DATA REPORT

An unclassified variant of *CHD7* activates a cryptic splice site in a patient with CHARGE syndrome

Yuko Katoh-Fukui[1], Shuichi Yatsuga[2], Hirohito Shima[1], Atsushi Hattori[1], Akie Nakamura[1], Kohji Okamura[3], Kumiko Yanagi[4], Manami Iso[4], Tadashi Kaname[4], Yoichi Matsubara[5] and Maki Fukami[1]

CHARGE syndrome is a rare autosomal dominant disease that is typically caused by heterozygous *CHD7* mutations. A *de novo* variant in a *CHD7* splicing acceptor site (NM_017780.3:c.7165−4A > G) was identified in a Japanese boy with CHARGE syndrome. This variant has been considered to be an "unclassified variant" due to its position outside the consensus splicing sites. In this study, abnormal splicing derived from this known variant was confirmed by cDNA sequencing.

CHARGE syndrome (CS) (OMIM 214800) is a rare, autosomal dominant syndromic disease that is characterized by a variety of anomalies, including choanal atresia and malformations of the heart, ears and eyes. These malformations include coloboma, microphthalmia, septal defects in the heart, growth retardation and delayed mental development, genital hypoplasia, external and middle ear abnormalities, and deafness. Furthermore, facial palsy, cleft palate and dysphagia are commonly associated with CS.[1–3] CS is estimated to occur at a rate of 1 in 8,500–12,000 live births.[2,4,5] CS is mainly caused by loss-of-function mutations within *CHD7* on chromosome 8q12. DNA sequencing detects *CHD7* mutations in 58–64% of patients clinically diagnosed with CS.[5–7] A large-scale study indicated that among pathogenic *CHD7* mutations, 78% constitute nonsense or frameshift mutations; 8% are missense mutations; and 11% are splice site mutations.[8] Currently, 581 mutations have been classified as pathogenic, whereas 93 mutations remain unclassified (CHD7 database: https://mol genis51.gcc.rug.nl/menu/main/dataexplorer?entity = Mutations&hi deselect = true#). These unclassified mutations mostly comprise missense substitutions or occur around the exon/intron junctions.

This study was approved by the Institutional Review Board Committees of the National Research Institute for Child Health and Development and Kurume University. The patient was a 7-year-old Japanese boy diagnosed with CS on the basis of the phenotypic expression of the disease, namely, choanal atresia, cleft palate, dysphagia, cryptorchidism, advanced hearing loss and cardiac septal defects. No coloboma or microphthalmia has been observed thus far. An apparent retardation of growth or development has also not been identified. The clinical details of the patient will be described in a subsequent report. Informed consent was obtained from the patient's parents.

Exome sequencing was performed as described earlier.[9] Using leukocytes obtained from the patient and his parents, genomic DNA was extracted using the DNeasy blood and tissue kit (Qiagen, Hilden, Germany). DNA fragments corresponding to protein-coding sequences were enriched using commercially available oligonucleotide libraries followed by next-generation sequencing (HiSeq 2500; Illumina, San Diego, CA, USA). The read lengths of the single- and paired-end libraries were 125 bases. The sequences obtained from the patient and his parents were aligned to the reference genome (GRCh37/hg19) using Burrows-Wheeler Aligner version 0.6.2 (Geeknet, Mountain View, CA, USA). Removal of potential PCR duplicates, recalibration of base quality values, local realignments and variation calls were performed using Samtools, Picard (http://broadinstitute.github.io/picard/) and GATK version 2.3–9. Filtration of putative false-positive variations was executed using optimization criteria, quality value recalibration, indel realignment, variant quality score recalibration, variant quality score recalibration PASS filter and GATK-recommended hard filters (GATK best practice v3/v4). Nucleotide substitutions with allele frequencies above 1% in the normal population were excluded as polymorphisms. A rare *CHD7* variant detected by exome sequencing was confirmed by Sanger sequencing using the *CHD7*-intron 33 FW1 (5′-CAGCTCTGTGCACCAGTCAT-3′) and *CHD7*-exon 34 RV1-2 (5′-AAGGCATCAGACACTGGTGT-3′) primer pair. We retrieved the mutation from previous reports or from public databases of normal populations (The ExAC browser, http://exac.broadinstitute. org/; 1000 Genome Browser, https://www.ncbi.nlm.nih.gov/varia tion/tools/1000genomes/).

Total RNA was extracted from lymphoblastoid cells from the patient's blood sample using the RNeasy RNA extraction kit (Qiagen). cDNA was reverse-transcribed from total RNA with the EMD Millipore Novagen oligo(dT) primer (Thermo Fisher Scientific, Waltham, MA, USA) and Applied Biosystems High-Capacity RNA-to-cDNA kit (Foster City, CA, USA). cDNA was PCR-amplified using *CHD7* coding sequence amplification with the *CHD7*-exon 33 FW2 (5′-GGCCTCATCCCAGGTTACAC-3′) and *CHD7*-exon 34 RV1-2 (5′-AAGGCATCAGACACTGGTGT-3′) primer pair. Amplified cDNA fragments were sub-cloned into the TOPO vector (Life Technologies, Carlsbad, CA, USA), and sequences were confirmed by Sanger sequencing.

[1]Department of Molecular Endocrinology, National Research Institute for Child Health and Development, Tokyo, Japan; [2]Department of Pediatrics and Child Health, Kurume University School of Medicine, Fukuoka, Japan; [3]Department of Systems BioMedicine, National Research Institute for Child Health and Development, Tokyo, Japan; [4]Department of Genome Medicine, National Research Institute for Child Health and Development, Tokyo, Japan and [5]National Research Institute for Child Health and Development, Tokyo, Japan.
Correspondence: Y Katoh-Fukui (fukui-y@ncchd.go.jp)
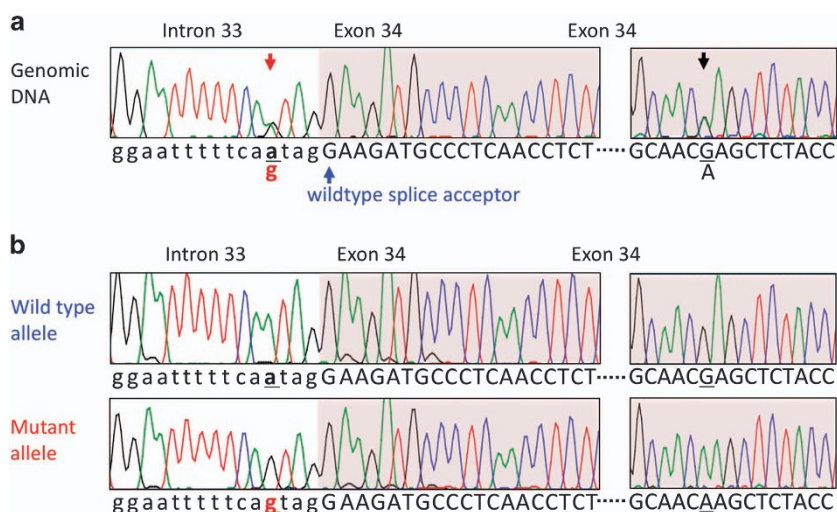Received 23 November 2017; revised 10 January 2018; accepted 10 January 2018

Figure 1. Single-nucleotide substitution in the patient's genome. (**a**) The electropherogram shows a sequence chromatogram of the genomic DNA of the patient with CHARGE syndrome. Rare substitutions in *CHD7*, namely, GRCh37, Chr8:61769000, NM_017780.3:c.7165−4A>G (red arrow) and the rs2272727 NM_017780.3:c.7356A/G polymorphism in exon 34 (black arrow), are shown. (**b**) The genomic linkage between the c.7165−4G substitution and the A polymorphism at c.7356 was confirmed using the cloned genomic sequence.

Figure 2. Single-nucleotide substitution-induced splicing abnormality in the patient. (**a**) *CHD7* cDNA was reverse-transcribed using total RNA from blood-derived lymphoblastoid cells and was then PCR-amplified, cloned and sequenced. Sequencing revealed the activation of a cryptic splice site three bases upstream of position c.7165 in exon 34. Furthermore, the variant RNA contains the UAG (TAG) (*) termination codon immediately after the cryptic splicing acceptor. (**b**) The schematic shows wild-type RNA and the consequences of the substitution at the −4 position of intron 33. In the wild-type gene, splicing occurs at the wild-type splice acceptor at the beginning of exon 34. The c.7165−4A>G substitution at the end of intron 33 activates a cryptic splice site immediately upstream of position c.7165 in exon 34. Moreover, all tested cDNA clones with the A polymorphism at c.7356 possessed a cryptic splice site.

Mutation Taster (http://www.mutationtaster.org/index.html), NNSPLICE (Berkeley Drosophila Genome Project) (http://fruitfly. org/seq_tools/splice.html), Human Splicing Finder (http://www. umd.be/HSF3/index.html) and MaxEntScan::score3ss (http://genes. mit.edu/burgelab/maxent/Xmaxentscan_scoreseq_acc.html) were used for predicting splice site alterations.

We detected a rare substitution that led to an alteration in the splicing acceptor and an ectopic premature termination within the thirty-third intron of *CHD7* (GRCh37, Chr8:61769000, NM_017780.3:c.7165−4A>G). This genomic substitution has been reported in patients with CS (HGMD number CS061264), but has been referred to as an "unclassified variant".[6,10] Exome trio analyses using leukocyte genomic DNA derived from the patient and his parents were performed. The NM_017780.3:c.7165−4A>G substitution was not detected in the parents; thus, this variant was considered to be a *de novo* mutation. Genomic DNA from the

patient was sequenced directly to confirm the NM_017780.3: c.7165−4A>G substitution (Figure 1a).

To identify the allele-specific sequence, the rs2272727 NM_017780.3:c.7356A>G polymorphism was used as a marker for each allele. Because the nucleotide position of rs2272727 is 195 bp away from the NM_017780.3:c.7165−4A>G substitution, the next-generation sequencing library could not determine the genomic linkage of these nucleotide variations. Thus, the genomic linkage between the c.7165−4G substitution and A polymorphism of rs2272727 (NM_017780.3: c.7356A/G) was confirmed by direct sequencing of genomic DNA sub-cloned into the TOPO vector (Life Technologies; Figure 1b). Exome sequencing indicated that the patient's mother possessed the G polymorphism of rs2272727, while the A polymorphism was assumed to be of paternal origin.

Moreover, to verify the single-nucleotide substitution-induced activation of cryptic splice sites, cDNA was reverse-transcribed using total RNA from lymphoblastoid cells derived from the patient's blood sample, and the PCR-amplified cDNA was subcloned and sequenced. Of 12 cDNA clones, 5 contained the cryptic splice acceptor site three bases upstream of the original site, and the remaining 7 had the wild-type (WT) sequence (Figure 2a). All of the WT cDNA clones had the G polymorphic variation at c.7356, whereas all of the mutated cDNA clones had an A residue at this site (Figure 2b). These data indicate that most of the RNA transcribed from the mutated allele possessed the cryptic splice acceptor. Furthermore, the available cryptic splice site prediction methods[11–13] were examined to evaluate the impact of this variation on splicing in silico. The NM_017780.3:c.7165−4A > G substitution was predicted to lead to "WT acceptor loss/acceptor decrease" and a "cryptic acceptor site" on the basis of model prediction, whereas the calculated values of each predictor were moderate compared to those of the typical splice acceptor site mutations associated with patients with CS, such as rs587783448, which are considered pathogenic (Supplementary Table S1).

We detected the NM_017780.3:c.7165−4A > G substitution four bases upstream of the WT splice acceptor site in exon 34 of CHD7. According to cDNA analyses, this substitution leads to the loss of the normal acceptor site and gain of a cryptic splicing acceptor three bases upstream of the WT splice acceptor site. Moreover, the variant RNA contains a termination codon UAG (TAG) immediately after the cryptic splicing acceptor, leading to a loss of 610 C-terminal amino acids from the full-length CHD7 protein (2,998 amino acids). The WT exons at positions 34–38 encode a nuclear localization signal and the BRK domain, a protein module associated with the CHROMO domain and DEAD/DEAH box helicase domain.[14,15] More than 20 nonsense and frameshift mutations have been reported in patients with CS in this genomic region.[8,10] Notably, de novo heterozygous frameshift mutations in exon 38 (last exon), which lead to the C-terminal truncation of CHD7 (1–2,988), are reported to be pathogenic.[10,16,17] Therefore, we propose that premature termination of CHD7 due to the NM_017780.3:c.7165−4A > G substitution, adversely affects normal growth and development.

Sequencing analyses of individual cDNA clones from the patient's RNA indicated that the activity of nonsense-mediated mRNA decay is probably subtle, presumably because of extremely long 3′ untranslated regions.[18–20] To date, five patients with CS, including the patient in our study, have been found to harbor this substitution,[6,10] whereas this substitution is not found in the general population. In previous reports,[6,10] neither genomic DNA from parents nor RNA from patients was examined for the presence of cryptic splicing sites. Therefore, the NM_017780.3: c.7165−4A > G substitution has been considered to be an "unclassified variant". This report is the first report of a laboratory-based assessment of this splicing variation and confirmation of the pathogenicity of the NM_017780.3:c.7165− 4A > G substitution. Although reliable predictor systems have been developed, the prediction of some atypical sequence variations in border splice regions is unclear. The NM_017780.3: c.7165−4A > G substitution did not fit with the consensus sites that disrupt the splicing code well. The prediction scores of the NM_017780.3:c.7165−4A > G substitution did not show marked changes compared to a typical splice site substitution, NM_017780 c.5666-2A > C (Supplementary Table S1). In addition to the NM_017780.3:c.7165−4A > G substitution, 17 mutations at the exon/intron junction within CHD7 are also considered "unclassified variants" in the CHD7 database. These substitutions are located at 1–11 base pairs from the exon/intron junction and may therefore include additional pathogenic variants.

In conclusion, we report a case of CS in a 7-year-old Japanese boy with a pathogenic de novo CHD7 mutation at an atypical splice site, which was previously described as an "unclassified variant". Experimental assessment of splicing variants is critical and contributes to the clinical diagnosis of certain unclassified mutations.

Availability of data and material
The data and materials described in this report are available on request. The identified mutation will be uploaded on the ClinVar website https://www.ncbi.nlm.nih.gov/clinvar/ (the accession SCV000599799). Supplementary Information (Supplementary Table S1; NM_017780.3:c.7165−4A > G evaluation using a splice site predictor) is available on the Human Genome Variation website.

HGV DATABASE
The relevant data from this Data Report are hosted at the Human Genome Variation Database at http://dx.doi.org/10.6084/m9.figshare.hgv.1890.

COMPETING INTERESTS
The authors declare no conflict of interest.

PUBLISHER'S NOTE
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES
1 Pagon RA, Graham JM Jr, Zonana J, Yong SL. Coloboma, congenital heart disease, and choanal atresia with multiple anomalies: CHARGE association. J Pediatr 1981; 99: 223–227.
2 Blake KD, Davenport SL, Hall BD, Hefner MA, Pagon RA, Williams MS et al. CHARGE association: an update and review for the primary pediatrician. Clin Pediatr (Phila) 1998; 37: 159–173.
3 Källén K, Robert E, Mastroiacovo P, Castilla EE, Källén B. CHARGE association in newborns: a registry-based study. Teratology 1999; 60: 334–343.
4 Issekutz KA, Graham JM Jr, Prasad C, Smith IM, Blake KD. An epidemiological analysis of CHARGE syndrome: preliminary results from a Canadian study. Am J Med Genet A 2005; 133A: 309–317.
5 Vissers LE, van Ravenswaaij CM, Admiraal R, Hurst JA, de Vries BB, Janssen IM et al. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. Nat Genet 2004; 36: 955–957.
6 Jongmans MC, Admiraal RJ, van der Donk KP, Vissers LE, Baas AF, Kapusta L et al. CHARGE syndrome: the phenotypic spectrum of mutations in the CHD7 gene. J Med Genet 2006; 43: 306–314.
7 Lalani SR, Safiullah AM, Fernbach SD, Harutyunyan KG, Thaller C, Peterson LE et al. Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. Spectrum of CHD7 mutations in 110 individuals with CHARGE syndrome and genotype-phenotype correlation. Am J Hum Genet 2006; 78: 303–314.
8 Janssen N, Bergman JE, Swertz MA, Tranebjaerg L, Lodahl M, Schoots J et al. Mutation update on the CHD7 gene involved in CHARGE syndrome. Hum Mutat 2012; 33: 1149–1160.
9 Higasa K, Miyake N, Yoshimura J, Okamura K, Niihori T, Saitsu H et al. Human genetic variation database, a reference database of genetic variations in the Japanese population. J Hum Genet 2016; 61: 547–553.
10 Bartels CF, Scacheri C, White L, Scacheri PC, Bale S. Mutations in the CHD7 gene: the experience of a commercial laboratory. Genet Test Mol Biomarkers 2010; 14: 881–891.

4

11 Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comp Biol* 1997; **4**: 311–323.

12 Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004; **11**: 377–394.

13 Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acid Res* 2009; **37**: e67.

14 Allen MD, Religa TL, Freund SM, Bycroft M. Solution structure of the BRK domains from CHD7. *J Mol Biol* 2007; **371**: 1135–1140.

15 Daubresse G, Deuring R, Moore L, Papoulas O, Zakrajsek I, Waldrip WR *et al*. The *Drosophila* kismet gene is related to chromatin-remodeling factors and is required for both segmentation and segment identity. *Development* 1999; **126**: 1175–1187.

16 Vuorela P, Ala-Mello S, Saloranta C, Penttinen M, Pöyhönen M, Huoponen K *et al*. Molecular analysis of the CHD7 gene in CHARGE syndrome: identification of 22 novel mutations and evidence for a low contribution of large CHD7 deletions. *Genet Med* 2007; **9**: 690–694.

17 Wincent J, Holmberg E, Strömland K, Soller M, Mirzaei L, Djureinovic T *et al*. CHD7 mutation spectrum in 28 Swedish patients diagnosed with CHARGE syndrome. *Clin Genet* 2008; **74**: 31–38.

18 Toma KG, Rebbapragada I, Durand S, Lykke-Andersen J. Identification of elements in human long 3′ UTRs that inhibit nonsense-mediated decay. *RNA* 2015; **22**: 887–897.

19 Ge Z, Quek BL, Beemon KL, Hogg JR. Polypyrimidine tract binding protein 1 protects mRNAs from recognition by the nonsense-mediated mRNA decay pathway. *Elife* 2016; **5**: e11155.

20 Popp MW, Maquat LE. Leveraging rules of nonsense-mediated mRNA decay for genome engineering and personalized medicine. *Cell* 2016; **165**: 1319–1322.

Supplemental Information for this article can be found on the Human Genome Variation website (http://www.nature.com/hgv).