that the affected siblings would be homozygous for the mutation obtained from a parent, and that both parents would be heterozygous for the mutation. However, we found that only the father expressed the del598-602(GAAGA) mutation, whereas no mutations were identified in any of the 10 exons of the *SIL1* gene in the mother.

We next confirmed the parent–child relationship for each sibling using microsatellite markers on chromosome 5. The mutation and microsatellite analyses suggested that the mother may be hemizygous around exon 6. Quantitative PCR analyses in all family members indicated that the unaffected sibling and father expressed two copies of exon 6 in the *SLI1* gene, whereas the three affected siblings and mother expressed only one copy of exon 6. Therefore, we attempted to define the copy number state for the entire *SIL1* gene using array CGH to confirm the break point of deletion. As it is possible to speculate break points from the array CGH results, we were able to design primers to amplify the deletion-specific product using PCR. Using this method, we found a 58 269-bp deletion in the three affected siblings and mother. The character of break points was not specific, and did not indicate the recombination between the repetitive sequence or low copy repeats.

Table 2 Primer sequences for real-time PCR

Target sequence (*SIL1* exon 2)	
Forward primer	5′-CTCTTGTTGGATGGCTGGAC-3′
Reverse primer	5′-TGTGATTCCCATGTCGTCAC-3′
Target sequence (*SIL1* exon 6)	
Forward primer	5′-GGCAGATGTCTCCAACCAAT-3′
Reverse primer	5′-CTTGTTGATCAGCCGTACCA-3′
Target sequence (*SIL1* exon 10)	
Forward primer	5′-AGAGCTAGCCAGGTGTGAGC-3′
Reverse primer	5′-AGGAGGTGTACCTGGCGATA-3′
Reference sequence (*NSD1*)	
Forward primer	5′-ATGCTTTTTCAGCCCAAATG-3′
Reverse primer	5′-CTCCCTGCAGTACAGCATCA-3′

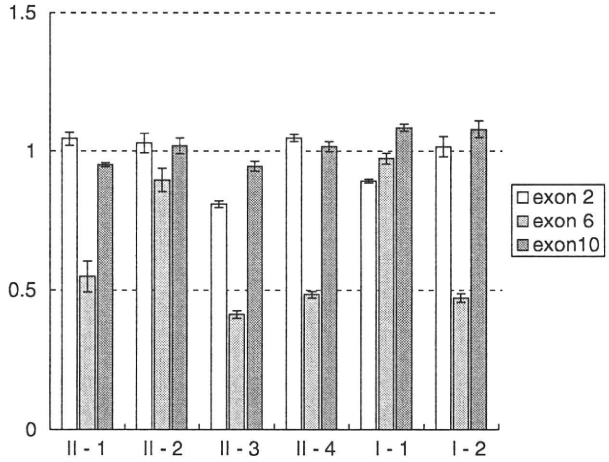


Figure 2 Copy number analysis. The *SIL1* to *NSD1* copy number ratio (*N*) for all family members. The sample with the deletion in *SIL1* is expected to yield *N*=0.5. The *N* values of exon 6 in affected siblings and the mother were 0.5472 (II-1), 0.414 (II-3), 0.483 (II-4) and 0.472 (I-1), respectively. The unaffected sibling was 0.897 (II-2) and the father (II-1) was 0.974. The fact that the *N* value of exons 2 and 10 for all family members was approximately 1.0 suggested that the deletion was not large enough to include the entire *SIL1* gene. The results are presented as mean ± s.d.

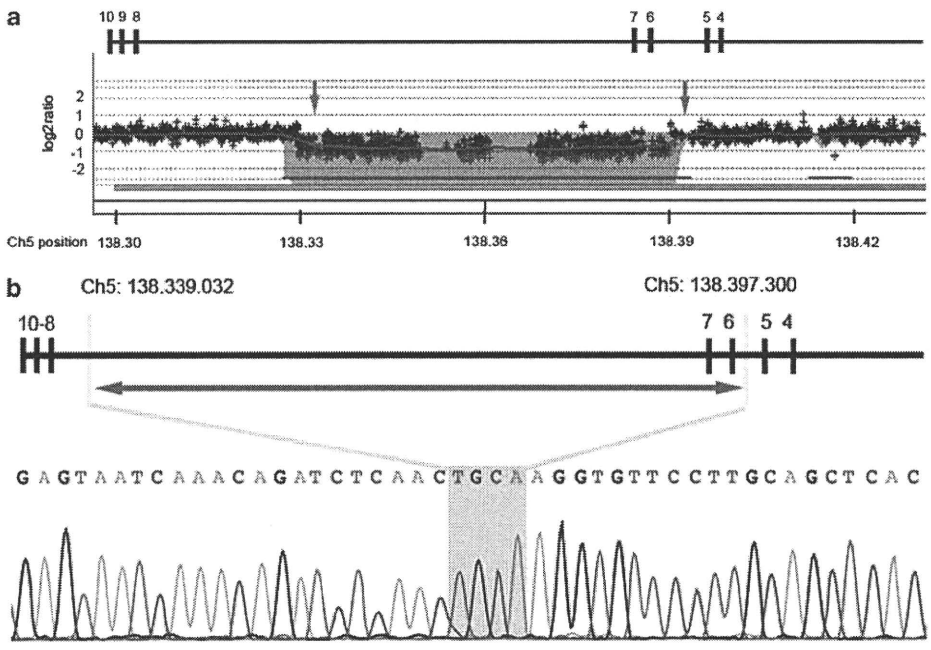


Figure 3 Break-point determination. (**a**) DNA Analytics view of the affected sibling (II-3) using the Agilent custom-designed array showing the approximately 58 kb deletion in the *SIL1* gene. Arrows indicate the break point. (**b**) The sequence results around the break point and the schematic drawing of the *SIL1* gene in the affected sibling (II-3) are shown. A 58 269 bp deletion at chr5: 138 339 032–138 397 300 (NCBI Build 36.1, hg18) and a 4b insertion (shaded region) were identified.

Table 3 Previously reported mutations in the *SIL1* gene of MSS patients

	Type	*Location*	*Nucleotide change*	*Amino-acid change*	*Origin*	P	*Ref.*
1	HM	Exon 3	212dupA	H71Qfs	France	1	6
2	HM	Exon 4	331C>T	R111X	Iran, Turkey, Italy	4	5,6,10
3	HM	Exon 6	506_509dupAAGA	D170fs	Finland, Norway	5	6
4	HM	Exon 6	645+1G>A	Skipping	Turkey	1	5
5	HM	Exon 9	936dupG	L313fs	Japan	2	8,9
6	HM	Exon 9	1029+1G>A	Skipping	Bosnia	1	5
7	HM	Exon 9	1030-9G>A	F345fs	Norway	2	9
8	HM	Exon 9	1249C>T	Q417X	Mali	1	5
9	HM	Exon 10	1312C>T	Q438X	Egypt	1	7
10	HM	Exon 10	1367T>A	L456X	Turkey	1	9
11	HM	Exon 10	1370T>C	L457Pro	Japan	1	9
12	CH	Exons 2, 4	178G>T 346delG	E60X G116fs	Vietnam	1	5
13	CH	Exon 6	506_509dupAAGA 645+2T>C	D170fs skipping	Sweden	1	6
14	CH	Exons 9, 10	947_948insT 1030-18G>A	L316fs M344fs	Germany	1	5
15	CH	Exons 9, 10	947_948insT 1366delT	L316fs 456fs	Russia	1	5

Abbreviations: CH, compound heterozygous; del, deletion; dup, duplication; fs, frameshift; HM, homozygous; ins, insertion; MSS, Marinesco–Sjögren syndrome; P, pedigree number; Ref., reference; X, stop.

MSS is a rare, autosomal recessive disorder. After the two initial groups independently identified several mutations in the *SIL1* gene in 2005,[5,6] only a few mutations in the *SIL1* gene have been reported since.[7–10] Karim *et al.*[7] located a novel mutation in an Egyptian family in 2006, and Eriguchi *et al.*[8] identified a novel mutation in three unrelated Japanese patients in 2008. All mutations in the *SIL1* gene reported previously to be associated with MSS are presented in Table 3. The mutation we found was located in exon 6, which encodes the BiP-interacting domain.[5] Zhao *et al.*[11] have reported that the SIL1 protein associates with the BiP chaperone to aid unfolded proteins in folding normally, and to help in the release of folded proteins. Thus, the loss of SIL1 protein function results in BiP recycling and the accumulation of unfolded proteins in the endoplasmic reticulum.[11–13]

Senderek *et al.*[5] were unable to identify any *SIL1* gene mutations in four individuals with typical MSS. These reports suggested genetic heterogeneity in MSS or that individuals exhibiting MSS may contain mutations that are difficult to detect. For example, compound heterozygous deletions that include different exons or intronic base changes affect the splicing process. In general, when gene mutations in a single gene defect syndrome are detected, it is essential to consider that deletion may not be detected using the PCR-direct sequencing protocol. Our results suggested that deletion assay, quantitative PCR, array CGH or multiple ligation-mediated PCR amplification should be performed to detect deletions of exons in MSS patients. It remains possible that some reported cases without base alterations in the *SIL1* gene are caused by small deletions rather than locus heterogeneity.

ACKNOWLEDGEMENTS

K Yamada was supported partly by a Grants-in-Aid for Scientific Research Category, no. 18791284 from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT). K Yoshiura was supported partly by a Grant-in-Aid for Scientific Research from the Ministry of Health, Labour and Welfare, and partly by grants from the Takeda Scientific Foundation and the Naito Foundation. We are greatly indebted to all the participants of this research. We also thank Ms M Ooga and C Hayashida for their excellent technical assistance.

1 Andersen, B. Marinesco-Sjogren syndrome: spinocerebellar ataxia, congenital cataract, somatic and mental retardation. *Dev Med Child Neurol.* **47,** 249–257 (1965).
2 Marinesco, G., Draganesco, S. & Vasiliu, D. Novelle maladie familiale caracterisee par Une cataracte congenitale et un arret du development somato-neuro-phychique. *Lencephale.* **26,** 97–109 (1931).
3 Sjögren, T. Hereditary congenital spinocerebellar ataxia accompanied by congenital cataracts and oligophrenia. *Confin. Neurol.* **10,** 293–308 (1950).
4 Lagier-Tourenne, C., Tranebjaerg, L., Chaigne, D., Gribaa, M., Dollfus, H., Silvestri, G. *et al.* Homozygosity mapping of Marinesco-Sjogren syndrome to 5q31. *Eur. J. Hum. Genet.* **11,** 770–778 (2003).
5 Senderek, J., Krieger, M., Stendel, C., North, K., Muntoni, F., Quijano-Roy, S. *et al.* Mutations in SIL1 cause Marinesco-Sjogren syndrome, a cerebellar ataxia with cataract and myopathy. *Nat. Genet.* **37,** 1312–1314 (2005).
6 Anttonen, A. K., Mahjneh, I., Hämäläinen, R. H., Lagier-Tourenne, C., Kopra, O., Waris, L. *et al.* The gene disrupted in Marinesco-Sjogren syndrome encodes SIL1, an HSPA5 cochaperone. *Nat. Genet.* **37,** 1309–1311 (2005).
7 Karim, M. A., Parsian, A J., Cleves, M. A., Bracey, J., Elsayed, M. S., Elsobky, E. *et al.* A novel mutation in BAP/SIL1 gene causes Marinesco-Sjögren syndrome in an extended pedigree. *Clin. Genet.* **70,** 420–423 (2006).
8 Eriguchi, M., Mizuta, H., Kurohara, K., Fujitake, J. & Kuroda, Y. Identification of a new homozygous frameshift insertion mutation in the SIL1gene in 3 Japanese patients with Marinesco-Sjögren syndrome. *J. Neurol. Sci.* **270,** 197–200 (2008).
9 Anttonen, A. K., Siintola, E., Tranebjaerg, L., Iwata, N. K., Bijlsma, E. K., Meguro, H. *et al.* Novel SIL1 mutations and exclusion offunctional candidate genes in Marinesco-Sjögren syndrome. *Eur. J. Hum. Genet.* **16,** 961–969 (2008).
10 Annesi, G., Aguglia, U., Tarantino, P., Annesi, F., De Marco, E. V., Civitelli, D. *et al.* SIL1 and SARA2 mutations in Marinesco-Sjögren and chylomicron retention diseases. *Clin. Genet.* **71,** 288–289 (2007).
11 Zhao, L., Longo-Guess, C., Harris, B. S., Lee, J. W. & Ackerman, S.L. Protein accumulation and neurodegeneration in the woozy mutant mouse is caused by disruption of SIL1, a cochaperone of BiP. *Nat Genet.* **37,** 974–979 (2005).
12 Weitzmann, A., Volkmer, J. & Zimmermann, J. The nucleotide exchange factor activity of Grp170 may explain the non-lethal phenotype of loss of Sil1 function in man and mouse. *FEBS Lett.* **580,** 5237–5240 (2006).
13 Weitzmann, A., Baldes, C., Dudek, J. & Zimmermann, R. The heat shock protein 70 molecular chaperone network in the pancreatic endoplasmic reticulum. *FEBS J.* **274,** 5175–5187 (2007).

Human Reproduction, Vol.25, No.4 pp. 1076–1080, 2010
Advanced Access publication on February 6, 2010 doi:10.1093/humrep/deq006

human reproduction **CASE REPORT** ***Reproductive genetics***

A case of Kallmann syndrome carrying a missense mutation in alternatively spliced exon 8A encoding the immunoglobulin-like domain IIIb of fibroblast growth factor receptor 1

Kiyonori Miura[1,*], Shoko Miura[1], Koh-ichiro Yoshiura[2], Stephanie Seminara[3], Daisuke Hamaguchi[1], Norio Niikawa[4], and Hideaki Masuzaki[1]

[1]Department of Obstetrics and Gynecology, Nagasaki University Graduate School of Biomedical Sciences, 1-7-1 Sakamoto, Nagasaki, Japan [2]Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan [3]Reproductive Endocrine Unit, Massachusetts General Hospital, Boston, MA 02114, USA [4]Research Institute of Personalized Health Sciences, Health Sciences University of Hokkaido, Hokkaido, Japan

[*]Correspondence address. Tel: +81-95-819-7363; Fax: +81-95-819-7365; E-mail: kiyonori@nagasaki-u.ac.jp

Submitted on August 24, 2009; resubmitted on December 17, 2009; accepted on January 4 2010

ABSTRACT: Fibroblast growth factor receptor 1 (*FGFR1*) is one of the causative genes for Kallmann syndrome (KS), which is characterized by isolated hypogonadotropic hypogonadism with anosmia/hyposmia. The third immunoglobulin-like domain (D3) of *FGFR1* has the isoforms *FGFR1-IIIb* and *FGFR1-IIIc*, which are generated by alternative splicing of exons 8A and 8B, respectively. To date, the only mutations to have been identified in D3 of *FGFR1* are in exon 8B. We performed mutation analysis of *FGFR1* in a 23-year-old female patient with KS and found a missense mutation (c.1072C>T) in exon 8A of *FGFR1*. The c.1072C>T mutation was not detected in her family members or in 220 normal Japanese and 100 Caucasian female controls. No mutation in other KS genes, KS 1, prokineticin-2, prokineticin receptor-2 and *FGF-8* was detected in the affected patient or in her family members. Therefore, this is the first case of KS carrying a *de novo* missense mutation in *FGFR1* exon 8A, suggesting that isoform *FGFR1-IIIb*, as well as isoform *FGFR1-IIIc*, plays a crucial role in the pathogenesis of KS.

Key words: Kallmann syndrome / *FGFR1b mutation* / fibroblast growth factor receptor 1 isoform expression

Introduction

Kallmann syndrome (KS), which is characterized by isolated hypogonadotropic hypogonadism (IHH) and anosmia/hyposmia, is a clinically and genetically heterogeneous disorder. To date, five causative genes for KS have been reported: KS 1 (*KAL1*, GenBank accession M97 252), prokineticin-2 (*PROK2*, GenBank accession NM 021935), prokineticin receptor-2 (*PROKR2*, GenBank accession NM 144773), fibroblast growth factor-8 (*FGF-8*, GenBank accession NM 033163) and fibroblast growth factor receptor 1 (*FGFR1*, GenBank accession NM 023110.2).

Although sporadic cases of KS are more frequent, families with KS have been reported with X-linked recessive or autosomal dominant or recessive modes of inheritance. Mutations in *KAL1* have been found in familial cases with X-linked recessive inheritance (Franco *et al.*, 1991; Legouis *et al.*, 1991). Mutations in *PROK2* were detected in the heterozygous state, whereas *PROKR2* mutations were found in the heterozygous, homozygous or compound heterozygous state (Dodé *et al.*, 2006). *PROKR2/PROK2* mutations with true pathogenic potential were found only in the homozygous state (Abreu *et al.*, 2008), and any dominant-negative effect of *PROKR2* mutations was ruled out (Monnier *et al.*, 2009). Mutations in *FGFR1* or *FGF8* underlie an autosomal dominant form with incomplete penetrance. Therefore, KS families harbouring heterozygous *FGFR1* or *FGF8* mutations display variable olfactory phenotypes (Dodé *et al.*, 2003; Falardeau *et al.*, 2008), and a few cases with heterozygous *FGFR1* mutations show a normosmic IHH (Pitteloud *et al.*, 2006a). The *FGFR1* gene, which is located on chromosome 8p12, comprises 18 exons (Ruta *et al.*, 1989), and various mutations, including missense and protein truncation mutations, have been reported (Trarbach *et al.*, 2007). The third immunoglobulin-like domain (D3) of *FGFR1* has the isoforms *FGFR1-IIIb* and *FGFR1-IIIc*, which are generated by alternative splicing

of exons 8A and 8B, respectively (Johnson *et al.*, 1991). To date, mutations in D3 of *FGFR1* have only been identified in exon 8B, which encodes immunoglobulin domain IIIc, suggesting that isoform FGFR1-IIIc plays a crucial role in the pathogenesis of KS (Pitteloud *et al.*, 2006b; Trarbach *et al.*, 2006; Dodé *et al.*, 2007).

Here, we report for the first time a KS case carrying a *de novo* missense mutation in the alternatively spliced exon 8A of *FGFR1-IIIb*.

Materials and Methods

Patient and family

Patient (Subject II-2) was a 23-year-old Japanese woman. When she was 18 years old, she was treated at Nagasaki University Hospital because of primary amenorrhea with anosmia. Her height was 159.2 cm and her weight was 72.0 kg. Her serum levels of luteinizing hormone (LH), follicle-stimulating hormone (FSH) and estradiol (E2) were less than 0.5 and 1.5m IU/ml and 10 pg/ml, respectively. Her LH frequent sampling study (sampling performed every 15 min) showed a low-amplitude pattern of LH pulsation (Fig. 1). Her brain magnetic resonance imaging (MRI) examination was negative for tumors and showed no anatomical abnormalities of the hypothalamic–pituitary region and olfactory bulbs. A scratch-and-sniff test (UPSIT, Sensonics, Haddon Hts, NJ, USA) (Doty *et al.*, 1985), which determines ability to smell, indicated anosmia. She was diagnosed as having KS and received hormone replacement therapy for 5 years. Her mother (Subject I-1) was normosmic and had normal puberty and regular menstrual cycles. Her father (Subject I-2), elder brother (Subject II-1) and younger brother (Subject II-3) were also normosmic and had normal puberty (Fig. 2).

Molecular analysis

DNA extraction

Whole blood samples were obtained from the KS patient and from her mother, father, elder and younger brothers. All samples were collected after obtaining written informed consent and the study protocol was approved by the Institutional Review Board of Nagasaki University. Genomic DNA from lymphocytes was extracted using a QIAamp DNA blood mini kit (Qiagen, Düsseldorf, Germany), according to the manufacturer's instructions.

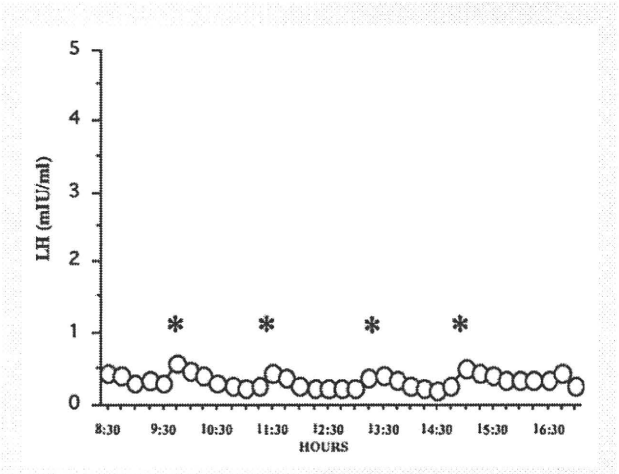


Figure 1 LH pulsation pattern in a case of KS, assayed using an LH frequent sampling study. LH frequent sampling was performed every 15 min. *Low-amplitude pattern of LH pulse.

Sequence analysis

FGFR1 consists of 18 coding exons. Intragenic mutations were investigated by PCR amplification and sequence analysis using 14 pairs of primers, as previously described (Dodé *et al.*, 2003; Sato *et al.*, 2004). Genomic DNA was PCR amplified using conditions of 95°C for 12 min followed by 95°C for 30 s, 59°C for 30 s and 72°C for 60 s for 35 cycles and a final cycle of 72°C for 10 min. PCR products were analyzed by agarose gel electrophoresis, purified with ExoSAP-IT and subjected to sequencing reactions. Sequencing reactions were performed using the BigDye terminator v.3.1 kit and analyzed with an ABI PRISM 3100 Genetic Analyzer™ (Applied Biosystems). The KS patient carrying a mutation in *FGFR1* and her family members were also screened for mutations in the other genes known to be involved in KS [*KAL1*, *PROK2*, *PROKR2* and *FGF8*]. Whether the mutation leads to a change in the protein structure and function was predicted bioinformatically using the ExPASy proteomics server (http://au.expasy.org/) and PolyPhen (http://genetics.bwh.harvard.edu/pph/).

Confirmation of the alternatively spliced exon

Isolation of a full-length murine *Fgfr1-IIIb* showed that Fgfr1-IIIb was a transmembrane receptor (Beer *et al.*, 2000). Although the mRNA encoding exon IIIb has been found in human (Johnson *et al.*, 1991), the presence of sequences encoding the intracellular domain has not yet been demonstrated. Therefore, to determine the splice site of exon 8A and to detect *FGFR1-IIIb* mRNA encoding the intracellular domain, we performed RT–PCR using specific primers to amplify the splice isoform containing exon 8A. Kal23 is designed to span exons 7 and 8A for specific annealing to the *FGFR1-IIIb* isoform, which is spliced from exon 7 to exon 8A (Fig. 3A). Kal5 is designed within exon 8B for specific annealing to the *FGFR1-IIIc* isoform, which is spliced from exon 7 to exon 8B. Kal2 and Kal6 are designed within exons 7 and 9, respectively, for annealing to the D3 isoforms of *FGFR1*. Primer sequences were as follows: kal2: 5′-GACAGAAGGTCGGTTATGTC-3′, Kal23: 5′-CAGATCTTGAAGC ATTCGGG-3′, Kal5: 5′-GGTGGTATTAACTCCAGCAG-3′ and Kal6: 5′-GTACAGGGGCGAGGTCATCA-3′. The BD multiple tissue complementary DNA (cDNA, MTC) panels Human I and Human II (BD Biosciences Clontech, Mountain View, CA, USA) were used to detect the expression of each isoform of *FGFR1*. PCR amplification was performed on cDNAs as follows: 94°C, 30 s; 62°C, 30 s; 72°C, 1 min; 40 cycles. PCR products were analyzed by agarose gel electrophoresis and sequenced using then ABI PRISM 3100 Genetic Analyzer™.

Results

Sequence analysis of the entire coding region of *FGFR1*, including exon–intron boundary regions, showed that the KS patient had a mutation (c.1072C>T) in exon 8A of *FGFR1-IIIb*, while the other family members did not (Fig. 2). However, the full-length *FGFR1* mRNA that includes exon 8A is not deposited in the full-length cDNA database (GenBank accession no. NM 023110.2). RT–PCR analysis indicated that most transcripts containing exon 8A were spliced to exon 8B in all adult tissues except bone marrow (data not shown). We wished to demonstrate the existence of an alternative transcript, exon 8A which was spliced to exon 9 encoding the transmembrane helix; therefore, RT–PCR products from human fetal brain were cloned and sequenced. In 1 of 27 clones exon 8A was spliced to exon 9 (designated here '*FGFR1-IIIb*', GenBank accession FJ809917, see Fig. 3B), while in the other clones exon 8A was spliced to exon 8B (designated here

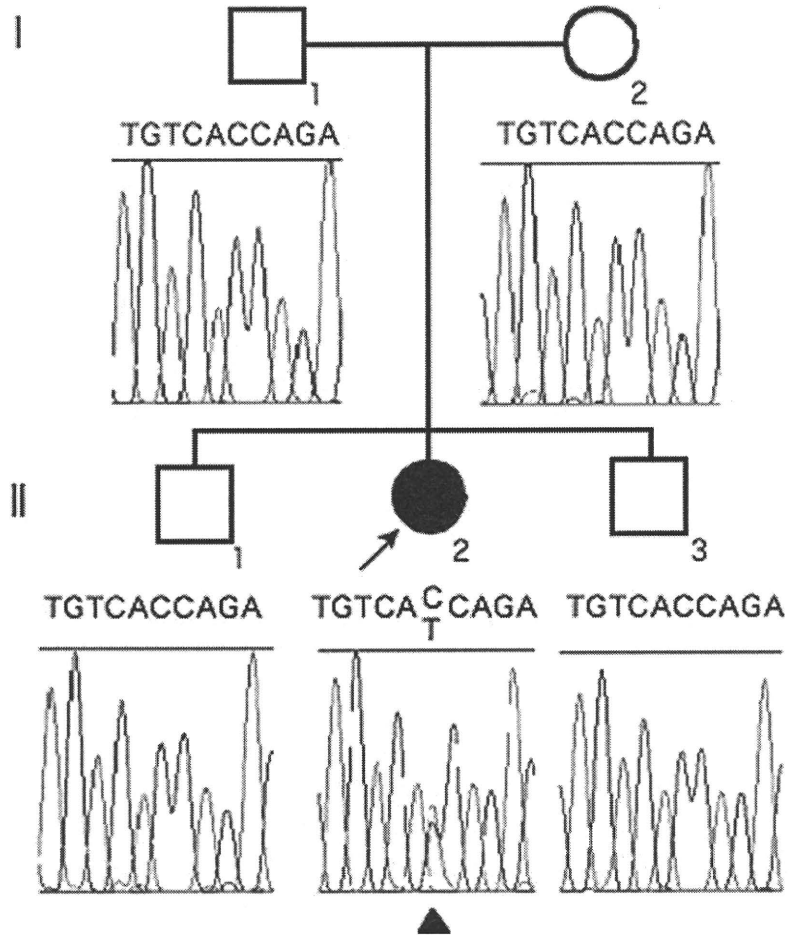


Figure 2 Pedigree of patient's family and the results of fibroblast growth factor receptor 1 (*FGFR1*) sequence analysis. II-2 is a 23-year-old woman with KS. The patient had a mutation in *FGFR1* (C>T) but the other family members did not. The mutation was *de novo* because parentage was assured. Arrowhead under the electropherogram indicates the mutation site.

'*FGFR1-secr*', GenBank accession FJ809916, see Fig. 3B). The exact acceptor and donor sites of exon 8A in '*FGFR1-IIIb*' mRNA, which produces a membrane-bound FGFR1-containing D3, were determined by sequence analysis of splice isoforms, '*FGFR1-IIIb*' and '*FGFR1-secr*', (Fig. 3). As most full-length *FGFR1* cDNAs in the database were transcripts containing exon 7–exon 8B–exon 9 (designated here '*FGFR1-IIIc*', GenBank accession NM 023110.2, see Fig. 2B) without exon 8A, '*FGFR1-IIIc*' is likely to be the most abundantly expressed human isoform. The EST, CA488712.1, was the only isoform in the EST database corresponding to '*FGFR1-IIIb*'. Although both '*FGFR1-IIIb*' and '*FGFR1-IIIc*' encode membrane-bound *FGFR1*, '*FGFR1-secr*' encodes a secreted form of *FGFR1* because of a sequence frameshift and a termination codon in exon 9.

The exons 8A and 8B of the human FGFR1 isoforms shared the amino acid sequence at 354–357; WLTV. However, exon 8A ends with six extra amino acids at 358–363, TRPVAK, whereas exon 8B ends with only two, LE. These sequences are identical in the mouse Fgfr1 isoforms (Beer *et al.*, 2000). The mutation in exon 8A of *FGFR1-IIIb* (GenBank accession no. FJ809917 bankit1193625) is c.1072C>T at the cDNA level and p.T358I at the amino acid level. Bioinformatic analysis shows the mutated amino acid residue to be conserved between human and mouse and to be located in D3 of *FGFR1-IIIb*, which is a critical region for FGF ligand binding. However, the mutation was not predicted to produce a change in the human protein structure. The c.1072C>T mutation was not detected in 220 normal Japanese women or in 100 normal Caucasian women. The patient had no mutation in any of the other four KS genes.

Discussion

Mouse *Fgfr1-IIIb* has a low level of expression in a wide variety of adult tissues, but a high level of expression in skin and brain, indicating the existence of specific splicing factors in skin and brain that recognize the relatively weak *Fgfr1-IIIb* splice site (Beer *et al.*, 2000). Consistent with the expression pattern of mouse *Fgfr1-IIIb*, we could isolate human *FGFR1-IIIb* from a fetal brain cDNA library but not from adult

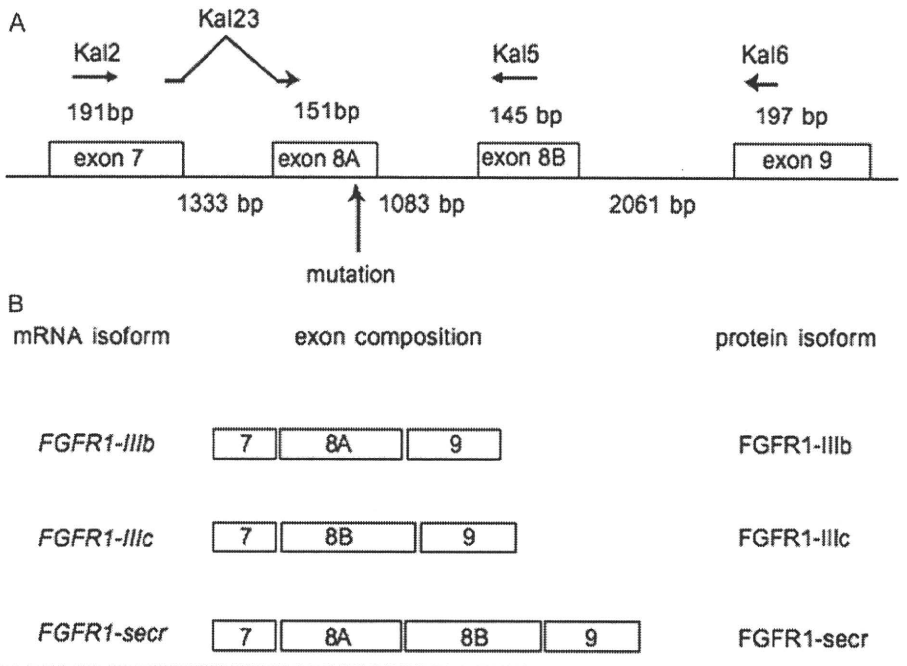


Figure 3 Genomic organization of *FGFR1* around exon 8A. (A) The numbers with bp indicate exon length (over the line) and intron length (under the line). The mutation is located near the end of exon 8A. Horizontal arrows indicate the locations of primers used to perform RT–PCR amplification of the isoform containing exons 7 and 8A. Primer Kal23 is designed to span exons 7 and 8A for specific annealing to the *FGFR1-IIIb* isoform, which is spliced from exon 7 to exon 8A (Fig. 3A). Kal5 is designed within exon 8B for specific annealing to the *FGFR1-IIIc* isoform, which is spliced from exon 7 to exon 8B. Kal2 and Kal6 are designed within exons 7 and 9, respectively, for annealing to the D3 isoforms of *FGFR1*. Vertical arrow indicates mutation site. (B) Composition of mRNA isoforms and of putative protein structures. FGFR1-IIIb: membrane-bound form of FGFR1 with immunoglobulin-like domain IIIb encoded by exon 8A, FGFR1-IIIc: membrane-bound form of FGFR1 with immunoglobulin-like domain IIIc encoded by exon 8B, FGFR1-secr: a secreted form of FGFR1.

tissues. Most full-length *FGFR1* cDNAs in the database represent *FGFR1-IIIc*. *FGFR1-IIIc* is expressed at high levels, but *FGFR1-IIIb* is expressed at very low levels (Johnson *et al.*, 1991). We can amplify only a tiny amount of *FGFR1-IIIb* that has exon 8A spliced to exon 9 by RT–PCR using the primers Kal23 and Kal6. Most of the sequenced RT–PCR products corresponded to *FGFR1-secr*, suggesting that the expression level of the three isoforms is '*FGFR1-IIIc*' ≫ '*FGFR1-IIIsecr*' ≫ '*FGFR1-IIIb*' (Fig. 3B).

Several studies suggested that mutations in exon 8B of isoform *FGFR1-IIIc* are implicated in the pathogenesis of KS (Pitteloud *et al.*, 2006b; Trarbach *et al.*, 2006; Dodé *et al.*, 2007). Mice homozygous for alleles with a stop codon in exon IIIc displayed phenotypes resembling those of embryos homozygous for null alleles, while mice carrying an in-frame stop codon in exon IIIb were viable and fertile (Partanen *et al.*, 1998). Therefore, *Fgfr1-IIIc* is the dominant isoform that carries out the majority of the biological functions of the *Fgfr1* gene, whereas *Fgfr1-IIIb* plays a minor and to some extent redundant role (Partanen *et al.*, 1998). A receptor-binding analysis revealed no difference in the binding specificity between the endogenous *Fgfr1-IIIb* and an artificially created *Fgfr1-IIIb*, which had two different amino acids in the 3′-end of the unique IIIb exon (Beer *et al.*, 2000), suggesting that the carboxyl terminus of D3 may not overtly influence binding specificity. However, being expressed at low levels does not imply that '*FGFR1-IIIb*' has an unimportant role. The mutation we found, p.T358I, was located in exon 8A of *FGFR1-IIIb*. Therefore, p.T358I may affect ligand binding and cause the KS phenotype, although how this mutation affects the loss of function of *FGFR1-IIIb* is unknown. The spatio-temporal expression of any gene involved in development is key; therefore, the expression and functional involvement of *FGFR1-IIIb* may be important in the early embryonic brain, in particular during GnRH neuronal development. KS missense mutations in *FGFR1* are distributed in the first, second and third immunoglobulin-like domains (D1–D3), in the tyrosine kinase domain and also in the intracellular domain (Dodé *et al.*, 2003; Sato *et al.*, 2004; Albuisson *et al.*, 2005; Pitteloud *et al.*, 2006a; Trarbach *et al.*, 2006; Dodé *et al.*, 2007); therefore, the membrane-bound form of *FGFR1* is probably important for the KS phenotype. The membrane-bound form of *FGFR1-IIIb* could, therefore, be the critical isoform for the KS phenotype.

We present here, for the first time, a case of KS carrying a missense mutation in exon 8A of *FGFR1*, suggesting that the minor isoform '*FGFR1-IIIb*' as well as the major isoform '*FGFR1-IIIc*' has a crucial role in the pathogenesis of KS. Therefore, immunoglobulin-like domain IIIb may have an essential role in GnRH neuronal migration, which is initiated from the nasal placode and runs towards the forebrain following the olfactory sensory neuron axonal connection with the developing olfactory bulb. Further experiments are needed to show that the mutation in

exon 8A causes KS, such as expression of the mutated isoform in transfected cells to analyze receptor stability and signaling efficiency. Although *FGFR1* containing immunoglobulin-like domain IIIb has not been analyzed intensively, our mutation report should encourage researchers to analyze immunoglobulin-like domain IIIb function and the spatio-temporal expression of exon 8A in fetal brain development.

Acknowledgements

We thank Ms Yasuko Noguchi and Miho Ooga for their technical assistance and we thank Drs Nelly Pitteloud and William Crowley for their valuable contribution.

Funding

K.M. was supported, in part, by Seeds (No.15-B09) from the Japan Science and Technology Agency (JST), by grants from the Naito Foundation, by Grant-in-Aid for Young Scientists (B) (no. 21791567) from the Ministry of Education, Sports, Culture, Science and Technology of Japan, by a Grant for Child Health and Development (20C-1) from the Ministry of Health, Labor and Welfare, and by Grant-in-Aid for Scientific Research from Nagasaki University, Japan.

References

Abreu AP, Trarbach EB, de Castro M, Frade Costa EM, Versiani B, Matias Baptista MT, Garmes HM, Mendonca BB, Latronico AC. Loss-of-function mutations in the genes encoding prokineticin-2 or prokineticin receptor-2 cause autosomal recessive Kallmann syndrome. *J Clin Endocrinol Metab* 2008;**93**:4113–4118.

Albuisson J, Pêcheux C, Carel JC, Lacombe D, Leheup B, Lapuzina P, Bouchard P, Legius E, Matthijs G, Wasniewska M *et al.* Kallmann syndrome: 14 novel mutations in KAL1 and FGFR1 (KAL2). *Hum Mutat* 2005;**25**:98–99.

Beer HD, Vindevoghel L, Gait MJ, Revest JM, Duan DR, Mason I, Dickson C, Werner S. Fibroblast growth factor (FGF) receptor 1-IIIb is a naturally occurring functional receptor for FGFs that is preferentially expressed in the skin and the brain. *J Biol Chem* 2000; **275**:16091–16097.

Dodé C, Levilliers J, Dupont JM, De Paepe A, Le Dû N, Soussi-Yanicostas N, Coimbra RS, Delmaghani S, Compain-Nouaille S, Baverel F *et al.* Loss-of-function mutation in *FGFR1* cause autosomal dominant Kallmann syndrome. *Nature Genet* 2003;**33**:463–465.

Dodé C, Teixeira L, Levilliers J, Fouveaut C, Bouchard P, Kottler ML, Lespinasse J, Lienhardt-Roussie A, Mathieu M, Moerman A *et al.* Kallmann syndrome: mutations in the genes encoding prokineticin-2 prokineticin receptor-2. *PLoS Genet* 2006;**2**:e175.

Dodé C, Fouveaut C, Mortier G, Janssens S, Bertherat J, Mahoudeau J, Kottler ML, Chabrolle C, Gancel A, François I *et al.* Novel FGFR1 sequence variants in Kallmann syndrome, and genetic evidence that the FGFR1c isoform is required in olfactory bulb and palate morphogenesis. *Hum Mutat* 2007;**28**:97–98.

Doty RL, Applebaum S, Zusho H, Settle RG. Sex differences in odor identification ability: a cross-cultural analysis. *Neuropsychologia* 1985; **23**:667–672.

Falardeau J, Chung WC, Beenken A, Raivio T, Plummer L, Sidis Y, Jacobson-Dickman EE, Eliseenkova AV, Ma J, Dwyer A *et al.* Decreased FGF8 signaling causes deficiency of gonadotropin-releasing hormone in humans and mice. *J Clin Invest* 2008;**118**:2822–2831.

Franco B, Guioli S, Pragliola A, Incerti B, Bardoni B, Tonlorenzi R, Carrozzo R, Maestrini E, Pieretti M, Taillon-Miller P *et al.* A gene deleted in Kallmann's syndrome shares homology with neural cell adhesion and axonal path-finding molecules. *Nature* 1991;**353**:529–536.

Johnson DE, Lu J, Chen H, Werner A, Williams LT. The human fibroblast growth factor receptor genes: a common structure arrangement underlies the mechanisms for generating receptor forms that differ in their third immunoglobulin domain. *Mol Cell Biol* 1991;**11**:4627–4634.

Legouis R, Hardelin JP, Levilliers J, Claverie JM, Compain S, Wunderle V, Millasseau P, Le Paslier D, Cohen D, Caterina D *et al.* The candidate gene for the X-linked Kallmann syndrome encodes a protein related to adhesion molecules. *Cell* 1991;**67**:423–435.

Monnier C, Dodé C, Fabre L, Teixeira L, Labesse G, Pin JP, Hardelin JP, Rondard P. PROKR2 missense mutations associated with Kallmann syndrome impair receptor signalling activity. *Hum Mol Genet* 2009; **18**:75–81.

Partanen J, Schwartz L, Rossant J. Opposite phenotypes of hypomorphic and Y766 phosphorylation site mutations reveal a function for Fgfr1 in anteroposterior patterning of mouse embryos. *Genes Dev* 1998; **12**:2332–2344.

Pitteloud N, Acierno JS Jr, Meysing A, Eliseenkova AV, Ma J, Ibrahimi OA, Metzger DL, Hayes FJ, Dwyer AA, Hughes VA *et al.* Mutations in fibroblast growth factor receptor 1 cause both Kallmann syndrome and normosmic idiopathic hypogonadotropic hypogonadism. *Proc Natl Acad Sci USA* 2006a;**103**:6281–6286.

Pitteloud N, Meysing A, Quinton R, Acierno JS Jr, Dwyer AA, Plummer L, Fliers E, Boepple P, Hayes F, Seminara S *et al.* Mutations in fibroblast growth factor receptor 1 cause Kallmann syndrome with a wide spectrum of reproductive phenotypes. *Mol Cell Endocrinol* 2006b; **254–255**:60–69.

Ruta M, Burgess W, Givol D, Epstein J, Neiger N, Kaplow J, Crumley G, Dionne C, Jaye M, Schlessinger J. Receptor for acidic fibroblast growth factor is related to the tyrosine kinase encoded by the fms-like gene (FLG). *Proc Natl Acad Sci USA* 1989;**86**:8722–8726.

Sato N, Katsumata N, Kagami M, Hasegawa T, Hori N, Kawakita S, Minowada S, Shimotsuka A, Shishiba Y, Yokozawa M *et al.* Clinical assessment and mutation analysis of Kallmann syndrome 1 (KAL1) and fibroblast growth factor receptor 1 (FGFR1, or KAL2) in five families and 18 sporadic patients. *J Clin Endorinol Matab* 2004; **89**:1079–1088.

Trarbach EB, Costa EM, Versiani B, de Castro M, Baptista MT, Garmes HM, de Mendonca BB, Latronico AC. Novel fibroblast growth factor receptor 1 mutations in patients with congenital hypogonadotropic hypogonadism with and without anosmia. *J Clin Endocrinol Metab* 2006;**91**:4006–4012.

Trarbach EB, Silveira LG, Latronico AC. Genetic insights into human isolated gonadotropin deficiency. *Pituitary* 2007;**10**:381–391.

nature genetics

Exome sequencing identifies *MLL2* mutations as a cause of Kabuki syndrome

Sarah B Ng[1,7], Abigail W Bigham[2,7], Kati J Buckingham[2], Mark C Hannibal[2,3], Margaret J McMillin[2], Heidi I Gildersleeve[2], Anita E Beck[2,3], Holly K Tabor[2,3], Gregory M Cooper[1], Heather C Mefford[2], Choli Lee[1], Emily H Turner[1], Joshua D Smith[1], Mark J Rieder[1], Koh-ichiro Yoshiura[4], Naomichi Matsumoto[5], Tohru Ohta[6], Norio Niikawa[6], Deborah A Nickerson[1], Michael J Bamshad[1–3] & Jay Shendure[1]

We demonstrate the successful application of exome sequencing[1–3] to discover a gene for an autosomal dominant disorder, Kabuki syndrome (OMIM%147920). We subjected the exomes of ten unrelated probands to massively parallel sequencing. After filtering against existing SNP databases, there was no compelling candidate gene containing previously unknown variants in all affected individuals. Less stringent filtering criteria allowed for the presence of modest genetic heterogeneity or missing data but also identified multiple candidate genes. However, genotypic and phenotypic stratification highlighted *MLL2*, which encodes a Trithorax-group histone methyltransferase[4]: seven probands had newly identified nonsense or frameshift mutations in this gene. Follow-up Sanger sequencing detected *MLL2* mutations in two of the three remaining individuals with Kabuki syndrome (cases) and in 26 of 43 additional cases. In families where parental DNA was available, the mutation was confirmed to be *de novo* ($n = 12$) or transmitted ($n = 2$) in concordance with phenotype. Our results strongly suggest that mutations in *MLL2* are a major cause of Kabuki syndrome.

Kabuki syndrome is a rare, multiple malformation disorder characterized by a distinctive facial appearance (**Supplementary Fig. 1**), cardiac anomalies, skeletal abnormalities, immunological defects and mild to moderate mental retardation. Originally described in 1981 (refs. 5,6), Kabuki syndrome has an estimated incidence of 1 in 32,000 (ref. 7), and approximately 400 cases have been reported worldwide. The vast majority of reported cases have been sporadic, but parent-to-child transmission in more than a half dozen instances[8] suggests that Kabuki syndrome is an autosomal dominant disorder. The relatively low number of cases, the lack of multiplex families and the phenotypic variability of Kabuki syndrome have made the identification of the gene(s) underlying this disorder intractable to conventional approaches of gene discovery, despite aggressive efforts.

We sequenced the exomes of ten unrelated individuals with Kabuki syndrome: seven of European ancestry, two of Hispanic ancestry and one of mixed European and Haitian ancestry (**Supplementary Fig. 1** and **Supplementary Table 1**). Enrichment was performed by hybridization of shotgun fragment libraries to custom microarrays followed by massively parallel sequencing[1–3]. On average, 6.3 gigabases of sequence were generated per sample to achieve 40× coverage of the mappable, targeted exome (31 Mb). As with our previous studies, we focused our analyses here primarily on nonsynonymous variants, splice acceptor and donor site mutations and coding indels, anticipating that synonymous variants were far less likely to be pathogenic. We also predicted that variants underlying Kabuki syndrome are rare, and therefore likely to be previously unidentified. We defined variants as previously unidentified if they were absent from all datasets used for comparison, including dbSNP129, the 1000 Genomes Project, exome data from 16 individuals previously reported by us[2,3] and 10 exomes sequenced as part of the Environmental Genome Project (EGP).

Under a dominant model in which each case was required to have at least one previously unidentified nonsynonymous variant, splice acceptor and donor site mutation or coding indel variant in the same gene, only a single candidate gene (*MUC16*) was shared across all ten exomes (**Table 1** and **Supplementary Table 2**). However, we considered *MUC16* as a likely false positive due to its extremely large size (14,507 amino acids). Potential explanations for our failure to find a compelling candidate gene in which newly identified variants were seen in all affected individuals included: (i) Kabuki syndrome is genetically heterogeneous and therefore not all affected individuals will have mutations in the same gene; (ii) we failed to identify all mutations in the targeted exome; and (iii) some or all causative mutations were outside of the targeted exome, for example, in noncoding regions or unannotated genes. To allow for a modest degree of genetic heterogeneity and/or missing data, we conducted a less stringent analysis by looking for candidate genes shared among subsets of affected individuals. Specifically, we searched

[1]Department of Genome Sciences, University of Washington, Seattle, Washington, USA. [2]Department of Pediatrics, University of Washington, Seattle, Washington, USA. [3]Seattle Children's Hospital, Seattle, Washington, USA. [4]Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan. [5]Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan. [6]Research Institute of Personalized Health Sciences, Health Sciences University of Hokkaido, Hokkaido, Japan. [7]These authors contributed equally to this work. Correspondence should be addressed to J.S. (shendure@u.washington.edu) or M.J.B. (mbamshad@u.washington.edu).

Received 28 April; accepted 21 July; published online 15 August 2010; doi:10.1038/ng.646

Table 1 Number of genes common to any subset of *x* affected individuals.

Subset analysis (any *x* of 10)	1	2	3	4	5	6	7	8	9	10
NS/SS/I	12,042	8,722	7,084	6,049	5,289	4,581	3,940	3,244	2,486	1,459
Not in dbSNP129 or 1000 Genomes	7,419	2,697	1,057	488	288	192	128	88	60	34
Not in control exomes	7,827	2,865	1,025	399	184	90	50	22	7	2
Not in either	6,935	2,227	701	242	104	44	16	6	3	1
Is loss-of-function (nonsense or frameshift indel)	753	49	7	3	2	2	1	0	0	0

The number of genes with at least one nonsynonymous variant (NS), splice-site acceptor or donor variants (SS) or coding indel (I) are listed under various filters. Variants were filtered by presence in dbSNP or 1000 Genomes (not in dbSNP129 or 1000 genomes) and control exomes (not in control exomes) or both (not in either); control exomes refer to those from 8 Hapmap[3], 4 FSS[3], 4 Miller[2] and 10 EGP samples. The number of genes found using the union of the intersection of *x* individuals is given.

for subsets of x out of 10 exomes having ≥1 previously unidentified variant in the same gene, with $x = 1$ to $x = 10$. For $x = 9$, $x = 8$ and $x = 7$, previously unidentified variants were shared in 3 genes, 6 genes and 16 genes, respectively (**Table 1**). However, there was no obvious way to rank these candidate genes.

We speculated that genotypic and/or phenotypic stratification would facilitate the prioritization of candidate genes identified by subset analysis. Specifically, we assigned a categorical rank to each individual with Kabuki syndrome based on a subjective assessment of the presence of, or similarity to, the canonical facial characteristics of Kabuki syndrome (**Supplementary Fig. 1**) and the presence of developmental delay and/or major birth defects (**Supplementary Table 1**). The highest-ranked individual was one of a pair of monozygotic twins with Kabuki syndrome. We then categorized the functional impact (that is, nonsense versus nonsynonymous substitution, splice-site disruption and frameshift compared to in-frame indel) of each newly identified variant in candidate genes shared by each subset of two or more ranked cases. Manual review of these data highlighted distinct, previously unidentified nonsense variants in *MLL2* in each of the four highest-ranked cases. After sequential analysis of phenotype-ranked cases with a loss-of-function filter, *MLL2* was the only candidate gene remaining after addition of the second individual (**Table 2**). We found no such variant in *MLL2* in the individual with Kabuki syndrome ranked fifth; hence, the number of candidate genes dropped to zero after the individual ranked fourth in the set (**Table 2**). However, we found a 4-bp deletion in the individual ranked sixth, and we found nonsense variants in the individuals ranked seventh and ninth. Thus, exome sequencing identified a nonsense substitution or frameshift indel in *MLL2* in seven of the ten individuals with Kabuki syndrome analyzed here.

Retrospectively, we applied a loss-of-function filter to the subset analysis of exome data (**Table 1**), and at $x = 7$, found *MLL2* to be the only candidate gene. We also developed a *post hoc* ranking of candidate genes based on the functional impact of the variants present (variant score) and the rank of the cases in which each variant was observed (case score). When this was applied to the exome data as a combined metric, *MLL2* emerged as the top candidate gene (**Supplementary Fig. 2**).

In parallel with these analyses, we applied genomic evolutionary rate profiling (GERP)[9] to the exome data. GERP uses mammalian genome alignments to define a rejected substitution score for each variant regardless of functional class. We have previously shown that the quantitative ranking of candidate genes by the rejected substitution scores of their variants can facilitate the exome-based analysis of Mendelian disorders[10]. Following subset analysis with GERP-based ranking, *MLL2* remained on the candidate list up to $x = 8$, ranking third in a list of 11 candidate genes at this threshold (**Table 3** and **Supplementary Fig. 3**). Notably, the additional *MLL2* variant contributing to this analysis (such that *MLL2* was still considered at $x = 8$) was a synonymous substitution with a rejected substitution score of 0.368 in the individual ranked fifth.

We sought to confirm all newly identified variants in *MLL2*, particularly because loss-of-function variants identified through massively parallel sequencing have a high prior probability of being false positives. All seven loss-of-function variants in *MLL2* were validated by Sanger sequencing. We further analyzed the three cases in which we did not initially find a loss-of-function variant in *MLL2*, first by array comparative genomic hybridization (aCGH) to determine any gross structural changes and then by Sanger sequencing of all exons of *MLL2* in case of false negatives by exome sequencing. Because an average of 96% of the coding bases in *MLL2* were called at sufficient quality and coverage for single nucleotide variant detection, we anticipated that any missed variants were more likely to be indels because of the higher coverage required for confident indel detection in short-read sequence data. Indeed, although aCGH did not find any structural variants in the region, Sanger sequencing did identify frameshift indels in two of these three cases (specifically, the cases ranked eighth and tenth).

Ultimately, loss-of-function mutations in *MLL2* were identified in nine out of ten cases in the discovery cohort (**Fig. 1**), making this gene a compelling candidate for Kabuki syndrome. For validation, we screened all 54 exons of *MLL2* in 43 additional cases by Sanger sequencing. Previously unidentified nonsynonymous, nonsense or frameshift mutations in *MLL2* were found in 26 of these 43 cases (**Fig. 1** and **Supplementary Table 3**). In total, through either exome sequencing or targeted sequencing of *MLL2*, 33 distinct *MLL2* mutations were identified in 35 of 53 families (66%) with Kabuki syndrome (**Fig. 1** and **Supplementary Table 3**). In each of 12 cases for which DNA from both parents was available, the *MLL2* variant was found to have occurred *de novo*. Three mutations were found in two individuals each. One of these three mutations was confirmed to have arisen *de novo* in one of the cases, indicating that some mutations in individuals with Kabuki syndrome are recurrent. In addition, *MLL2* mutations (resulting in p.4527K>X and p.5464T>M) were also identified in each of two families in which Kabuki syndrome was transmitted from parent to child.

Table 2 Number of genes common in sequential analysis of phenotypically ranked individuals

Sequential analysis	1	+2	+3	+4	+5	+6	+7	+8	+9	+10
NS/SS/I	5,282	3,850	3,250	2,354	2,028	1,899	1,772	1,686	1,600	1,459
Not in dbSNP129 or 1000 Genomes	687	214	145	84	63	54	42	40	39	34
Not in control exomes	675	134	50	26	13	13	8	5	4	2
Not in either	467	89	34	18	9	8	4	4	3	1
Is loss-of-function (nonsense/frameshift indel)	25	1	1	1	0	0	0	0	0	0

Variants were filtered as in **Table 1**. Exomes were added sequentially to the analysis by ranked phenotype; for example, column "+3" shows the number of genes at the intersection of the three top ranked cases (**Supplementary Fig. 1**). The gene with at least one NS/SS/I in all individuals is *MUC16*, which is very likely to be a false positive due to its extreme length (14,507 amino acids).

Table 3 Analysis of exome variants using genomic evolutionary rate profiling

GERP score analysis (at least *x* of 10)	1	2	3	4	5	6	7	8	9	10
Variant RS score > 0	7,176	2,360	754	269	106	39	20	11	3	1
MLL2 rank	3,732	1,232	399	136	47	14	6	3	NA	NA

The number of genes with at least a single previously unidentified variant with a rejected substitution score[10] > 0 in at least *x* individuals is given. A gene rank is assigned based on the average GERP score[9] over all newly identified variants with rejected substitution score > 0 in all affected individuals.

None of the additional *MLL2* mutations was found in 190 control chromosomes from individuals of matched geographical ancestry.

Our results strongly suggest that mutations in *MLL2* are a major cause of Kabuki syndrome. *MLL2* encodes a large 5,262-residue protein that is part of the SET family of proteins, of which Trithorax, the *Drosophila* homolog of MLL, is the best characterized[11]. The SET domain of MLL2 confers strong histone 3 lysine 4 methyltransferase activity and is important in the epigenetic control of active chromatin states[12]. In mice, loss of *Mll2* on a mixed 129Sv/C57BL/6 background slows growth, increases apoptosis and retards development, leading to early embryonic lethality due in part to misregulation of homeobox gene expression[13]. However, no morphological defects have been reported in $Mll2^{+/-}$ mice[13].

Most of the *MLL2* variants identified in individuals with Kabuki syndrome are predicted to truncate the polypeptide chain before translation of the SET domain. Though it is not certain whether Kabuki syndrome results from haploinsufficiency or from a gain of function at *MLL2*, haploinsufficiency seems to be the more likely mechanism. Deletion of chromosome 12q12–q13.2, which encompasses *MLL2*, has been reported in a child with characteristics of Noonan syndrome[14]. However, we re-analyzed this case using oligo aCGH (including 21 probes that cover *MLL2*) and found the distal breakpoint to be located ~700 kb proximal to *MLL2* (data not shown). Also, all of the pathogenic missense variants identified here are located in regions of *MLL2* that encode C-terminal domains. This suggests that missense variants elsewhere in *MLL2* may be better tolerated or, alternatively, may be embryonically lethal.

For the 18 of 53 cases for which no previously unidentified protein-altering variant was found, it is possible that noncoding or other missed mutations in *MLL2* are responsible for this disorder. Alternatively, Kabuki syndrome could be genetically heterogeneous, and further analysis of these cases by exome sequencing may elucidate additional genes for Kabuki syndrome and potentially explain some of the phenotypic heterogeneity seen in this disorder. Notably, 9 of 10 individuals in the discovery cohort (90%), but only 26 of 43 individuals in the replication cohort (60%), were ultimately found to have mutations in *MLL2*. It is therefore possible that the careful selection of canonical Kabuki cases for the discovery cohort enriched for a shared genetic basis. This underscores the importance of access to deeply phenotyped and well-characterized cases.

In summary, we applied exome sequencing of a small number of unrelated individuals with Kabuki syndrome to discover that mutations in *MLL2* underlie this disorder. As predicted in previous analyses[2,3], allowing for even a small degree of genetic heterogeneity or missing data substantially confounds exome analysis by increasing the number of candidate genes consistent with the model of inheritance. To facilitate the prioritization of genes under such criteria, we stratified data by ranked phenotypes and found that *MLL2* was prominent in the higher ranked cases. However, nine of the ten individuals with Kabuki syndrome in the discovery cohort were ultimately found to have *MLL2* mutations, such that stratification by phenotype was of less importance than originally appeared to have been the case. Nonetheless, the sequential analysis of ranked cases may have reduced the probability of confounding due to genetic heterogeneity. All of the *MLL2* mutations found in the discovery set via exome sequencing were loss-of-function variants. As a result, *MLL2* ranked highly among candidate genes assessed by predicted functional impact. Such a pattern will likely occur for some, but not all, Mendelian phenotypes subjected to this approach. We anticipate that the further development of strategies to stratify data at both the genotypic and phenotypic level will be critical for exome and whole-genome sequencing to reach their full potential as tools for discovery of genes underlying Mendelian and complex diseases.

URLs. RefSeq 36.3, ftp://ftp.ncbi.nlm.nih.gov/genomes/MapView/Homo_sapiens/sequence/BUILD.36.3/updates/seq_gene.md.gz; Phaster, http://www.phrap.org; SeattleSeq Annotation, http://gvs.gs.washington.edu/SeattleSeqAnnotation/; 1000 Genomes Project, http://www.1000genomes.org/page.php/; dbGaP accession, http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000295.v1.p1.

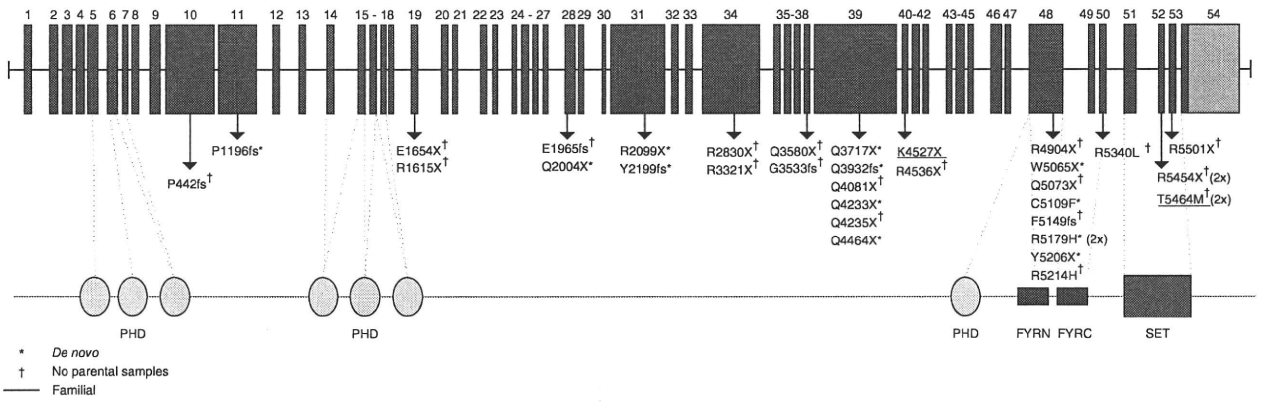


Figure 1 Genomic structure and allelic spectrum of *MLL2* mutations that cause Kabuki syndrome. *MLL2* is composed of 54 exons that encode untranslated regions (orange) and protein coding sequence (blue) including 7 PHD fingers (yellow), FYRN (green), FYRC (green) and a SET domain (red). Arrows indicate the locations of 32 different mutations found in 53 families with Kabuki syndrome including 20 nonsense mutations, 7 indels and 5 amino acid substitutions. Asterisks indicate mutations that were confirmed to be *de novo* and crosses indicate cases for which parental DNA was unavailable. The two underlined mutations were transmitted each within a family, from an affected parent to an affected child.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Accession codes. Exome data for the discovery cohort is available via the NCBI dbGaP repository under accession number phs000295.v1.p1.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank the families for their participation and the Kabuki Syndrome Network for their support. We thank J. Allanson, J. Carey and M. Golabi for referral of cases and M. Emond for helpful discussion. We thank the 1000 Genomes Project for early data release that proved useful for filtering out common variants. Our work was supported in part by grants from the US National Institutes of Health (NIH)–National Heart, Lung, and Blood Institute (5R01HL094976 to D.A.N. and J.S.), the NIH–National Human Genome Research Institute (5R21HG004749 to J.S., 1RC2HG005608 to M.J.B., D.A.N. and J.S.; and 5RO1HG004316 to H.K.T.), NIH–National Institute of Environmental Health Sciences (HHSN273200800010C to D.N. and M.J.R.), Ministry of Health, Labour and Welfare (K.Y., N.M., T.O. and N.N.), Japan Science and Technology Agency (N.M.), Society for the Promotion of Science (N.M.), the Life Sciences Discovery Fund (2065508 and 0905001), the Washington Research Foundation and the NIH–National Institute of Child Health and Human Development (1R01HD048895 to M.J.B.). S.B.N. is supported by the Agency for Science, Technology and Research, Singapore. A.W.B. is supported by a training fellowship from the NIH–National Human Genome Research Institute (T32HG00035).

AUTHOR CONTRIBUTIONS

The project was conceived and the experiments were planned by M.J.B., D.A.N. and J.S. The review of phenotypes and the sample collection were performed by M.J.B., M.C.H., M.J.M., K.Y., N.M., T.O. and N.N. Experiments were performed by S.B.N., K.J.B., A.E.B., C.L., H.C.M., J.D.S., M.J.R., E.H.T. and H.I.G. Ethical consultation was provided by H.K.T. Data analysis was performed by A.W.B., M.J.B., K.J.B., G.M.C., S.B.N. and J.S. The manuscript was written by M.J.B., S.B.N. and J.S. All aspects of the study were supervised by M.J.B. and J.S.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Choi, M. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. USA* **106**, 19096–19101 (2009).
2. Ng, S.B. *et al.* Exome sequencing identifies the cause of a Mendelian disorder. *Nat. Genet.* **42**, 30–35 (2010).
3. Ng, S.B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272–276 (2009).
4. FitzGerald, K.T. & Diaz, M.O. MLL2: A new mammalian member of the *trx/MLL* family of genes. *Genomics* **59**, 187–192 (1999).
5. Niikawa, N., Matsuura, N., Fukushima, Y., Ohsawa, T. & Kajii, T. Kabuki make-up syndrome: a syndrome of mental retardation, unusual facies, large and protruding ears, and postnatal growth deficiency. *J. Pediatr.* **99**, 565–569 (1981).
6. Kuroki, Y., Suzuki, Y., Chyo, H., Hata, A. & Matsui, I. A new malformation syndrome of long palpebral fissures, large ears, depressed nasal tip, and skeletal anomalies associated with postnatal dwarfism and mental retardation. *J. Pediatr.* **99**, 570–573 (1981).
7. Niikawa, N. *et al.* Kabuki make-up (Niikawa-Kuroki) syndrome: a study of 62 patients. *Am. J. Med. Genet.* **31**, 565–589 (1988).
8. Courtens, W., Rassart, A., Stene, J.J. & Vamos, E. Further evidence for autosomal dominant inheritance and ectodermal abnormalities in Kabuki syndrome. *Am. J. Med. Genet.* **93**, 244–249 (2000).
9. Cooper, G.M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* **15**, 901–913 (2005).
10. Cooper, G.M. *et al.* Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. *Nat. Methods* **7**, 250–251 (2010).
11. Prasad, R. *et al.* Structure and expression pattern of human *ALR*, a novel gene with strong homology to *ALL-1* involved in acute leukemia and to *Drosophila* trithorax. *Oncogene* **15**, 549–560 (1997).
12. Issaeva, I. *et al.* Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. *Mol. Cell. Biol.* **27**, 1889–1903 (2007).
13. Glaser, S. *et al.* Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development* **133**, 1423–1432 (2006).
14. Tonoki, H., Saitoh, S. & Kobayashi, K. Patient with del(12)(q12q13.12) manifesting abnormalities compatible with Noonan syndrome. *Am. J. Med. Genet.* **75**, 416–418 (1998).

ONLINE METHODS

Cases and samples. For exome sequencing, we selected ten individuals of self-reported European, Hispanic or mixed European and Haitian ancestry with Kabuki syndrome from ten unrelated families. Phenotypic data were collected from review of medical records, phone interviews and photographs. All participants provided written consent, and the Institutional Review Boards of Seattle Children's Hospital and the University of Washington approved all studies. The clinical characteristics of the 43 individuals in the validation cohort who had been diagnosed with Kabuki syndrome have been reported previously[7]. Subjective assessment and ranking of the Kabuki phenotype was based on pictures of each subject (**Supplementary Fig. 1**) and clinical information (**Supplementary Table 1**). Informed consent was obtained for publication of each of the facial photos shown.

Exome definition, array design and target masking. We targeted all protein-coding regions as defined by RefSeq 36.3. Entries were filtered for the following: (i) CDS as the feature type, (ii) transcript name starting with "NM_" or "-", (iii) reference as the group_label, (iv) not being on an unplaced contig (for example, 17|NT_113931.1). Overlapping coordinates were collapsed for a total of 31,922,798 bases over 186,040 discontiguous regions. A single custom array (Agilent, 1M features, aCGH format) was designed to have probes over these coordinates as previously described[3], except here, the maximum melting temperature (T_m) was raised to 73 °C.

The mappable exome was also determined as previously described[3] using this RefSeq exome definition instead. After masking for 'unmappable' regions, 30,923,460 bases were left as the mappable target.

Targeted capture and massive parallel sequencing. Genomic DNA was extracted from peripheral blood lymphocytes using standard protocols. Five micrograms of DNA from each of ten individuals with Kabuki syndrome was used for construction of a shotgun sequencing library as described previously[3] using paired-end adaptors for sequencing on an Illumina Genome Analyzer II (GAII). Each shotgun library was hybridized to an array for target enrichment; this was then followed by washing, elution and additional amplification. Enriched libraries were then sequenced on a GAII to get either single-end or paired-end reads.

Read mapping and variant analysis. Reads were mapped and processed largely as previously described[3]. In brief, reads were quality recalibrated using Eland and then aligned to the reference human genome (hg18) using Maq. When reads with the same start site and orientation were filtered, paired-end reads were treated like separate single-end reads; this method is overly conservative and hence the actual coverage of the exomes is higher than reported here. Sequence calls were performed using Maq and these calls were filtered to coordinates with ≥8× coverage and consensus quality ≥ 20.

Indels affecting coding sequences were identified as previously described[3], but we used phaster instead of cross_match and Maq. Specifically, unmapped reads from Maq were aligned to the reference sequence using phaster (version 1.100122a) with the parameters -max_ins:21 -max_del:21 -gapextend_ins:-1 -gapextend_del:-1 -match_report_type:1. Reads were then filtered for those with at most two substitutions and one indel. Reads that mapped to the negative strand were reverse complemented and, together with the other filtered reads, were remapped using the same parameters to reduce ambiguity in the called indel positions. These reads were then filtered for (i) having a single indel more than 3 bp from the ends and (ii) having no other substitutions in the read. Putative indels were then called per individual if they were supported by at least two filtered reads that started from different positions. An 'indel reference' was generated as previously described[3], and all the reads from each individual were mapped back to this reference using phaster with default settings and -match_report_type:1. Indel genotypes were called as previously described[3].

To determine the novelty of the variants, sequence calls were compared against 16 individuals for whom we had previously reported exome data[2,3] and 10 EGP exomes. Annotations of variants were based on NCBI and UCSC databases using an in-house server (SeattleSeqAnnotation). Loss-of-function variants were defined as nonsense mutations (premature stop) or frame-shifting indels. For each variant, we also generated constraint scores as implemented in GERP[10].

***Post hoc* ranking of candidate genes.** Candidate genes were ranked by summation of a case score and variant score. The case score was calculated by counting the total number of Kabuki exomes in which a variant was identified at a given gene, weighted for case rank from 1 to 10. For example, the top ranked case was weighted by a factor of 10, whereas the case ranked tenth was weighted by a factor of 1. The variant score was calculated by first counting the total number of nonsense, nonsynonymous and synonymous variants across the ten Kabuki exomes and assigning a prior probability of the occurrence of each variant type per gene based upon the target of 18,918 genes. Next, for each candidate gene shared among two or more Kabuki exomes, the scores for each newly identified variant were summed across the gene. The case score and variant score were summed as the candidate gene score.

Mutation validation. Sanger sequencing of PCR amplicons from genomic DNA was used to confirm the presence and identity of variants in the candidate gene identified via exome sequencing and to screen the candidate gene in additional individuals with Kabuki syndrome.

Array comparative genomic hybridization (CGH). Samples were hybridized to commercially available whole-genome tiling arrays consisting of one million oligonucleotide probes with an average spacing of 2.6 kb throughout the genome (SurePrint G3 Human CGH Microarray 1x1M, Agilent Technologies). Twenty-one probes on this array covered *MLL2* specifically. Data were analyzed using Genomics Workbench software according to the manufacturer's instructions.

doi:10.1038/ng.646

Exome sequencing identifies *MLL2* mutations as a cause of Kabuki syndrome

Sarah B. Ng[1*], Abigail W. Bigham[2*], Kati J. Buckingham[2], Mark C. Hannibal[2,3], Margaret McMillin[2], Heidi Gildersleeve[2], Anita E. Beck[2,3], Holly K. Tabor[2,3], Greg M. Cooper[1], Heather C. Mefford[2], Choli Lee[1], Emily H. Turner[1], Josh D. Smith[1], Mark J. Rieder[1], Koh-ichiro Yoshiura[4], Naomichi Matsumoto[5], Tohru Ohta[6], Norio Niikawa[6], Deborah A. Nickerson[1], Michael J. Bamshad[1,2,3†], Jay Shendure[1†]

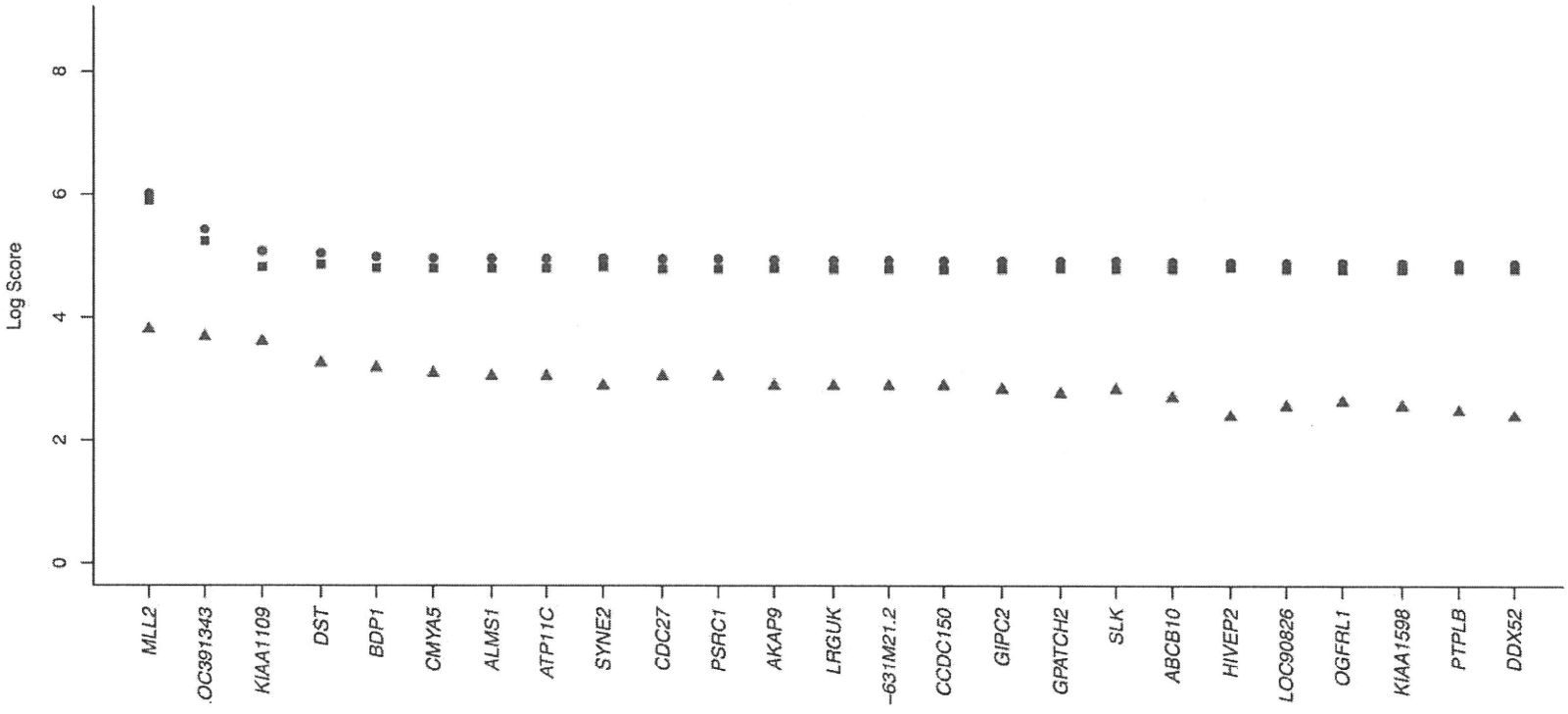
Departments of [1]Genome Sciences and [2]Pediatrics, University of Washington, Seattle, Washington, USA. [3]Seattle Children's Hospital, Seattle, Washington, USA. [4]Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan. [5]Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan. [6]Research Institute of Personalized Health Sciences, Health Sciences University of Hokkaido, Hokkaido, Japan

Supplementary Figures 1-3 and Supplementary Tables 1-3

Nature Genetics: doi:10.1038/ng.646

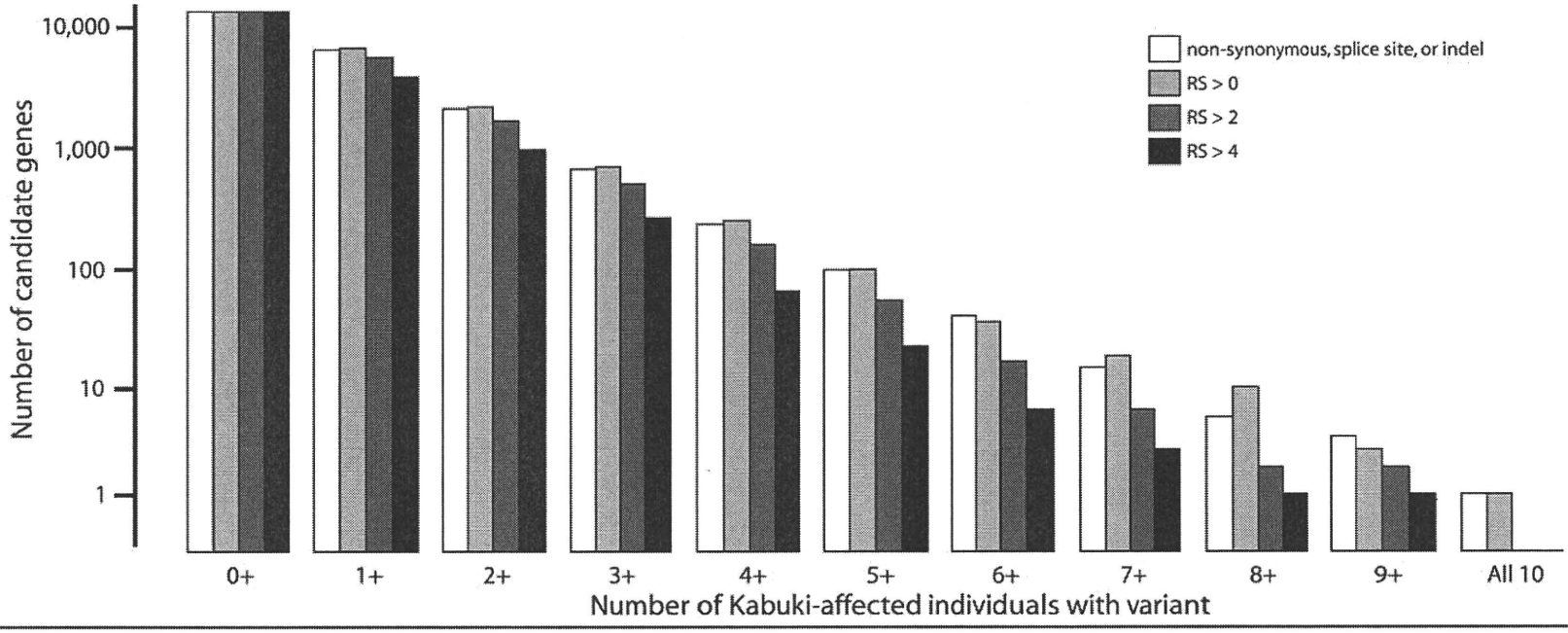
Supplementary Figure 1. Photographs of the facial characteristics used to determine the subjective ranking of Kabuki phenotypes. The phenotype ranking of the ten children with Kabuki syndrome is listed here from 1-10 based on similarity to the canonical phenotype of Kabuki syndrome. Asterisks indicate cases in which *MLL2* mutations were identified. Informed consent was obtained for publication of each of the facial photos shown.

Nature Genetics: doi:10.1038/ng.646



Supplementary Figure 2. Ranking of candidate genes identified by exome sequencing based on shared variants and functional annotation. Plot of the log of the top twenty-five candidate gene scores (green circles) in 10 Kabuki cases ranked by sum of case score (red triangles) and variant score (blue squares).

Nature Genetics: doi:10.1038/ng.646



MLL2 rank :	3,732	3,732	1,232	399	136	47	14	6	3	NA	NA
Total (RS > 0) :	14,407	7,176	2,360	754	269	106	39	20	11	3	1

Supplementary Figure 3. Constraint scores enrich candidate gene pool for variants causing Kabuki syndrome. Number of candidate genes in which at least the given number of individuals with Kabuki syndrome has a rare variant that is functionally defined or with RS scores as indicated. Total number of candidate genes identified at RS > 0 and rank of *MLL2* among those genes are also given.

Nature Genetics: doi:10.1038/ng.646

Supplementary Table 1. Clinical characteristics used to determine the subjective phenotype ranking of the 10 children with Kabuki syndrome.

	Cardiovascular	Spleen/Liver abnormality	Kidney abnormality	Kidney dysfunction	Hearing loss	Preauricular pits/tags	Cleft palate	High arched palate	Hypotonia	Developmental delay
1	ASD/VSD, aortic coarctation, bicuspid valves, dysrhythmia	np	X	np	X	X	X	X	X	X
2	aortic coarctation, bicuspid valves	-	X	X	X	X	X	X	X	X
3	ASD/VSD, aortic coarctation, bicuspid valves	-	-	-	X	np	np	X	X	X
4	VSD, aortic coarctation	X	np	np	X	-	X	X	X	X
5	ASD, VSD	np	np	np	np	np	X	X	X	X
6	np	np	np	np	-	np	np	X	X	X
7	np	-	X	-	X	np	np	X	X	X
8	np	X	X	np	np	X	X	X	X	X
9	np	np	np	np	np	np	np	np	np	X
10	ASD, VSD, dysrhythmia	X	X	X	X	np	np	X	X	X

X denotes abnormality present
np denotes abnormality not present
- denotes data not found in medical record

Nature Genetics: doi:10.1038/ng.646

Supplementary Table 2. Lists of certain candidate genes as identified in Table 1 and 2.

Gene names given for cells in Table 1 where number of genes < 20

A. Subset analysis (any *x* of 10)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
NS/SS/I	12,042	8,722	7,084	6,049	5,289	4,581	3,940	3,244	2,486	1,459
Not in dbSNP129 or 1000 genomes	7,419	2,697	1,057	488	288	192	128	88	60	34
Not in control exomes	7,827	2,865	1,025	399	184	90	50	22	7 MUC2, LIN37, MUC6, AHNAK2, UGT2B10, MUC16, MUC5B	2 MUC16, AHNAK2
Not in either	6,935	2,227	701	242	104	44	16 AHNAK2, MUC2, EVPL, LOC100128468, MUC6, MLL2, PCDHGB1, DSPP, LOC391343, MYCBP2, PCDHGA3, PCDHGA2, MUC16, PCDHGA1, LYST, PKHD1L1	6 MUC2, MUC6, DSPP, MUC16, AHNAK2, LYST	3 MUC16, AHNAK2, MUC2	1 MUC16
Is loss-of-function (nonsense / frameshift indel)	753	49	7 FLJ34443, GJB4, MLL3, MLL2, SQLE, LOC391343, ZNF598	3 MLL2, ZNF598, FLJ34443	2 MLL2, ZNF598	2 MLL2, ZNF598	1 MLL2	0	0	0

Nature Genetics: doi:10.1038/ng.646

Supplementary Table 2 continued. Lists of certain candidate genes as identified in Table 1 and 2.

Gene names given for cells in Table 2 where number of genes < 20

B. Sequential analysis	**1**	**+2**	**+3**	**+4**	**+5**	**+6**	**+7**	**+8**	**+9**	**+10**
NS/SS/I	5,282	3,850	3,250	2,354	2,028	1,899	1,772	1,686	1,600	1,459
Not in dbSNP129 or 1000 genomes	687	214	145	84	63	54	42	40	39	34
Not in control exomes	675	134	50	26	13 OBSCN, LIN37, MUC16, PCDHGB1, PCDHGA2, PCDHGA3, PCDHGA1, PCDHGA4, MUC2, MUC6, UGT2B10, AHNAK2, PKHD1L1	13 OBSCN, LIN37, MUC16, PCDHGB1, PCDHGA2, PCDHGA3, PCDHGA1, PCDHGA4, MUC2, MUC6, UGT2B10, AHNAK2, PKHD1L1	8 OBSCN, LIN37, MUC16, MUC2, MUC6, UGT2B10, AHNAK2, PKHD1L1	5 LIN37, MUC16, MUC2, MUC6, AHNAK2	4 MUC16, MUC2, MUC6, AHNAK2	2 MUC16, AHNAK2
Not in either	467	89	34	18 SRRM2, ATG2B, VPS13A, FNDC1, MLL2, MUC4, UBR1, MUC16, PCDHGB1, PCDHGA2, PCDHGA3, PCDHGA1, LYST, A26C1A, MUC2, MUC6, AHNAK2, PKHD1L1	9 MUC16, PCDHGB1, PCDHGA2, PCDHGA3, PCDHGA1, MUC2, MUC6, AHNAK2, PKHD1L1	8 MUC16, PCDHGB1, PCDHGA2, PCDHGA3, PCDHGA1, MUC2, MUC6, AHNAK2	4 MUC16, MUC2, MUC6, AHNAK2	4 MUC16, MUC2, MUC6, AHNAK2	3 MUC16, MUC2, AHNAK2	1 MUC16
Is loss-of-function (nonsense / frameshift indel)	25	1 MLL2	1 MLL2	1 MLL2	0	0	0	0	0	0

Nature Genetics: doi:10.1038/ng.646

Supplementary Table 3. Annotation of all *MLL2* mutations found in 53 Kabuki cases screened.

Kindred	Indiv	Exome Sequenced	Mutation	Exon	Predicted Amino Acid Change	Confirmed as *de novo*	Position[o]
1		yes	c.G15195A	48	p.W5065X	+	chr12:47706821
2		yes	c.C6010T	28	p.Q2004X	+	chr12:47722238
3		yes	c.C12697T	39	p.Q4233X	+	chr12:47712058
4		yes	c.C8488T	34	p.R2830X	-	chr12:47718918
5		yes	--	--	--	--	--
6		yes	c.11794_11797delCAAC	39	p.Q3932SfsX46	+	chr12:47712958-61
7		yes	c.T15618G	48	p.Y5206X	+	chr12:47706398
8		yes	c.3585_3586insA	11	p.P1196TfsX11	+	chr12:47730053-54
9		yes	c.C6295T	31	p.R2099X	+	chr12:47721525
10		yes	c.6595delT	31	p.Y2199IfsX65	+	chr12:47721225
11		no	c.G15326T	48	p.C5109F	+	chr12:47706690
12		no	c.G15536A	48	p.R5179H	+	chr12:47706480
13		no	c.C11149T	39	p.Q3717X	+	chr12:47713606
14		no	c.15444_15445delTT	48	p.F5149CfsX9	-	chr12:47706571-72
15		no	c.C15217T	48	p.Q5073X	-	chr12:47706799
16		no	c.C9961T	34	p.R3321X	-	chr12:47717445
17		no	c.C14710T	48	p.R4904X	-	chr12:47707306
18		no	c.5875_5891dup17	28	p.E1965GfsX88	-	chr12:47722341-57
19		no	c.G15536A	48	p.R5179H	-	chr12:47706480
20		no	c.C12703T	39	p.Q4235X	-	chr12:47712152
21		no	c.C12241T	39	p.Q4081X	-	chr12:47712514
22		no	c.C13390T	39	p.Q4464X	+	chr12:47711365
23		no	c.G15641A	48	p.R5214H	-	chr12:47706375
24		no	--	--	--	--	--
25	1	no	c.A13580T	40	p.K4527X	*	chr12:47711035
	2	no	c.A13580T	40	p.K4527X	-	chr12:47711035
26		no	c.C16501T	53	p.R5501X	-	chr12:47702113
27	1	no	--	--	--	--	--
	2	no	--	--	--	--	--
28		no	--	--	--	--	--
29		no	c.C10738T	38	p.Q3580X	-	chr12:47714119
30		no	--	--	--	--	--
31		no	c.C16360T	52	p.R5454X	-	chr12:47702382
32		no	--	--	--	--	--
33		no	--	--	--	--	--
34		no	--	--	--	--	--
35		no	c.4956_4957insG	19	p.E1654X	-	chr12:47724800-01
36		no	c.10599_10630del32	38	p.V3534QfsX11	-	chr12:47714227-58
37		no	--	--	--	--	--
38		no	c.C13606T	40	p.R4536X	-	chr12:47711008
39		no	c.G16019T	50	p.R5340L	-	chr12:47704661
40		no	c.C4843T	19	p.R1615X	-	chr12:47724914
41		no	c.C16391T	52	p.T5464M	-	chr12:47702351
42		no	--	--	--	--	--
43		no	--	--	--	--	--
44		no	--	--	--	--	--
45		no	c.C16360T	52	p.R5454X	-	chr12:47702382
46		no	--	--	--	--	--
47		no	--	--	--	--	--
48		no	--	--	--	--	--
49		no	--	--	--	--	--
50		no	c.1324delC	10	p.P442HfsX487	-	chr12:47732409
51		no	--	--	--	--	--
52	1	no	c.C16391T	52	p.T5464M	*	chr12:47702351
	2	no	c.C16391T	52	p.T5464M	-	chr12:47702351
53		no	--	--	--	--	--

-- no mutation identified
\+ confirmed *de novo*
\- no parental samples available
X stop codon
fs frameshift
[o] chromosomal position was determined using March 2006 assembly from UCSC (hg18)
* confirmed as inherited

Kindreds 25, 27 and 52 show dominant transmission of Kabuki syndrome from parent to child. Both affected individuals are listed here.

Nature Genetics: doi:10.1038/ng.646