoku Medical Megabank) and using Haploview 4.1 software. To determine CNVs in *KDM2B*, primer pairs and probes were designed against exons 1, 2, 3, 4, 5, 9, 12, 13, 14, and 18–19 of *KDM2B*, according to the LD map. All primers and probes were purchased from Applied Biosystems. We performed CNV analysis of *KDM2B* using TaqMan Copy Number Assays (Thermo Fisher Scientific) on a 7300 real-time PCR system (Thermo Fisher Scientific).

2.6. Lymphoblastoid cells

Peripheral blood mononuclear cells were obtained from five normal controls and the proband. The peripheral blood mononuclear cells were immortalized by an EBV infection according to the protocol of the SRL Medisearch Incorporation and were transformed into lymphoblastoid cell lines (LCLs). LCLs were cultured in RPMI1640 medium (GIBCO, Grand Island, NY, USA) containing 10% FBS (Sigma-Aldrich, St. Louis, MO, USA) and 1% penicillin/streptomycin (Nacalai Tesque, Kyoto, Japan) in a 5% CO_2 humidified incubator at 37 °C. LCLs were seeded at 3.0×10^5 cells. After five days of incubation, cells were harvested for RNA extraction and cDNA synthesis, and for histone extraction.

185

186 **2.7. Real-time quantitative PCR (qPCR) for *KDM2B* mRNA expression**

187 Total RNA was purified from LCLs of five normal controls and the proband using the QIAamp RNA

188 Blood Mini kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. The reverse

189 transcription reaction was performed using SuperScript III (Thermo Fisher Scientific). The cDNA

190 was amplified by real-time PCR using the 7300 Real-time PCR System (Applied Biosystems) with

191 the THUNDERBIRD SYBR qPCR Mix (Toyobo Co., Osaka, Japan). Relative changes in target gene

192 levels were calculated using the ΔΔCt method and normalized to the cDNA levels of *β-actin* and

193 *HPRT*. Triplicated measurements were taken for each sample.

194

195 **2.8. Histone H3K4 tri-methyl quantification**

196 Histones were extracted from pellets of 1.0×10^7 lymphoblastoid cells using the Histone Extraction

197 Kit (ab113476; Abcam, Cambridge, UK) according to the manufacturer's protocol. The protein

198 concentration was adjusted using the Pierce Microplate BCA Protein Assay Kit (Thermo Fisher

199 Scientific) measured on an Infinite M200 Plate Reader (TECAN, Männedorf, Switzerland). The total

200 amount of H3K4me3 in the extracted histone fraction was measured using the Histone H3 (tri-

201 methyl K4) Quantification Kit (ab115057; Abcam) according to the manufacturer's protocol. The

fluorescence (at an excitation wavelength of 530 nm and emission wavelength of 600 nm) was measured using a TriStar LB941 Microplate Reader (Berthold, Osaka, Japan).

3. Results

3.1. Patients information and clinical symptoms

The proband (Fig. Sa–h, IV-9 in Fig. 1) was a 32-year-old woman. She had poor grades in elementary and middle school. She was admitted to a psychiatric hospital three times between the ages of 17 and 19 with a diagnosis of SCZ. She continuously showed wandering behaviors, silly smiles, and soliloquies. She developed urinary incontinence at the age of 21. Because of her social withdrawal, tendency to be bedridden, and her deteriorating nutritional status, she was admitted to another psychiatric hospital at the age of 25. At the beginning of the hospitalization, she developed catatonia syndrome, including stupor, rejection, echolalia, incoherence, auditory hallucinations, misidentification of persons, soliloquies, and insomnia. At 31 years old, her intelligence quotient score was 39 on the Suzuki-Binet intelligence scale. Fig. S1a–h shows her Marfan syndrome-like MPAs, including epicanthus, retrognathia, pectus excavatum, high arched palate with crowding of teeth, thumb sign (entire thumbnail protrudes beyond the ulnar border of hand), wrist sign (thumb and fifth finger overlap when encircling the wrist), plain flat foot, and medial displacement of the medial malleolus. Her arm span and height were 175 cm and 165 cm, respectively, resulting in a 1.06

arm span to height ratio. The revised Ghent criteria for Marfan syndrome require a systemic score of 7 or more for systemic findings (15). In the proband, we estimated her systemic score was at least 6 points. This close, but insufficient to meet diagnostic systemic score has been due to lack of information about the length of the lower segment, which is defined as the distance from the top of the symphysis pubis to the floor in the standing position. A systolic murmur and incomplete right bundle branch block were found, although the echocardiograph showed no abnormalities. A chromosomal study with G-banding showed a normal karyotype of 46, XX.

The proband's mother (III-4 in Fig. 1) was a 69-year-old woman. She worked as a dressmaker after graduating from elementary school. After the birth of her fifth baby at age 29, she began to talk incoherently and spent two months in a psychiatric hospital. After discharge, she continued with outpatient visits until she was 38 years old. Then, she discontinued her hospital visits. When we visited her home during the investigation, she was 69 years old and presented with mutism, echolalia, and soliloquies. She also showed Marfan-syndrome like MPAs, including mild pectus excavatum, mild high arched palate and medial displacement of the medial malleolus (data not shown). Her arm span and height were 159 cm and 149 cm, respectively, resulting in a 1.06 arm span to height ratio.

The proband's older sister (Fig. S1i–o, IV-5 in Fig. 1) was a 39-year-old woman. She attended a

resource room class in middle school because of her intellectual disability. She was employed in a spinning factory after graduating from middle school. At age 22, she already exhibited soliloquies and silly smiles. She was unable to communicate appropriately and showed incoherent behavior. She was admitted to a psychiatric hospital for SCZ on at least ten separate occasions. She occasionally showed epileptic seizures. She also presented similar MPAs to the proband. Fig. S1i–o shows her MPAs, including retrognathia, pectus excavatum, mild high arched palate with crowding of teeth, wrist sign, plain flat foot, and medial displacement of the medial malleolus. Her arm span and height were 158 cm and 152.5 cm, respectively, resulting in a 1.04 arm span to height ratio.

The proband's older brother (IV-8 in Fig. 1) was a 34-year-old man. He was born with cerebral paresis with unsmooth foot movements and left blindness. His left eye was obscured by an opaque cornea. He started to walk and talk around the age of five. He moved to a school for handicapped children at the age of nine, and graduated from a middle school for handicapped children. He had never worked. His intelligence quotient score was 35 on the Tanaka-Binet intelligence scale at the age of 20. He did not present with any hallucinations or delusions. He spoke in a high-pitched stammering voice using only simple words. He showed similar MPAs to his mother (III-4) and two sisters (IV-5, -9). He showed pectus carinatum, medial displacement of the medial malleolus, and high arched palate (data not shown). His arm span and height were 175 cm and 160 cm, respectively,

253 resulting in a 1.05 arm span to height ratio.

254 The proband's older sisters (IV-1, -3) were 45-year-old and 42-year-old women, respectively. They

255 suffered a transient psychotic episode in their 20s. Thereafter, they have remained in remission. They

256 did not show a series of Marfan-like MPAs.

257

258 **3.2. Linkage mapping**

259 We performed a genome-wide scan using the DNA arrays to confirm the evidence for linkage.

260 Multipoint analysis using the dominant and parametric models at 100% penetration resulted in two

261 potential linkage regions on chromosomes 12q24.23–q24.31 and 17p11.2–p12 (Fig. 2a). Maximum

262 multipoint LOD scores of 2.414 ($\theta = 0.00$) were obtained in both regions. The multipoint linkage

263 analysis with the addition of microsatellite markers as well as SNPs yielded consistently high LOD

264 scores of 2.414 at around D12S2073 on chromosome 12p (Fig. 2a).

265 **3.3. Haplotype analysis**

266 Haplotype analysis using SNPs and additional microsatellite markers analysis revealed a common

267 haplotype delimited by D12S84 and D12S1609 containing *KDM2B* on chromosome 12q24.11–

268 12q24.32, which segregated in all the available patients (Fig. 1). A minor allele in D12S2073 was

269 co-separated only in patients and not found in 50 controls, suggesting that the minor allele was

strongly associated with linkage disequilibrium of the disease-causing mutation. The proximal recombination site was between rs12318877 and rs12369523, and the distal site was between rs12369523 and D12S1609.

3.4. Whole-exome sequencing and co-segregation analysis

A region of approximately 84.61 Mb was captured from the proband (IV-9) exome. The average reading depth was 169.6, and about 97.7% of the target area was covered by more than 10 reads. In IV-9, 21,008 mutations including 9373 nonsynonymous SNPs and 460 indels were detected in the coding region. Filtering was performed on the detected mutations, as described in the Materials and Methods section. In total, 65 mutations were under the LOD elevated regions on chromosomes 12 and 17 (Fig. 2b). Subsequently, four mutations in *KCN12* on chromosome 17 and one mutation in *KDM2B* on chromosome 12q were found in the filtering conditions with a depth of more than 25 and with a minor allele frequency of less than 0.001. However, the subsequent Sanger sequencing for *KCN12* revealed that no mutations were present. A coding strand heterozygous mutation in *KDM2B* (NM_032590.4: rs1555289644: exon15, c.2173G>A, p.A725T) was re-confirmed by Sanger sequencing in the proband (IV-9), her mother (III-4), and her two siblings (IV-5, -8) (Fig. 3a,b). The MPAs and psychiatric symptoms of these patients were more severe than those of other family

members not carrying c.2173G>A mutation.

Additional Sanger sequencing revealed that the c.2173G>A mutation was not detected in 141 patients with SCZ or 50 healthy controls. According to 8.3KJPN, this is an extremely rare mutation with a frequency of 0.0004. The former version, 3.5KJPN database, provided clinical information on one individual with this mutation by self-reported questionnaire data but no information on psychiatric symptoms or MPAs was available (14). This mutation was not found in The Genome Aggregation Database (gnomAD). Additionally, in *KDM2B*, Sanger sequencing for all 29 coding exons and exon-intron boundaries was performed for 50 patients with SCZ. Another *KDM2B* missense mutation (exon 19 c.3007G>A, p.G1003S, NM_032590.4) was identified in one of the 100 alleles.

We evaluated the p.A725T and p.G1003S mutations in *KDM2B* using the following five *in silico* tools: poyphen2, SHIFT, PANTHER, PROVEAN and Mutation Taster.

The functional predictions for p.A725T were “probably damaging (1.0)” by PolyPhen2, “tolerated (0.45)” by SIFT, “NA” by PANTHER, “deleterious (−3.458)” by PROVEAN, and “disease causing” by Mutation Taster. For the second *KDM2B* mutation, the predictions for p.G1003S were defined as “possibly damaging (0.466)” by PolyPhen2, “tolerated (0.99)” by SIFT, “NA” by PANTHER, “deleterious (−4.508)” by PROVEAN, and as a “polymorphism” by Mutation Taster.

3.5. CNV analysis

In all chromosomes, the signal intensity of both the replication and deletion copy-number predictions was examined in detail from the raw data. However, no specific CNVs were found in affected individuals. Detailed CNV analysis using real-time qPCR of *KDM2B* showed no duplicates or deletions.

3.6. Mutant *KDM2B* mRNA expression level

In LCLs from the proband with a heterozygous c.2173G>A mutation, the expression level of *KDM2B* mRNA was approximately halved (N = 1, n = 6 technical replicates, mean ± 95% confidential interval [CI] = 0.75 ± 0.20) in comparison with that from normal controls (N = 5, n = 6 technical replicates, mean ± 95% CI = 1.39 ± 0.11) (Fig. 4a). On the other hand, in LCLs from the patient with a heterozygous c.3007G>A mutation (N = 1, n = 6 technical replicates, mean ± 95% CI = 1.55 ± 0.23), there was no significant difference in comparison with that from normal controls (Fig. 4a). Each normalization was performed with *HPRT*. The expression level of *KDM2B* mRNA in the proband was also halved when normalized with *β-actin* (data not shown).

3.7. The *KDM2B* mutation has no effect on histone H3K4 methylation

The quantification of the amount of trimethylated H3K4 extracted from LCLs showed no significant difference between the proband, the patient with heterozygous c.3007G>A, and controls (Fig. 4b).

4. Discussion

KDM2B plays an important role as a histone lysine demethylase that removes methyl from H3K36me2 and H3K4me3(16). The roles of KDM2B are associated with normal cellular processes such as cell senescence, cell differentiation, and stem cell self-renewal (17). In addition, it has been reported that KDM2B is associated with embryonic neural development through the regulation of cell proliferation and cell death (18). In the present study, in a Japanese family with autosomal-dominant Marfan syndrome-like MPAs, intellectual disability, and SCZ, although with the limitation of only two generations of linkage analysis, we identified an extremely rare genetic coding strand mutation (rs1555289644: c.2173G>A, p.A725T) in *KDM2B*. It was detected using complementary positional cloning methods, validated by Sanger sequencing, and rarely found in either large sequencing databases or in the Sanger sequences of 100 alleles of controls. When considering the c.2173G>A mutation in *KDM2B* as a risk for Marfan syndrome-like MPAs, intellectual disability, and SCZ; three platforms (polyphen2, PROVEAN, and Mutation Taster) defined the mutation as

338 “probably damaging”, “deleterious”, and “disease causing”, respectively. The c.2173G>A mutation
339 is in the functionally important PHD domain of *KDM2B* (Fig. 3b), which specifically recognizes
340 H3K4me3 and binds to histones H3 and DNA (16,17). In addition, our expression analysis revealed
341 that *KDM2B* mRNA of LCLs from the proband with c.2173G>A was expressed at half the level of
342 controls, suggesting a loss-of-function mutation. In silico, the Residual Variation Intolerance Score
343 (RVIS) assesses an intolerance to loss-of-function mutation. The RVIS of *KDM2B* was as low as
344 −2.17, which was lower than the average value (RVIS = −0.56) of the neurodevelopmental disorder
345 group (21), suggesting that a high level of harm could be caused by haploinsufficiency.
346 There have been several reports of *KDM2B* mutations causing neurodevelopmental disorders. The de
347 novo mutation p.G745S causes SCZ (22). A microdeletion of 12q24.31, including *KDM2B*, results in
348 intellectual disability, autism, and epilepsy (23). A homozygous p.R1017H mutation causes
349 developmental delay, microcephaly, hypotonia, and neonatal convulsion (24). The results of these
350 previous studies and the present study suggest that functional mutations of *KDM2B* are likely to
351 cause SCZ, other neurodevelopmental disorders, and skeletal abnormalities.
352 In the *KDM2B* gene knockout mouse, neural tube defect and severe developmental delays were
353 observed (25). KDM2B binds to nonmethylated DNA via a ZF-CxxC DNA binding domain (Fig. 3b)
354 (26). Mice with a heterozygous deletion of the CxxC domain in Kdm2b exhibited skeletal alterations

355 with homeotic transformations in the cervical to sacral regions (27). In the present study, affected
356 patients who possibly had half the typical expression level of KDM2B, a situation that can be
357 considered a genetic dose-equivalent to a heterozygous deletion of the CxxC domain, also exhibited
358 skeletal abnormalities with Marfan syndrome-like MPAs. In the present study, we also found an
359 extremely rare c.3007G>A mutation in a patient with SCZ who showed no Marfan-like MPAs.
360 Although no difference was detected in *KDM2B* expression, our in silico analysis using polyphen2
361 and PROVEAN revealed that the mutation was “possibly damaging” and “deleterious”, respectively,
362 suggesting it as a possible disease-causing mutation for SCZ.
363 Regulation of H3K4me3 in the central nervous system is associated with common
364 neurodevelopmental genetic risk structures, including SCZ and autism (1). Approximately 120
365 H3K4me3 peaks are completely conserved between human postmortem brains, non-human primates,
366 and the mouse prefrontal cortex. Abnormal H3K4 methylation patterns, including the width of the
367 H3K4me3 peaks, have been also implicated in other complex neurological and psychiatric disorders.
368 For example, genes with broad domains of H3K4me3 are preferentially down-regulated in a mouse
369 model of Huntington's disease (28). Abnormal excess spreading of H3K4me3 from the transcription
370 start sites into the gene bodies and upstream promoters has been shown to occur in some cases of
371 autism spectrum disorder (6).

372 Embryos with homozygous mutations in *NSD1,* which has an intrinsic histone methyltransferase
373 activity specific for histone H3K36 and histone H4K20, have a high incidence of apoptosis upon
374 initiation of mesoderm formation and are unable to complete gastrulation (29). KDM4A/C is
375 essential for selective H3K36 demethylation and loss of RNA polymerase II mobilization in the
376 transcribed region of the astrocyte trait gene *GFAP*, and suppresses astrocyte differentiation in
377 parallel by controlling polymerase extension (30).
378 There have been a number of reports of diseases associated with neuropsychiatric disorders and
379 MPAs due to mutations in histone H3K4 and H3K36 modulators. These include Kleefstra syndrome
380 2, which is caused by mutations in *KMT2C*; Kabuki syndrome, caused by mutations in *KMT2D;* and
381 other diseases involving MPAs and neurodevelopmental disorders, occasionally with additional
382 comorbid medical conditions such as motor dysfunction and epilepsy, caused by mutations in other
383 H3K4 methyltransferase associated genes (28, 29, 30, 31). By analyzing the whole-exome sequences
384 of 4264 schizophrenia cases, 9343 controls and 1077 trios, loss-of-function variants in *KMT2F,* also
385 known as *SETD1A* and which produces mono-, di-, and trimethylated histone H3 at Lys 4, were
386 identified in patients with SCZ and developmental disorders (35). *Nkx2-5*, a histone H3K36 tri-
387 methyltransferase, is associated with Wolf-Hirschhorn syndrome, which causes the characteristic
388 facial malformations known as "Greek warrior helmet", moderate to severe intellectual disability,

389 and growth retardation (36).

390 There have also been a number of reports of diseases associated with neuropsychiatric disorders and

391 MPAs due to mutations of histone demethylases. In 2005, X-linked Claes-Jensen syndrome was

392 determined to be caused by a mutation in *KDM5C,* a histone demethylase. Interestingly, there was a

393 patient with a hemizygous mutation in *KDM5C* who lacked the behavioral and facial features of

394 Claes-Jensen syndrome and was instead diagnosed with autism spectrum disorder (34, 35). Cleft

395 palate, psychomotor retardation, and distinctive facial features syndrome are caused by *KDM1A*

396 mutations (29, 36, 37). *KDM5B* mutations have also been associated with both autism and

397 intellectual disability (2, 38, 39). Recently, haploinsufficiency has been predicted using %HI, with

398 lower values being more likely to produce haploinsufficiency (40). The %HI of *KDM2B* by

399 DECIPHER is 18.61 (https://decipher.sanger.ac.uk/browser). Interestingly, in many diseases, which

400 are caused by haploinsufficiency or suspected haploinsufficiency of genes belonging to the KDM or

401 KTM families, the %HI of each disease-causing gene is not consistently low, but ranges from 5.56 to

402 53.58. For example, Kabuki syndrome and Kleefstra syndrome 2 are caused by haploinsufficiency of

403 the *KMT2D* and *KMT2C*, whose %HI are 17.49 and 53.58, respectively.

404 Although the experimental material with c.2173G>A mutation was derived only one patient, using

405 patient derived LCLs, we demonstrated that *KDM2B* variants are not associated with alterations in

406 the overall levels of H3K4 methylation. Haploinsufficiency for other KDM family members has not
407 been convincingly shown to affect global H3K4 methylation levels. Compensation due to other
408 KDM family members or the fact that samples could not be obtained during the developmental stage
409 or cell types where the methylation may be most relevant could be the possible cause of the lack of
410 methylation differences. The functional mutation in *KDM2B* in the present study is possibly disease-
411 causing, and may be responsible for the Marfan syndrome-like MPAs, intellectual disability, and
412 SCZ observed, although further studies are required to evaluate the effects of the *KDM2B* point
413 mutation during development and to assess its functions in the brain.
414 Taken together with a number of previous results, genetic mutations involved in the demethylation or
415 methylation of histone H3K4 and/or H3K36 often result in neuropsychiatric disorders, including
416 intellectual disabilities, developmental disorders, and/or SCZ with or without MPAs by dominant
417 inheritance due to haploinsufficiency. The familial patients with the functional *KDM2B* mutation
418 found in the present study may be an example of such a group of cases of intellectual disability,
419 developmental disability, and schizophrenia with MPAs.
420
421 **Acknowledgments**
422 The authors thank all involved patients and their families for their participation. The authors thank
423 Ms. Hiwatashi, Ms. Meguro, Ms. Nishimura, Ms. Shimomura and Ms. Yokoyama for their technical

assistance. We thank Dr. Kasamo and Dr. Urata for their technical advisees. We also thank Dr. Ishida for editing this manuscript. We would like to thank Uni-edit (https://uni-edit.net/) for editing and proofreading this manuscript. This work was supported by MEXT KAKENHI Grant Number JP15K09811.

References

1. Dincer A, Gavin DP, Xu K, Zhang B, Dudley JT, Schadt EE, et al. Deciphering H3K4me3 broad domains associated with gene-regulatory networks and conserved epigenomic landscapes in the human brain. Transl Psychiatry. 2015;5:e679–e679.

2. Collins BE, Greer CB, Coleman BC, Sweatt JD. Histone H3 lysine K4 methylation and its role in learning and memory. Epigenetics Chromatin. 2019;12:7.

3. Zaghi M, Broccoli V, Sessa A. H3K36 Methylation in Neural Development and Associated Diseases. Front Genet. 2020;10:1–9.

4. Faundes V, Newman WG, Bernardini L, Canham N, Clayton-Smith J, Dallapiccola B, et al. Histone Lysine Methylases and Demethylases in the Landscape of Human Developmental Disorders. Am J Hum Genet. 2018;102:175–87.

5. Huang HS, Matevossian A, Whittle C, Se YK, Schumacher A, Baker SP, et al. Prefrontal dysfunction in schizophrenia involves mixed-lineage leukemia 1-regulated histone methylation at GABAergic gene promoters. J Neurosci. 2007;27:11254–62.

6. Shulha HP, Cheung I, Whittle C, Wang J, Virgil D, Lin CL, et al. Epigenetic signatures of autism: Trimethylated H3K4 landscapes in prefrontal neurons. Arch Gen Psychiatry. 2012;69:314–24.

7. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014;515:216–21.

8. Lumish HS, Wynn J, Devinsky O, Chung WK. Brief Report: SETD2 Mutation in a Child with Autism, Intellectual Disabilities and Epilepsy. J Autism Dev Disord. 2015;45:3764–70.

9. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin — Rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet. 2002;30:97–101.

10. Wigginton JE, Abecasis GR. PEDSTATS: Descriptive statistics, graphics and quality assessment for gene mapping data. Bioinformatics. 2005;21:3445–7.

11. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet. 1995;11:241–7.

12. Kasamo K, Nakamura M, Daimou Y, Sano A. A PRIMPOL mutation and variants in multiple genes may contribute to phenotypes in a familial case with chronic progressive external ophthalmoplegia symptoms. Neurosci Res. 2020;157:58–63.

13. Hiromi U, Masayuki N, Mio N, Kato Y, Shu-ichi U, Akira S. Positional cloning and comprehensive mutation analysis of a Japanese family with lithium-responsive bipolar disorder identifies a novel DOCK5 mutation. J Hum Genet. 2020;

14. Tadaka S, Saigusa D, Motoike IN, Inoue J, Aoki Y, Shirota M, et al. jMorp: Japanese Multi

Omics Reference Panel. Nucleic Acids Res. 2018 Jan 4;46:D551–7.

15. Loeys BL, Dietz HC, Braverman AC, Callewaert BL, De Backer J, Devereux RB, et al. The revised Ghent nosology for the Marfan syndrome. J Med Genet. 2010;47(7):476–85.

16. Tsukada Y-I, Fang J, Erdjument-Bromage H, Warren ME, Borchers CH, Tempst P, et al. Histone demethylation by a family of JmjC domain-containing proteins. Nature. 2006;439:811–6.

17. Yan M, Yang X, Wang H, Shao Q. The critical role of histone lysine demethylase KDM2B in cancer. Am J Transl Res. 2018;10:2222–33.

18. Fukuda T, Tokunaga A, Sakamoto R, Yoshida N. Molecular and Cellular Neuroscience Fbxl10 / Kdm2b de fi ciency accelerates neural progenitor cell death and leads to exencephaly. Mol Cell Neurosci. 2011;46:614–24.

19. Peña P V., Davrazou F, Shi X, Walter KL, Verkhusha V V., Gozani O, et al. Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. Nature. 2006;442:100–3.

20. Li H, Ilin S, Wang W, Duncan EM, Wysocka J, Allis CD, et al. Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature. 2006;442:91–5.

21. Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic Intolerance to Functional Variation and the Interpretation of Personal Genomes. Williams SM, editor. PLoS Genet. 2013 Aug 22;9:e1003709.

22. Girard SL, Gauthier J, Noreau A, Xiong L, Zhou S, Jouan L, et al. Increased exonic de novo mutation rate in individuals with schizophrenia. Nat Genet. 2011;43:860–3.

23. Labonne JDJ, Han K, Shigeki L, Il I, Kong K, Diamond MP, et al. An atypical 12q24 . 31 microdeletion implicates six genes including a histone demethylase KDM2B and a histone methyltransferase SETD1B in syndromic intellectual disability. Hum Genet. 2016;135:757–71.

24. Charng W-L, Karaca E, Coban Akdemir Z, Gambin T, Atik MM, Gu S, et al. Exome sequencing in mostly consanguineous Arab families with neurologic disease provides a high potential molecular diagnosis rate. BMC Med Genomics. 2016;9:42.

25. Salminen A, Kaarniranta K, Kauppinen A. Hypoxia-Inducible Histone Lysine Demethylases: Impact on the Aging Process and Age-Related Diseases. Aging Dis. 2016;7(2):180–200.

26. Long HK, Blackledge NP, Klose RJ. ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. Biochem Soc Trans. 2013;41(3):727–40.

27. Blackledge NP, Farcas AM, Kondo T, King HW, McGouran JF, Hanssen LLP, et al. Variant

PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. Cell. 2014;157:1445–59.

28. Vashishtha M, Ng CW, Yildirim F, Gipson TA, Kratter IH, Bodai L, et al. Targeting H3K4 trimethylation in Huntington disease. Proc Natl Acad Sci. 2013;110:E3027–36.

29. Rayasam GV, Wendling O, Angrand PO, Mark M, Niederreither K, Song L, et al. NSD1 is essential for early post-implantation development and has a catalytically active SET domain. EMBO J. 2003;22:3153–63.

30. Cascante A, Klum S, Biswas M, Antolin-fontes B, Barnabé-heider F, Hermanson O. Gene-Specific Methylation Control of H3K9 and H3K36 on Neurotrophic BDNF versus Astroglial GFAP Genes by KDM4A/ C Regulates Neural Stem Cell Differentiation. J Mol Biol. 2014;426:3467–77.

31. Koemans TS, Kleefstra T, Chubak MC, Stone MH, Reijnders MRF, de Munnik S, et al. Functional convergence of histone methyltransferases EHMT1 and KMT2C involved in intellectual disability and autism spectrum disorder. FitzPatrick DR, editor. PLOS Genet. 2017 Oct 25;13(10):e1006864.

32. Tunovic S, Barkovich J, Sherr EH, Slavotinek AM. De novo ANKRD11 and KDM1A gene mutations in a male with features of KBG syndrome and Kabuki syndrome. Am J Med Genet

Part A. 2014;164:1744–9.

33. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet. 2010;42:790–3.

34. Miyake N, Koshimizu E, Okamoto N, Mizuno S, Ogata T, Nagai T, et al. MLL2 and KDM6A mutations in patients with Kabuki syndrome. Am J Med Genet Part A. 2013;161:2234–43.

35. Singh T, Kurki MI, Curtis D, Purcell SM, Crooks L, McRae J, et al. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. Nat Neurosci. 2016;19:571–7.

36. Nimura K, Ura K, Shiratori H, Ikawa M, Okabe M, Schwartz RJ, et al. LETTERS A histone H3 lysine 36 trimethyltransferase links Nkx2-5 to Wolf – Hirschhorn syndrome. Nature. 2009;460:287–91.

37. Iwase S, Lan F, Bayliss P, de la Torre-Ubieta L, Huarte M, Qi HH, et al. The X-Linked Mental Retardation Gene SMCX/JARID1C Defines a Family of Histone H3 Lysine 4 Demethylases. Cell. 2007;128:1077–88.

38. Abidi FE, Holloway L, Moore CA, Weaver DD, Simensen RJ, Stevenson RE, et al. Mutations in JARID1C are associated with X-linked mental retardation, short stature and hyperreflexia. J

Med Genet. 2008 Jul 8;45:787–93.

39. Chong JX, Yu JH, Lorentzen P, Park KM, Jamal SM, Tabor HK, et al. Gene discovery for Mendelian conditions via social networking: De novo variants in KDM1A cause developmental delay and distinctive facial features. Genet Med. 2016;18:788–95.

40. Huang N, Lee I, Marcotte EM, Hurles ME. Characterising and Predicting Haploinsufficiency in the Human Genome. Schierup MH, editor. PLoS Genet. 2010 Oct 14;6:e1001154.

Figure legends

Figure 1

Pedigree and haplotypes of a Japanese family with minor malformations, mental retardation, and schizophrenia. An arrow indicates the proband. Blackened symbols at the upper right, lower right, lower left, and upper left indicate symptoms of schizophrenia, medial displacement of the medial malleolus, arm span/height ratio > 1.03, and pectus excavatum, respectively. Grayed symbol at the upper right indicates schizoaffective disorder. Black dots indicate individuals who were sampled and analyzed. The haplotypes are shown under each individual, with the disease haplotype boxed. The proximal recombination site was between rs12318877 and rs12369523, and the distal site was between rs12369523 and D12S1609.

Figure 2

The multipoint logarithm of odds (LOD) scores for genetic locations on chromosome 12 and 17 and filtering steps. **a** The x-axis represents the genetic distance between the markers, and the y-axis represents the LOD score. The maximum multipoint LOD scores of 2.414 ($\theta = 0.00$) were found on chromosomes 12q24.23–q24.31 and 17p11.2–p12. **b** Filtering steps applied to variations called from the whole exome sequencing of the proband (VI-9).

Figure 3

The results of DNA Sanger sequencing and a schematic representation of the domain architecture of the KDM2B protein. **a** The DNA Sanger sequence shows a heterozygous c.2173G>A nucleotide change (arrow) in exon 15 of *KDM2B*, which leads to the replacement of alanine (GCC) with threonine (ACC) at codon 725 (p.A725T). **b** The domain architecture of the KDM2B protein is shown in a schematic figure. The box and arrow indicate the position of the missense change p.A725T. Functional domains were identified using the UniProtKB site (https://www.uniprot.org/uniprot/Q8NHM5#family_and_domains).

Figure 4

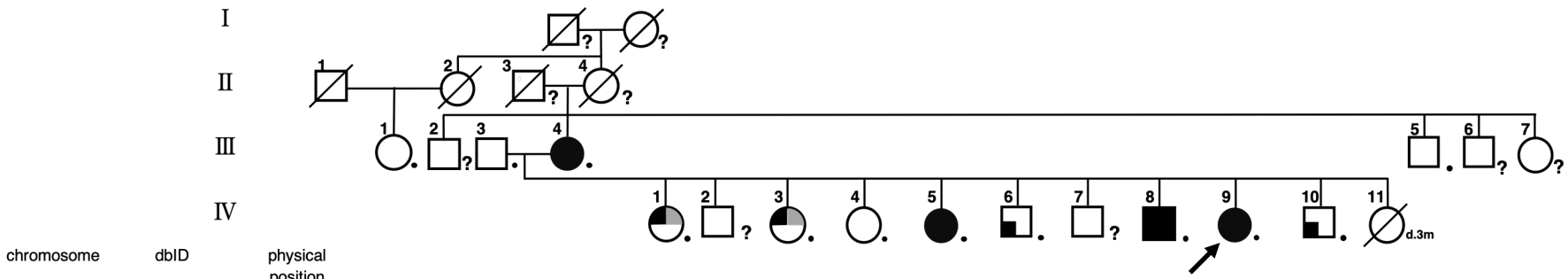
KDM2B expression and effects on histone H3K4 methylation. **a** RT–qPCR indicates that the proband with a heterozygous c.2173G>A *KDM2B* mutation has decreased lymphoblastoid mRNA levels of *KDM2B* when compared to controls (n = 5) (mean ± 95% confidential interval [CI] of controls = 1.39 ± 0.11; mean ± 95% CI of the proband = 0.75 ± 0.20). The patient with a heterozygous c.3007G>A mutation had similar lymphoblastoid mRNA levels of *KDM2B* to controls (mean ± 95% CI of the patient = 1.55 ± 0.23). **b** Histone methylation was assayed independently in the proband sample taken

from each patient-derived lymphoblastoid cell line on different days and compared with control cell lines (n = 5). Methylation values were normalized to pan histone H3 levels. The bar graphs show the means with error bars showing ± 95% CI. Differences between controls and patient-derived samples are not significant (H3K4me3: mean ± 95% CI of controls = 1.48 ± 0.12; mean ± 95% CI of the proband with heterozygous c.2173G>A = 1.26 ± 0.47; mean ± 95% CI of the patient with heterozygous c.3007G>A mutation = 1.44 ± 0.17)

Figure S1

Marfan syndrome-like minor physical anomalies of IV-9 (proband) and IV-5. The proband exhibited **a** long upper limbs (in a whole-body lateral view), **b** epicanthus, **c** retrognathia, **d** pectoris excavatum, **e** high-arched palate, **f** wrist sign and thumb signs, **g** flatfoot, and **h** medial displacement of the medial malleolus. Patient IV-5 showed **i** long upper limbs (in a whole-body lateral view), **j** retrognathia, **k** pectoris excavatum, **l** high-arched palate, **m** wrist sign and thumb signs, **n** flatfoot, and **o** medial displacement of the medial malleolus.

Figure 1



chromosome	dbID	physical position																					
12q24.11	**D12S84**	107546196	10	10	8	9	10	3	10	9	10	8	10	8	10	8	3	8	3	8	10	8	10	9	10	11
	rs3741992	111001020	A	B	A	B	A	A	A	A	A	A	A	A	A	B	A	A	A	A	A	A	A	B	A	B
	rs2264886	112010845	A	A	A	A	A	B	A	A	A	A	A	A	A	A	A	A	B	A	A	A	A	A	A	A
	rs1598978	113008816	A	B	B	B	B	A	B	B	B	B	B	B	B	B	B	B	A	B	B	B	B	B	B	B
	rs2384568	114043120	B	A	A	A	A	B	A	A	A	A	A	A	A	A	A	A	B	A	A	A	A	A	B	A
	rs11612304	115002087	B	B	A	A	B	A	B	A	B	A	B	A	B	A	A	A	A	A	B	A	B	A	B	B
	rs904660	116042182	B	A	A	A	B	A	B	A	B	A	B	A	B	A	A	A	A	A	B	A	B	A	B	B
	rs11068692	116753774	A	A	A	A	A	B	B	A	B	A	B	A	B	A	A	A	A	A	A	A	B	A	B	A
	rs12318877	116766164	A	A	A	A	A	B	B	A	B	A	B	A	B	A	A	A	A	A	A	A	B	A	B	A
	rs12369523	116793391	B	A	A	B	B	B	B	B	B	B	B	B	B	A	B	B	B	A	B	A	B	A	B	B
	rs11068933	117322508	B	A	A	A	A	B	B	A	B	A	B	A	A	A	B	A	A	A	A	A	B	A	B	A
	rs12316703	117324840	B	B	B	B	A	B	B	B	B	B	B	B	A	B	B	B	A	B	A	B	B	B	B	B
	rs11068937	117341487	B	A	A	A	A	B	B	A	B	A	B	A	A	A	B	A	A	A	A	A	B	A	B	A
12q24.23	**D12S321**	118631762	3	2	3	4	3	3	3	4	3	4	3	4	3	3	3	4	3	3	3	3	3	3	3	3
12q24.31	**D12S1721**	119170517			6	10	7	8	8	6	8	6	8	6	7	10	8	6	7	10	7	10	8	10	6	10
	rs9657942	119245175	A	A	B	B	A	A	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	A
	rs2859287	119305722	B	A	A	A	B	B	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A
	rs787820	119410702	A	A	A	A	B	A	A	A	A	A	A	A	B	A	A	A	B	A	B	A	A	A	A	A
	rs674240	119533318	B	A	A	A	B	B	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	B
	rs509152	119670932	A	A	B	A	A	A	A	A	A	A	A	A	A	B	A	B	A	B	A	B	A	B	A	A
	D12S2073	119767662	5	6	6	5	**1**	5	5	5	5	5	5	5	**1**	6	5	5	**1**	6	**1**	6	5	6	5	6
	rs1794899	120027356	A	B	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
	rs1696326	121877934	A	A	A	B	B	B	B	B	B	B	B	B	B	A	B	B	B	A	B	A	B	A	B	A
	rs897393	121878004	B	B	B	A	A	A	A	A	A	A	A	A	A	B	A	A	A	B	A	B	A	B	A	B
***KDM2B* mutation**	**rs1555289644**	**121882270**	A	A	A	A	**B**	A	A	A	A	A	A	A	**B**	A	A	A	**B**	A	**B**	A	A	A	A	A
	rs921808	121889088	B	B	B	A	A	A	A	A	A	A	A	A	A	B	A	A	A	B	A	B	A	B	A	B
	rs12313006	121908838	A	A	B	A	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	A	B	A
	rs4275659	122013881	A	B	B	A	A	A	A	A	A	A	A	A	A	B	A	A	A	B	A	B	A	B	A	A
	rs3759111	122253727	B	B	A	B	B	B	B	B	B	B	B	B	B	A	B	B	B	A	B	A	B	A	B	B
	rs1609520	122301890	A	A	B	A	A	A	A	A	A	A	A	A	A	B	A	A	A	B	A	B	A	B	A	A
	rs7139193	122737129	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
	rs12371640	123202281	B	B	A	A	A	B	B	A	B	A	B	A	A	A	B	A	A	A	A	A	B	A	B	B
12q24.31	**D12S378**	123228943	4	5	5	5	5	4	4	5	4	5	4	5	5	5	4	5	5	5	5	5	4	5	4	5
12q24.31	**D12S342**	124449793	5	9	6	4	5	8	8	6	4	8	4	8	5	6	4	8	5	6	5	6	8	6	8	10
	rs2270141	125192685	B	B	B	B	A	A	A	B	A	B	A	B	B	B	A	B	B	B	B	B	A	B	A	B
	rs882393	127601677	A	A	A	A	A	B	A	A	B	A	B	A	A	A	B	A	A	A	A	A	B	A	B	A
	rs4882806	127649609	B	B	A	A	B	A	B	A	A	A	A	A	B	A	B	A	B	A	B	A	A	A	B	A
12q24.32	**D12S1609**	127617802	12	7	7	7	12	2	12	7	2	7	2	7	12	7	2	7	12	7	12	7	2	7	2	12
	rs11060081	128077685	B	A	A	A	B	A	A	A	B	A	B	A	A	A	A	A	A	A	A	A	B	A	B	A

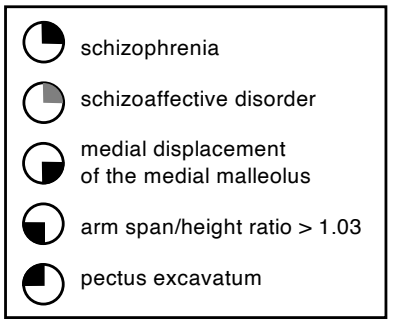
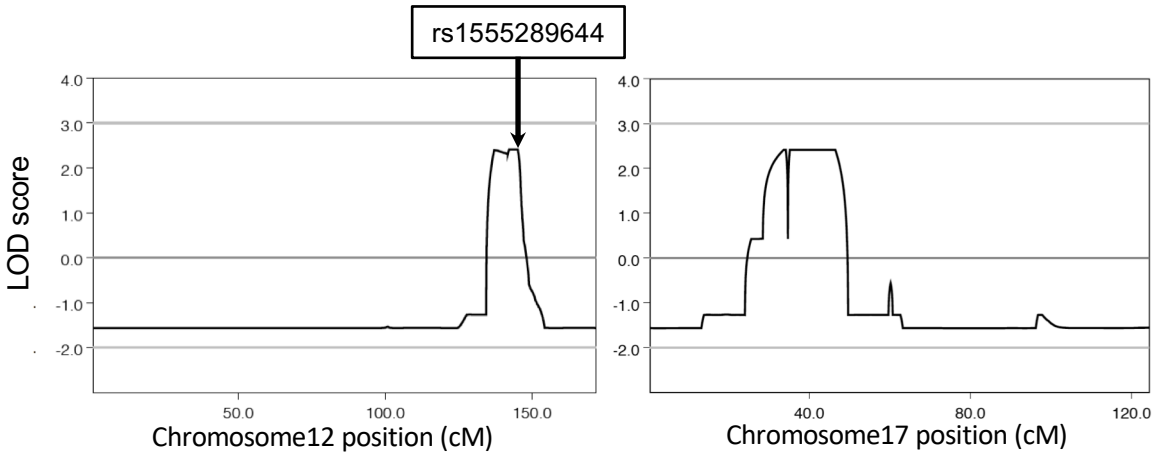


Figure 2

A


rs1555289644
LOD score
4.0
3.0
2.0
1.0
0.0
-1.0
-2.0
50.0
100.0
150.0
Chromosome12 position (cM)
40.0
80.0
120.0
Chromosome17 position (cM)

B

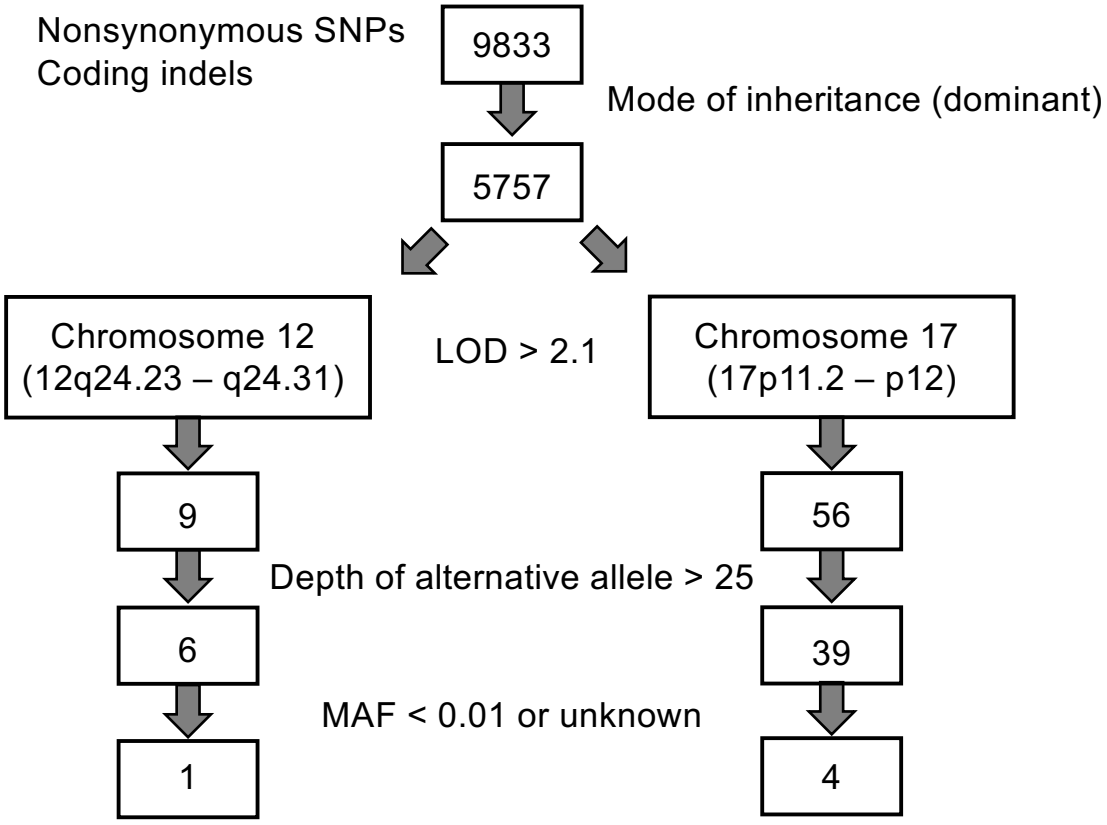

Nonsynonymous SNPs
Coding indels
9833
Mode of inheritance (dominant)
5757
Chromosome 12
(12q24.23 – q24.31)
LOD > 2.1
Chromosome 17
(17p11.2 – p12)
9
56
Depth of alternative allele > 25
6
39
MAF < 0.01 or unknown
1
4

Figure 3

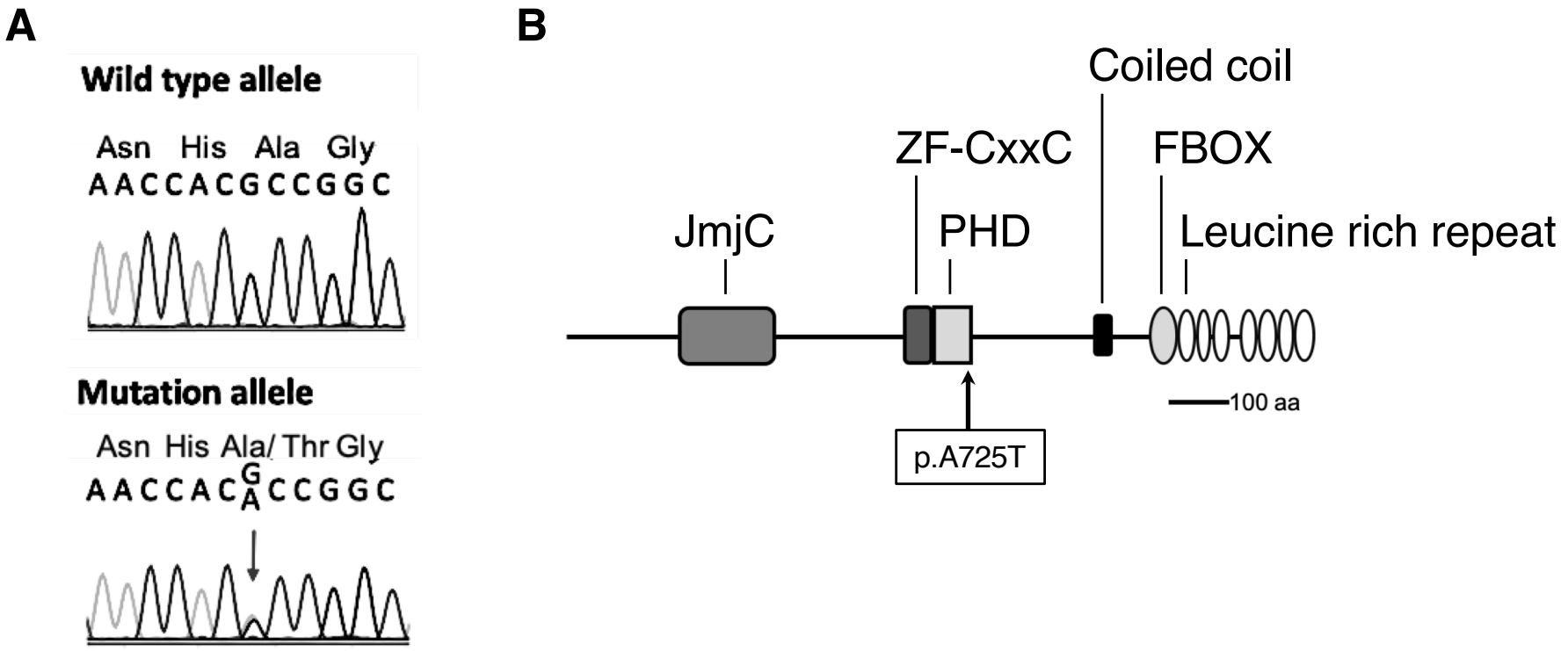

A
Wild type allele
Asn His Ala Gly
AACCACGCCGGC
Mutation allele
Asn His Ala/ Thr Gly
AACCAC G/A CCGGC
B
Coiled coil
ZF-CxxC
FBOX
JmjC
PHD
Leucine rich repeat
100 aa
p.A725T

Figure 4

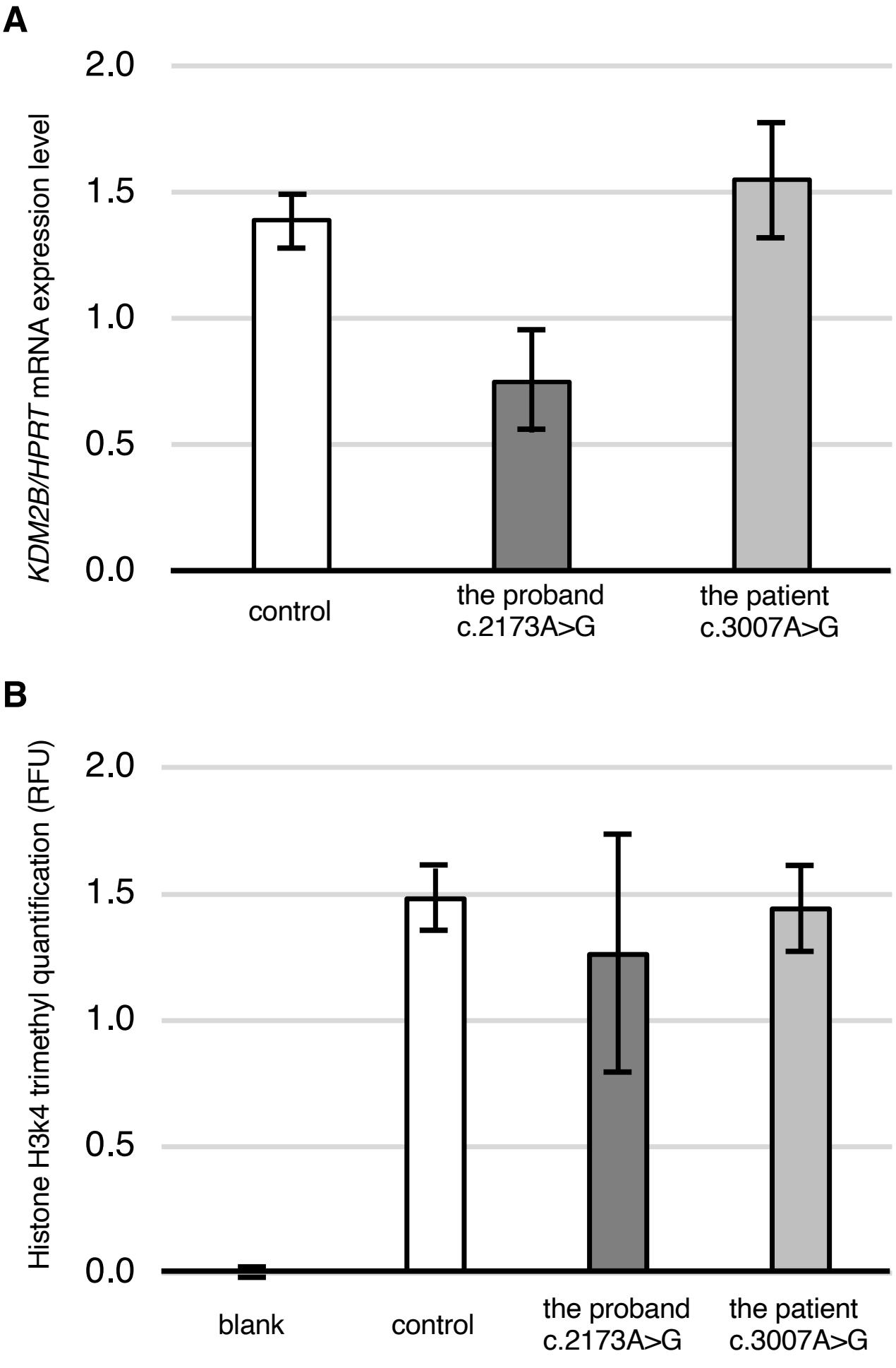

A
2.0
1.5
1.0
0.5
0.0
KDM2B/HPRT mRNA expression level
control
the proband
c.2173A>G
the patient
c.3007A>G
B
2.0
1.5
1.0
0.5
0.0
Histone H3k4 trimethyl quantification (RFU)
blank
control
the proband
c.2173A>G
the patient
c.3007A>G

Figure S1