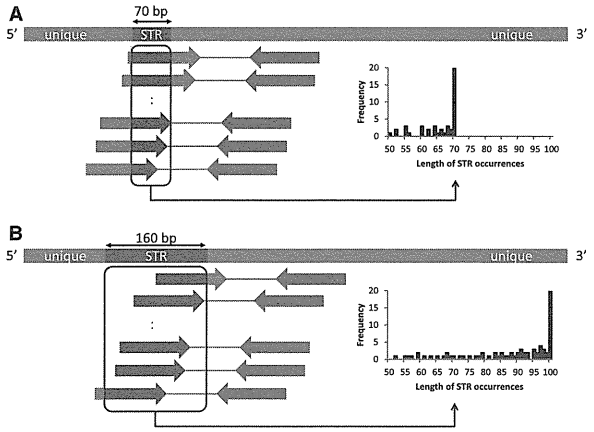

Fig. 2. Select positions where STR occurrences are expanded significantly. (**A**) We generate the frequency distribution of lengths of STR occurrences in paired-end reads. This picture shows the case of a 70-bp long STR. The histogram of the frequency distribution peaks at 70 bp. (**B**) When the STR is 160-bp long, the distribution has a significant peak at 100 bp. We test if the peak is a significant outlier in the frequency distribution using the Smirnov–Grubbs' test

threshold according to the Smirnov–Grubbs' test. For example, the $P<5\times10^{-9}$ when the t-score is >5.27.

(4) We consider ~10 million non-overlapping regions of length 300 bp (the average insert size of paired-end reads) in the human genome. We perform multiple hypothesis testing using the Bonferroni correction to test if each 300-bp region has a significant STR expansion in the case sample at a significance level of 5% divided by 10 million (i.e. 5×10^{-9}). We select positions such that $P<5\times10^{-9}$ in the case sample but no 100-bp STR occurrences are present in any of the control samples. We can relax the condition to consider more candidates with less evidence.

Sequencing candidate STR positions

SMRT^{TM} sequencing of expanded STRs is performed using information on the boundaries of individual STR positions.

3.1 A rare STR significantly expanded in the case sample

To demonstrate the effectiveness of this approach, we first examined a well-characterized case sample, SCA31 (Sato *et al.*, 2009), which contains long expansions of two STRs, (AAAATAGAAT) repeat and (AATGG) repeat, in the introns of genes BEAN1 and TK2 (Chr.16 66,524,303 in hg19), where the reference genome has an (AAAAT) repeat.

We resequenced the genome of a sample from an individual whose parent is a case of SCA31 using an Illumina HiSeq2000 (Supplementary Table S4). All primary sequencing data of the SCA31 sample will be made available under controlled access through the DNA Databank of Japan (DDBJ; accession number JGAS00000000002). We examined whether we could find these STRs with no prior information. We applied the *ab initio* procedure to SCA31 as the case sample, and NA12877, NA12878 and NA18507 as control samples (Fig. 3A). Our procedure detected only one STR; AAAATAGAAT ($P=1.07\times10^{-19}$).

Figure 3B shows the frequency distributions of the (AAAATAGAAT) repeat, supporting the presence of long occurrences of the STR in SCA31 and the absence of long occurrences of length >60 bp in the other control samples. Supplementary Figure S3A shows the distributions of the (AATGG) repeat, but the difference between SCA31 and the other samples was unclear because the (AATGG) repeat is enriched in human centromeres (Grady *et al.*, 1992). Therefore, our *ab initio* analysis suggests that long occurrences of the (AAAATAGAAT) repeat characterize SCA31, consistent with reported observations (Sato *et al.*, 2009). Arguably, we could detect the (AAAATAGAAT) repeat as an approximate (AAAAT) repeat because the last half, AGAAT, is identical to (AAAAT), except for the second base G; therefore, we analyzed the frequency distribution of the (AAAAT) repeat to determine the remarkable expansion of the (AAAAT) repeat in SCA31. This failed due to numerous long instances of the (AAAAT) repeat in all samples (Supplementary Fig. S3B). This example indicates the importance of looking at STRs of repeat units longer than 2–6-base units, to determine expansions of STRs associated with cases.

We also examined the frequency distributions of other well-characterized repeats, such as the (GGGTTA) repeat in telomeres (Supplementary Fig. S3C), (CAG) repeat encoding polyglutamine stretches in protein coding regions (La Spada *et al.*, 1991; The Huntington's Disease Collaborative Research Group, 1993; Walker, 2007 and Supplementary Fig. S4A), (CCTG) repeat associated with myotonic dystrophy type 2 (DM2; Liquori *et al.*, 2001 and Supplementary Fig. S4B) and (ATTCT) repeat associated with spinocerebellar ataxia type 10 (SCA10; Matsuura *et al.*, 2000 and Supplementary Fig. S4C). For the last three repeats, no significant differences were detected between SCA31 and the three control samples, suggesting that these three repeats are not associated with SCA31.

Using paired-end reads with AAAATAGAAT repeats at their 5′ ends and uniquely mapped reads at their 3′ ends, we could determine the 3′ end of the insertion. Figure 4A shows how we locate a ~2.5–3.8 kb insertion of the repeat associated with the SCA31 sample (Sato *et al.*, 2009).

We sequenced the repeat region in 11 SCA31 samples using SMRT^{TM} sequencing. We designed a pair of PCR primers around the candidate repeat region in the SCA31 sample the right boundary of which could be determined. As illustrated in Figure 4A, we could identify the right boundary in the reference genome because the right ends of many paired-end reads mapped to the downstream region of the right boundary, whereas the left ends did not. We could sequence the candidate repeat region. Supplementary Table S6 presents the statistics of filtered subreads, corrected subreads and assembled contigs. Previously, Sato *et al.* estimated a 2.5–3.8 kb insertion of the following form for an SCA31 sample (Sato *et al.*, 2009):

$$\mathrm{TCACTAAAA(TAGAA)}_i\mathrm{(TGGAA)}_j$$
$$\mathrm{(TAGAA|TGGAA|TGGGAA)}_k$$
$$\mathrm{(TAGAA)}_l\mathrm{(TAAAATAGAA)}_m,$$

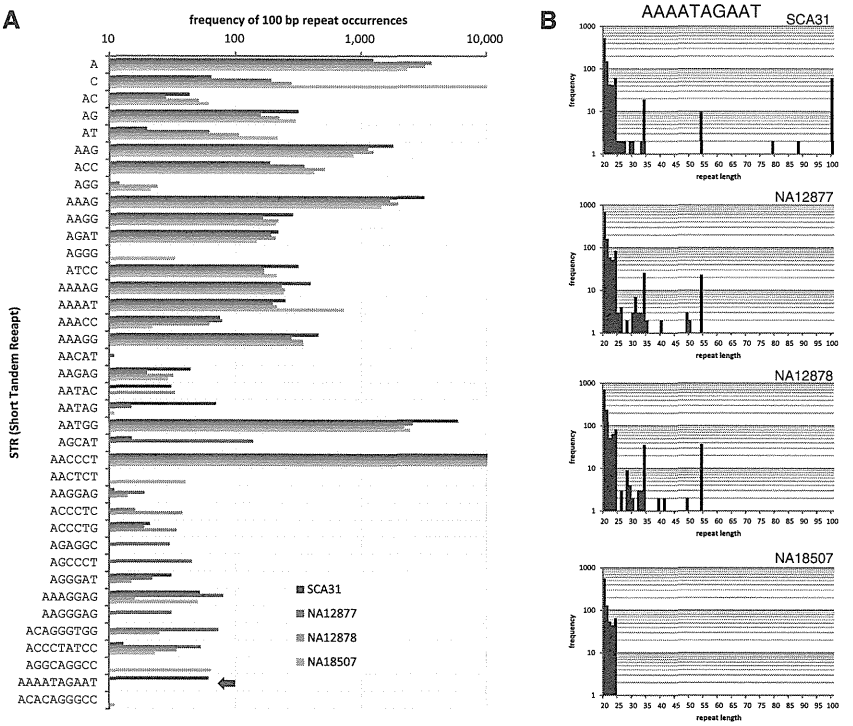

Fig. 3. Sensing expanded STRs associated with SCA31. (**A**) Frequencies of 100-bp STRs that have >10 occurrences in one of SCA31, NA12877, NA12878 or NA18507. For example, the arrow in the second lowest row shows that the (AAAATAGAAT) repeat is expanded only in SCA31. Our *ab initio* procedure analyzes this bar chart and selects STRs that are significantly abundant in the case sample (e.g., SCA31) but absent in all of the control samples. The bar chart is also useful for confirming the abundance of (AATGG) and (AACCCT) repeats, equivalent to the (GGGTTA) repeat, where the former and latter motifs are known to be enriched in centromeres and telomeres, respectively. (**B**) Frequency distributions of the (AAAATAGAAT) repeat. SCA31 has many 100-bp occurrences, whereas no occurrences of length >55 bp were observed in NA12877, NA12878 and NA18507

where (TAGAA | TGGAA | TGGGAA)$_k$ is a series of k occurrences of TAGAA, TGGAA and TGGGAA. In their sample, they determined that $i=2$, $k=10$ and $l=46$, but left j and m undetermined because both appeared to be extremely long. In our 11 SCA31 samples, we could determine the values of j and m. We found that the numbers of individual repeats varied markedly ($i=1\sim2, j=220\sim321, k=9\sim13, l=42\sim78$ and $m=90\sim118$) and the insertion size ranged from 2350 to 3088 b (Fig. 4C and Supplementary Table S5), demonstrating the instability of the STR expansion in SCA31. In particular, two STRs, (TAGAA)j and (TAAAA TAGAA)m, form ~90% of the entire repeat expansion, and the values of j and m are positively correlated (correlation coefficient $r=0.70$), implying that these two values are the determinants of the instability of the repeat expansions in SCA31 (Fig. 4D). In all samples, the repeat expansion was present in one allele, but was absent in the other. Note that the numbers of STR units might not be exact because PCR for repeat regions can introduce more replication errors than those produced by bacterial DNA replication (Loomis *et al.*, 2013).

3.2 Common STRs significantly expanded in the case sample

We also applied our procedure to the SCA31 data, and examined common STRs, AAAG, ATCC, AAAAG, AATAG and AATGG, present in both the case and control samples but significantly expanded in the case sample. We identified STR expansions at 11 genomic locations that were significantly expanded in the case sample ($P<5\times10^{-19}$ and Supplementary Fig. S5). We then used SMRT^{TM} sequencing to confirm the four expanded STRs in the case sample that were significantly longer than the corresponding STR occurrences in the reference genome (Fig. 5 and Supplementary Fig. S6). No false-positive expansions were found in this experiment, suggesting that the false-positive rate of the procedure is generally low.

4 DISCUSSION

STRs in personal genomes remain largely uncharacterized. We proposed a novel method for listing long approximate STRs with mutations in personal genomes using a massive number of short reads of length ~100 bp. Here, we discuss some situations in which detecting a long expansion of STRs specific to disease samples is inherently problematic. As genomic regions of GC content >70% are difficult to cover with an ample number of Illumina reads, our method is unlikely to detect long expansions of STRs with high GC contents. STRs in reads originating in centromeres, telomeres or retrotransposons are too numerous to map to unique genomic positions. As illustrated in

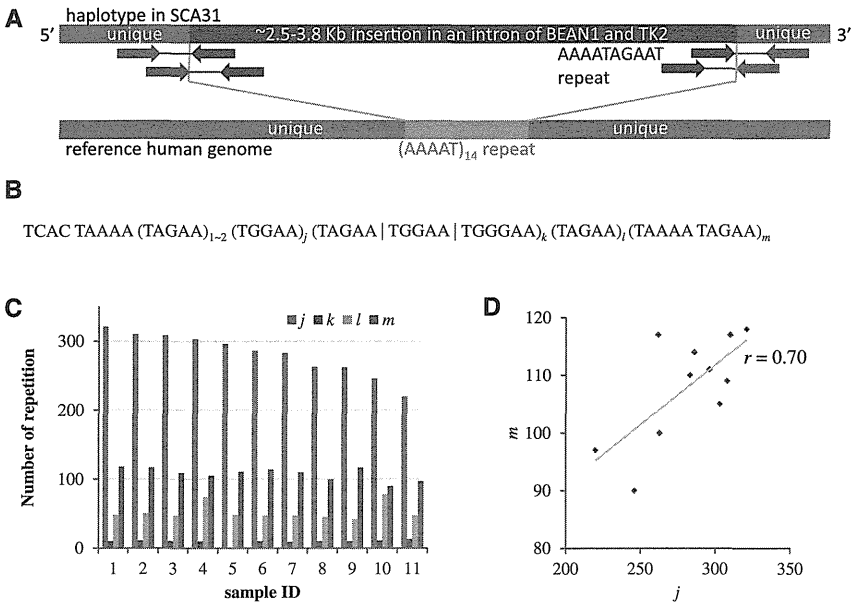

Fig. 4. Locating and sequencing expanded STRs associated with SCA31. (**A**) A real example from SCA31. One haplotype contains a ~2.5–3.8 kb insertion at Chr.16 66 524 303 in hg19 in an intron of BEAN1 and TK2. The right boundary of the insertion could be identified using paired-end reads with AAAATAGAAT repeats at their left ends and uniquely mapped reads at their right ends. The lower bar illustrates the reference genome (hg19) with an AAAAT repeat. (**B**) A form of expanded repeat associated with SCA31 samples. The values of *i*, *j*, *l* and *m* vary in the individual SCA31 samples. (**C**) We determined the values of *i*, *j*, *l* and *m* in 11 SCA31 samples using SMRT^{TM} sequencing. This shows that ~90% of the repeat expansion are $(TAGAA)_j$ and $(TAAAA\ TAGAA)_m$. (**D**) The values of ***j*** and ***m*** are positively correlated ($r = 0.70$). These two values are the determinants of the instability of the repeat expansions in SCA31

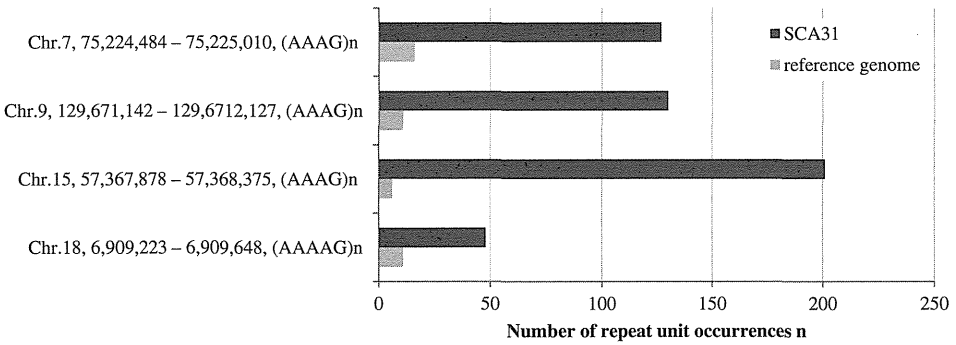

Fig. 5. Sizes of the common STRs, (AAAG)*n* and (AAAAG)*n*, at four genomic positions in the SCA31 sample and reference genome. Note that individual STR occurrences are significantly expanded in the SCA31 sample. The PCR primers used for amplifying individual regions and the sequences of amplicons can be found in Supplementary Figure S6

Supplementary Figure S3, massive numbers of long expansions of these STRs can be found in any sample.

We also presented an *ab initio* procedure for detecting significant expansions of STRs in case samples that are absent in control samples via comparisons between the frequency distributions of STRs in case and control samples. We demonstrated the potential applicability of this method using three publicly available control samples. To exploit this approach, however, constructing a large-scale database of the frequency distributions of STRs collected from a number of control samples is necessary.

The variety of expanded STRs of length >1 kb in disease remains unexplored. Also, examining whether expansions of STRs are more pronounced in germline and somatic cells would be intriguing. Thus, after locating STRs, sequencing expanded STRs is a promising direction of study. For this purpose, SMRT^{TM} sequencing enables the sequencing DNA fragments averaging ~5 kb long as of 2013. Using SMRT^{TM} sequencing, we were able to determine a divergent set of 2.3–3.1 kb STR sequences in 11 SCA31 samples, showing the instability of STR expansions. Analysis of the stability of STR expansions

in germline and somatic cells of a specific disease might eventually lead to the recognition of a functional role of STRs.

In the near future, the typical lengths of short reads in the majority of commercial sequencers should increase to 150–500 bases. Our method is ready to process longer reads in a straightforward manner. Furthermore, our method was designed so that it could output STRs of repeat units of any length, and we presented an illustrative case in which detecting STRs of a 10-base repeat unit from an SCA31 sample was essential. Our program will serve as a valuable tool for discovering unknown STRs in a variety of diseases, even with future advances in sequencing technology.

ACKNOWLEDGEMENTS

The authors are grateful to an anonymous reviewer for suggesting a way of detecting expansions of common STRs and for many valuable comments on the manuscript.

Funding: Grant-in-Aid for Scientific Research on Innovative Areas (22129008, 221S0002 to S.M.) (in part) from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT) and by the Global COE program (Deciphering Biosphere from Genome Big Bang) to S.M. from the MEXT.

Conflict of Interest: none declared.

REFERENCES

Ballantyne,K.N. *et al.* (2010) Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. *Am. J. Hum. Genet.*, **87**, 341–353.

Benson,G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.*, **27**, 573–580.

Brook,J.D. *et al.* (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. *Cell*, **69**, 385.

Conrad,D.F. *et al.* (2011) Variation in genome-wide mutation rates within and between human families. *Nat.Genet.*, **43**, 712–714.

DeJesus-Hernandez,M. *et al.* (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*, **72**, 245–256.

DePristo,M.A. *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*, **43**, 491–498.

Domanic,N.O. and Preparata,F.P. (2007) A novel approach to the detection of genomic approximate tandem repeats in the Levenshtein metric. *J. Comput. Biol.*, **14**, 873–891.

Eid,J. *et al.* (2009) Real-time DNA sequencing from single polymerase molecules. *Science*, **323**, 133–138.

Grady,D.L. *et al.* (1992) Highly conserved repetitive DNA sequences are present at human centromeres. *Proc. Natl Acad. Sci. USA*, **89**, 1695–1699.

Gymrek,M. *et al.* (2012) lobSTR: a short tandem repeat profiler for personal genomes. *Genome Res.*, **22**, 1154–1162.

Jorda,J. and Kajava,A.V. (2009) T-REKS: identification of Tandem REpeats in sequences with a K-meanS based algorithm. *Bioinformatics*, **25**, 2632–2638.

Kobayashi,H. *et al.* (2011) Expansion of intronic GGCCTG hexanucleotide repeat in NOP56 causes SCA36, a type of spinocerebellar ataxia accompanied by motor neuron involvement. *Am. J. Hum. Genet.*, **89**, 121–130.

Kolpakov,R. *et al.* (2003) mreps: Efficient and flexible detection of tandem repeats in DNA. *Nucleic Acids Res.*, **31**, 3672–3678.

Kong,A. *et al.* (2012) Rate of *de novo* mutations and the importance of father's age to disease risk. *Nature*, **488**, 471–475.

Kremer,E.J. *et al.* (1991) Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. *Science*, **252**, 1711–1714.

La Spada,A.R. *et al.* (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature*, **352**, 77–79.

Li,H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Lim,K.G. *et al.* (2013) Review of tandem repeat search tools: a systematic approach to evaluating algorithmic performance. *Brief. Bioinform.*, **14**, 67–81.

Liquori,C.L. *et al.* (2001) Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science*, **293**, 864–867.

Loomis,E.W. *et al.* (2013) Sequencing the unsequenceable: expanded CGG-repeat alleles of the fragile X gene. *Genome Res.*, **23**, 121–128.

Lupski,J.R. (2007) Genomic rearrangements and sporadic disease. *Nat. Genet.*, **39**, S43–S47.

Mahadevan,M. *et al.* (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3′ untranslated region of the gene. *Science*, **255**, 1253–1255.

Main,M.G. (1989) Detecting leftmost maximal periodisities. *Discrete Appl. Math.*, **25**, 145–153.

Main,M.G. and Lorentz,R.J. (1984) an O(n log n) algorithm for finding all repetitions in a string. *J. Algorithm.*, 422–432.

Matsuura,T. *et al.* (2000) Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. *Nat. Genet.*, **26**, 191–194.

Mirkin,S.M. (2007) Expandable DNA repeats and human disease. *Nature*, **447**, 932–940.

Mudunuri,S.B. and Nagarajaram,H.A. (2007) IMEx: Imperfect Microsatellite Extractor. *Bioinformatics*, **23**, 1181–1187.

Orr,H.T. (2011) FTD and ALS: genetic ties that bind. *Neuron*, **72**, 189–190.

Pellegrini,M. *et al.* (2010) TRStalker: an efficient heuristic for finding fuzzy tandem repeats. *Bioinformatics*, **26**, i358–i366.

Renton,A.E. *et al.* (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron*, **72**, 257–268.

Sato,N. *et al.* (2009) Spinocerebellar ataxia type 31 is associated with "inserted" penta-nucleotide repeats containing (TGGAA)n. *Am. J. Hum. Genet.*, **85**, 544–557.

Sherman,S.L. *et al.* (1985) Further segregation analysis of the fragile X syndrome with special reference to transmitting males. *Hum Genet*, **69**, 289–299.

The Huntington's Disease Collaborative Research Group. (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. *Cell*, **72**, 971–983.

Verkerk,A.J. *et al.* (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*, **65**, 905–914.

Walker,F.O. (2007) Huntington's disease. *Lancet*, **369**, 218–228.

Warner,J.P. *et al.* (1996) A general method for the detection of large CAG repeat expansions by fluorescent PCR. *J. Med. Genet.*, **33**, 1022–1026.

Wexler,Y. *et al.* (2005) Finding approximate tandem repeats in genomic sequences. *J. Comput. Biol.*, **12**, 928–942.

Wojciechowska,M. and Krzyzosiak,W.J. (2011) Cellular toxicity of expanded RNA repeats: focus on RNA foci. *Hum. Mol. Genet.*, **20**, 3811–3821.

Please cite this article in press as: Takahashi et al., *ERBB4* Mutations that Disrupt the Neuregulin-ErbB4 Pathway Cause Amyotrophic Lateral Sclerosis Type 19, The American Journal of Human Genetics (2013), http://dx.doi.org/10.1016/j.ajhg.2013.09.008

REPORT

ERBB4 Mutations that Disrupt the Neuregulin-ErbB4 Pathway Cause Amyotrophic Lateral Sclerosis Type 19

Yuji Takahashi,[1] Yoko Fukuda,[1] Jun Yoshimura,[2] Atsushi Toyoda,[3] Kari Kurppa,[4,5] Hiroyoko Moritoyo,[6] Veronique V. Belzil,[7] Patrick A. Dion,[7,8] Koichiro Higasa,[2] Koichiro Doi,[2] Hiroyuki Ishiura,[1] Jun Mitsui,[1] Hidetoshi Date,[1] Budrul Ahsan,[1] Takashi Matsukawa,[1] Yaeko Ichikawa,[1] Takashi Moritoyo,[6] Mayumi Ikoma,[9] Tsukasa Hashimoto,[9] Fumiharu Kimura,[10] Shigeo Murayama,[11] Osamu Onodera,[12] Masatoyo Nishizawa,[12] Mari Yoshida,[13] Naoki Atsuta,[14] Gen Sobue,[14] JaCALS,[15] Jennifer A. Fifita,[16,17,18] Kelly L. Williams,[16,17,18] Ian P. Blair,[16,17,18] Garth A. Nicholson,[16,17] Paloma Gonzalez-Perez,[19] Robert H. Brown, Jr.,[19] Masahiro Nomoto,[6] Klaus Elenius,[4,20] Guy A. Rouleau,[7,21,22] Asao Fujiyama,[3] Shinichi Morishita,[2] Jun Goto,[1] and Shoji Tsuji[1,23,*]

Amyotrophic lateral sclerosis (ALS) is a devastating neurological disorder characterized by the degeneration of motor neurons and typically results in death within 3–5 years from onset. Familial ALS (FALS) comprises 5%–10% of ALS cases, and the identification of genes associated with FALS is indispensable to elucidating the molecular pathogenesis. We identified a Japanese family affected by late-onset, autosomal-dominant ALS in which mutations in genes known to be associated with FALS were excluded. A whole- genome sequencing and parametric linkage analysis under the assumption of an autosomal-dominant mode of inheritance with incomplete penetrance revealed the mutation c.2780G>A (p. Arg927Gln) in *ERBB4*. An extensive mutational analysis revealed the same mutation in a Canadian individual with familial ALS and a de novo mutation, c.3823C>T (p. Arg1275Trp), in a Japanese simplex case. These amino acid substitutions involve amino acids highly conserved among species, are predicted as probably damaging, and are located within a tyrosine kinase domain (p. Arg927Gln) or a C-terminal domain (p. Arg1275Trp), both of which mediate essential functions of ErbB4 as a receptor tyrosine kinase. Functional analysis revealed that these mutations led to a reduced autophosphorylation of ErbB4 upon neuregulin-1 (NRG-1) stimulation. Clinical presentations of the individuals with mutations were characterized by the involvement of both upper and lower motor neurons, a lack of obvious cognitive dysfunction, and relatively slow progression. This study indicates that disruption of the neuregulin-ErbB4 pathway is involved in the pathogenesis of ALS and potentially paves the way for the development of innovative therapeutic strategies such using NRGs or their agonists to upregulate ErbB4 functions.

Amyotrophic lateral sclerosis (ALS) is a devastating neurological disorder in which the degeneration of motor neurons leads to progressive weakness and wasting of limb, bulbar, and respiratory muscles. Familial ALS (FALS) comprises 5%–10% of ALS cases, and the remaining cases are simplex cases of ALS (SALS). To date, more than 20 genes have been shown to be associated with ALS,[1] and these account for 75% of FALS and 14% of SALS cases.[2] Mutations that are found in FALS-associated genes but that are also identified in individuals with SALS are considered mutations with reduced penetrance or de novo mutations. Further discovery of genes associated with FALS is indispensable to elucidating the molecular backgrounds of both FALS and SALS.

Identification of genes associated with familial diseases has been accomplished through identification of the disease loci on the chromosomes by linkage analysis of large pedigrees and subsequent positional cloning of the genes. The majority of the FALS pedigrees, however, are not large and do not have multiple affected members as a result of the poor prognosis of the disease and the late age of onset, which makes it difficult to sufficiently narrow the candidate regions by linkage analyses and means that it takes a tremendous effort to identify the genes associated with FALS. The recent development of massively parallel

[1]Department of Neurology, Graduate School of Medicine, The University of Tokyo, Tokyo 113-8655, Japan; [2]Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo, Chiba 277-8561, Japan; [3]Comparative Genomics Laboratory, National Institute of Genetics, Shizuoka 411-8540, Japan; [4]Department of Medical Biochemistry and Genetics, University of Turku, Turku, 20014, Finland; [5]Turku Doctoral Programme of Biomedical Sciences, Turku, 20520, Finland; [6]Department of Neurology and Clinical Pharmacology, Ehime University Hospital, Ehime 791-0295, Japan; [7]Research Center, Centre Hospitalier Universitaire Sainte-Justine, Université de Montreal, Montreal, QC, H3T 1C5, Canada; [8]Department of Pathology and Cellular Biology, Université de Montréal, Montreal, QC H3T 1C5, Canada; [9]Department of Neurology, National Hospital Organization Ehime National Hospital, Ehime 791-0281, Japan; [10]Division of Neurology, First Department of Internal Medicine, Osaka Medical College, Osaka 569-8686, Japan; [11]Department of Neurology and Neuropathology and Brain Bank for Aging Research, Tokyo Metropolitan Geriatric Hospital and Institute of Gerontology, Tokyo 173-0015, Japan; [12]Department of Neurology, Brain Research Institute, Niigata University, Niigata 951-8520, Japan; [13]Department of Neuropathology, Institute for Medical Science of Aging, Aichi Medical University, Aichi 480-1195, Japan; [14]Department of Neurology, Nagoya University Graduate School of Medicine, Nagoya 466-8560, Japan; [15]Japanese Consortium for Amyotrophic Lateral Sclerosis Research; [16]Northcott Neuroscience Laboratory, ANZAC Research Institute, Sydney, New South Wales 2139, Australia; [17]Sydney Medical School, University of Sydney, New South Wales 2006, Australia; [18]Austrian School of Medicine, Macquarie University, Sydney, New South Wales 2109, Australia; [19]Department of Neurology, University of Massachusetts Medical School, Worcester, MA 01655-0318, USA; [20]Department of Oncology, Turku University Hospital, Turku 20521, Finland; [21]Montreal Neurological Institute, McGill University, Montreal, QC H3A 2B4, Canada; [22]Department of Neurology and Neurosurgery, McGill University, Montreal, QC H3A 2B4, Canada; [23]Medical Genome Center, The University of Tokyo Hospital, The University of Tokyo, Tokyo 113-8655, Japan
*Correspondence: tsuji@m.u-tokyo.ac.jp
http://dx.doi.org/10.1016/j.ajhg.2013.09.008.

Please cite this article in press as: Takahashi et al., *ERBB4* Mutations that Disrupt the Neuregulin-ErbB4 Pathway Cause Amyotrophic Lateral Sclerosis Type 19, The American Journal of Human Genetics (2013), http://dx.doi.org/10.1016/j.ajhg.2013.09.008

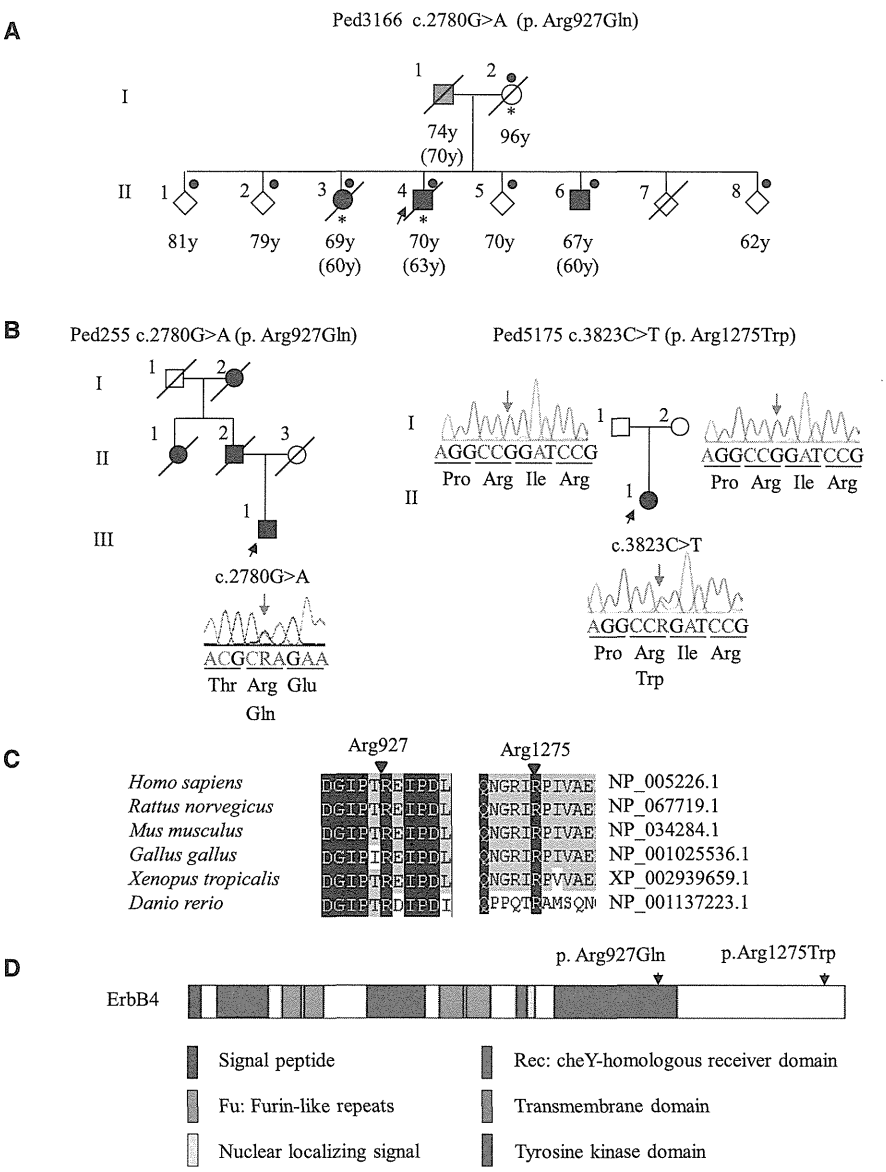

Figure 1. Pedigrees of ALS and Characterization of Mutations

(A) Pedigree charts of the index family. Filled symbols indicate affected individuals. The arrow indicates the proband. For confidentiality purposes, all unaffected siblings are indicated by diamonds. Dots or asterisks indicate individuals included in the linkage study or WGS, respectively. Age at present or age at death is shown under each individual, and ages at onset are shown in parentheses. The box with gray shading indicates that the individual's clinical information obtained from the family members strongly supports the diagnosis of ALS, although detailed neurological evaluations have not been conducted for this individual.

(B) Additional Canadian (Ped255) and Japanese (Ped5175) pedigrees with *ERBB4* mutations. The electropherograms of mutational data are shown beside each member. Nucleotide colors correspond to the colors in the electropherograms. The amino acids are designated below the nucleotide sequences. The blue arrows indicate the nucleotide positions of the mutations. In the electropherograms (Ped5175), nucleotide sequences of the reverse complementary strand are shown.

(C) Amino acid conservation. The amino acids Arg927 and Arg1275 are highly conserved among species.

(D) The protein structure along with the locations of amino acid substitutions are shown; amino acid substitutions are indicated by arrows. The amino acid substitution p. Arg927Gln resides in the tyrosine kinase domain, which mediates the key functions of ErbB4. The amino acid substitution p. Arg1275Trp resides in the C-terminal domain in the vicinity of multiple phosphorylation sites, which mediate downstream signaling pathways.

sequencing technologies has allowed us to overcome the difficulty by means of whole-genome sequencing (WGS) or exome analysis.

We identified a Japanese family with three affected siblings presenting with late-onset ALS (Figure 1A and Table 1). The familial history indicated that the mode of inheritance is probably an autosomal-dominant one. Mutational analysis of the proband (II-4) employing direct nucleotide sequence analysis, a microarray-based resequencing, or a repeat-primed PCR analysis excluded *SOD1*[MIM 147450], *ALS2*[MIM 606352], *DCTN1*[MIM 601143], *CHMP2B*[MIM 609512], *ANG*[MIM 105850], *TARDBP*[MIM 605078], *FUS*[MIM 137070] and *C9ORF72* [MIM 1614260] as the genes associated with FALS.[3,4] To identify a gene associated with FALS, we applied WGS in combination with a linkage analysis to the pedigree. Written informed consent was obtained from all the participants. This study was approved by the institutional review board at the University of Tokyo.

WGS was performed on three individuals (I-2, II-3 and II-4, as shown in Figure 1A) in the index pedigree. Paired-end DNA libraries were generated and subjected to massively parallel sequencing with a GAII Illumina Genome Analyzer in accordance with the manufacturer's instructions. The short read sequences obtained were aligned to the reference genome (NCBI37/hg19 assembly) via the Burrows-Wheeler Aligner.[5] Downstream analyses in which potential PCR duplicates were removed were processed with SAMtools.[6] Aligned reads were viewed on an Integrative Genomics Viewer.[7] Genomic sequence variations were identified with the SAMtools pileup command and annotated with Refseq, dbSNP135, 1000 Genomes, personal genome databases, the NHLBI GO Exome Sequencing Project (NHLBI-ESP) database, and an in-house variant database containing 41 whole genomes and 1,408 exomes in the Japanese population. The numbers of nonsynonymous variants that were identified in individuals I-2, II-3, and II-4 but that were not present in any of the

Please cite this article in press as: Takahashi et al., *ERBB4* Mutations that Disrupt the Neuregulin-ErbB4 Pathway Cause Amyotrophic Lateral Sclerosis Type 19, The American Journal of Human Genetics (2013), http://dx.doi.org/10.1016/j.ajhg.2013.09.008

Table 1. Clinical Characteristics of Affected Individuals

Pedigree Number	Pedigree 3166				Pedigree 255	Pedigree 5175
Ethnicity	Japanese				Canadian	Japanese
Inheritance	familial (autosomal dominant)				familial (autosomal dominant)	simplex
Mutation	c.2780G>A				c.2780G>A	c.3823C>T
Amino acid substitution	p. Arg927Gln				p. Arg927Gln	p. Arg1275Trp
Members	I-1	II-3	II-4 (proband)	II-6	III-3	II-1
Age at onset	70	60	63	60	67	45
Initial symptoms	bulbar	N.D.	upper limbs	respiration	upper limbs	upper limbs
Diagnostic criteria[a]	N.D.	N.D.	definite	definite	probable	probable
Progression	unable to walk after 3 years	ventilator -dependent after 5 years, locked-in state after 8 years	locked-in state after 5 years	ventilator- dependent after 1 year, locked-in state after 5 years	slow progression that significantly decelerated and finally stopped after 8 years	wheelchair- bound, MRS 1-2/5 in upper extremities after 5 years
Cognitive function	N.D.	N.D.	normal	normal	N.D.	normal
Age at death	74	69	70	66	N/A	N/A

Abbreviations are as follows: N.D., not described; MRS, manual muscle testing rating scale; and N/A, not applicable.
[a]El Escorial and Airlie House revised criteria.

databases (hereafter, variants not found in the databases are referred to as "novel") were 411, 404, and 382, respectively (Table S1). No novel nonsynonymous variants in genes known to be associated with FALS were included. Among the identified variants, 57 were identified both in the proband and in the affected sibling, but not in the mother, and were subjected to further analysis.

The individuals indicated by dots in Figure 1A were genotyped with Genome-Wide Human SNP Array 6.0 (Affymetrix). Linkage analysis and haplotype reconstruction were conducted with the pipeline software SNP-HiTLink[8] and Allegro version 2[9] under the assumption of an autosomal-dominant mode of inheritance and a disease-allele frequency of 0.000001. Parametric multipoint linkage analysis under the assumption of complete penetrance revealed three loci spanning 23.6 Mb on chromosomes 1, 6, and 13, having a maximum LOD score of 1.8 (Figure S1; penetrance = 1.0), and containing 88 annotated genes. However, no novel nonsynonymous variants were identified in the candidate regions. We then considered the possibility of reduced penetrance. When penetrance was reduced to 0.8 (Figure S1), seven additional loci had LOD scores > 0.7 and were thus shown to support linkage; these loci contained 809 annotated genes. Three heterozygous novel nonsynonymous variants were identified in these regions; among these variants, only c.2780G>A (p. Arg927Gln; dbSNP SubSNP ID ss831884245) substituting glutamine for arginine at codon 927 (p. Arg927Gln) in v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian) (*ERBB4* [MIM 600543; RefSeq accession number NM_005235.2]) was not present in 477 controls (Table S2). When we allowed further reduced penetrance, we identified 19 additional loci with LOD > 0; these loci contained 1,265 annotated genes. In these regions, we identified seven heterozygous novel nonsynonymous variants, among which three variants in *OR2D3* (RefSeq NM_001004684.1), *FTCD* (MIM 606806; RefSeq NM_206965.1), and *TJP2* (MIM 607709; RefSeq NM_001170414.2) were not present in 477 controls (Table S2). *OR2D3* is an olfactory receptor gene; the substituted amino acid in *OR2D3* is not conserved, and the substitution is predicted as benign by PolyPhen-2 analysis. *FTCD* and *TJP2* are associated with autosomal-recessive glutamate formiminotransferase deficiency (MIM 229100) and familial hypercholanemia (MIM 607748), respectively, and heterozygous carriers have not been described as exhibiting ALS. Taken together, the results pointed to c.2780G>A in *ERBB4* as the most likely pathogenic mutation.

We used a direct nucleotide sequence analysis method to conduct mutational analysis of *ERBB4* in 364 FALS and 818 SALS individuals by using an ABI 3100 sequencer and BigDye Terminator ver3.1 (Applied Biosystems). We used the ExonPrimer website to design oligonucleotide primers (Table S3). The mutation c.2780G>A was also identified in one Canadian FALS individual (Figure 1B). Unfortunately, DNA from other family members was not available to confirm segregation. To investigate a possibility that the c.2780G>A mutation identified in the Japanese and Canadian families is a common founder mutation, we compared the haplotypes with the c.2780G>A mutation in *ERBB4* of the Japanese and Canadian families (Figure S2). Different SNPs were observed 14 kbp and 5 kbp centromeric and telomeric to the mutation, respectively, indicating that disease haplotypes of the Japanese and Canadian families are different and that

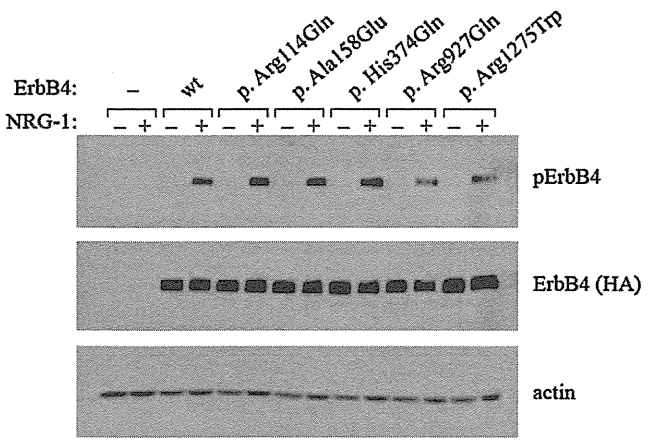

Figure 2. Functional Analysis of Wild-Type and Mutant ErbB4 upon Neuregulin-1 Stimulation
COS-7 cells transfected with an empty-vector control or plasmids encoding either wild-type (wt) or mutant HA-tagged ErbB4 (p. Arg114Gln, p. Ala158Glu, p. His374Gln, p. Arg927Gln, or p. Arg1275Trp) were stimulated with or without NRG-1, and the autophosphorylation activity of ErbB4 was analyzed by immunoblot analysis with antibodies against phospho-ErbB4 (Tyr1284) (Cell Signaling) and HA tag (Abcam), respectively. For loading controls, immunoblotting was performed with an anti-actin antibody (Santa Cruz Biotechnology). Three amino acid substitutions, including p.Arg114Gln, p.Ala158Glu, and p.His374Gln (rs760369), identified through mutational analysis of FALS and SALS individuals, were included in autophosphorylation assay. The substitutions p.Arg114Gln and p.Ala158Glu were not considered to be relevant to ALS because neither recurrence nor cosegregation was confirmed.

mutation occurred independently. We identified a de novo mutation of c.3823C>T (dbSNP SubSNP ID ss831884246), substituting tryptophan for arginine at codon 1275 (p. Arg1275Trp), in a Japanese SALS individual (Figure 1B) in whom a biological parent-descendant relationship was confirmed (Table S4) by the PLINK[10] algorithm. These mutations were neither present in the 477 Japanese controls nor registered in the in-house database containing 41 whole genomes and 1408 exomes, the 1000 Genomes database, or the NHLBI-ESP database, containing 6503 exomes. Furthermore, c.2780G>A was not present in 190 Canadian controls. The identification of c.2780G>A in two independent families of different ethnic backgrounds strongly supported c.2780G>A as the causative mutation for ALS. Given that de novo mutation rates have been estimated to be 1.20×10^{-8} per nucleotide per generation[11] and less than one nonsynonymous single-nucleotide variant (SNV)/generation,[12] the observation of the de novo mutation further supports the idea that c.3823C>T is likely to be the causative mutation for ALS in this individual. The mutation's substituted arginine residues, Arg927 and Arg1275, are highly conserved among species (Figure 1C), and the substitutions are predicted to be probably damaging by PolyPhen-2 analysis. The amino acid residue Arg927 resides in a tyrosine kinase domain, which is essential for the receptor tyrosine kinase activity, and Arg1275 is located in a C-terminal domain in the vicinity of multiple phosphorylation sites, which mediate downstream signaling pathways (Figure 1D). The clinical presentations of these ALS individuals with the *ERBB4* mutations are summarized in Table 1. The common clinical characteristics of the individuals included both upper and lower motor-neuron involvement diagnosed as definite or probable ALS according to El Escorial and Airlie House revised criteria, relatively slow disease progression, and no obvious cognitive impairment. The individuals with the c.2780G>A mutation were characterized by relatively late onset (the ages at onset ranged from 60–70 years) and a slightly reduced penetrance. In contrast, the individual with the c.3823C>T mutation was characterized by early onset (45 years of age).

ErbB4 is a member of the epidermal growth factor (EGF) subfamily of receptor tyrosine kinases (RTKs). It forms a homodimer or a heterodimer with ErbB2 or ErbB3 and is activated upon binding of neuregulins (NRGs) to the extracellular ligand-binding domain of ErbB4.[13] Activation of ErbB4 is mediated by increased tyrosine kinase activity upon NRG binding, resulting in autophosphorylation of the C-terminal tail.[14] To determine how the two mutations identified in the ALS individuals affect ErbB4 functions, we investigated the autophosphorylation of ErbB4 in cells expressing either wild-type or mutant (c.2780G>A or c.3823C>T) *ERBB4* in the presence of NRG-1. The *ERBB4* mutations were introduced into the pBABE-puro*ERBB4JM-aCYT-2HA* plasmid encoding HA-tagged ErbB4 JM-a CYT-2[15] by site-directed mutagenesis according to the protocol described in the Phusion Site-Directed Mutagenesis Kit (Thermo Fisher Scientific). After mutagenesis, all the constructs were verified by sequencing. The plasmids were transiently transfected into COS-7 cells via FuGENE 6 transfection reagent (Roche) in accordance with the manufacturer's instructions. Transfected cells were starved of serum overnight and stimulated with 0 or 50 ng/ml NRG-1 (R&D Systems) for 10 min at 37°C. After stimulation, the cells were lysed, and samples equivalent to 50 μg of total protein were separated through 8% SDS-PAGE gels. For detectiin of ErbB4 phosphorylation and total ErbB4 protein levels, immunoblotting was performed antibodies against phospho-ErbB4 (Tyr1284) (Cell Signaling) and HA-tag (Abcam), respectively. The two amino acid substitutions, p. Arg927Gln and p. Arg1275Trp, showed a clearly reduced autophosphorylation of ErbB4 (Figure 2). On the basis of these genetic and functional data, we concluded that the two mutations are causative mutations for ALS (ALS19).

This study revealed that a reduced autophosphorylation of ErbB4 upon NRG-1 stimulation is involved in the pathogenesis of ALS. *Erbb4* is specifically expressed in the soma of large motor neurons of the rat spinal cord.[16] The lack of *Erbb4* is embryonically lethal in mice, which displayed the derangement of motor-neuron axon guidance and pathfinding during embryogenesis.[17] Heterozygous-null mice showed a reduced body weight and delayed motor development, and brain-specific conditional knock-out mice

demonstrated reduced spontaneous motor activity and grip strength of the hindlimbs.[18] Mice lacking cysteine-rich domain (CRD) isoforms of *Nrg-1* (*CRD-NRG-1*$^{-/-}$) die perinatally as a result of respiratory failure, lack detectable limb movement, and exhibit a loss of ~60% of spinal motor neurons.[19] Similarly, motor and sensory neuron-specific conditional *Nrg-1* knockout mice die at birth and showed marked retraction of motor-neuron axons.[20] Furthermore, a decrease in the amount of CRD-NRG-1 has been detected in the spinal motor neurons in FALS and SALS individuals and *Sod1* mutant mice at disease onset,[21] raising the possibility that disruption of the NRG-ErbB pathway is commonly involved in the motor-neuron degeneration underlying ALS. This study provides insight into ALS pathogenesis and is expected to pave the way for the development of innovative therapeutic strategies such as using NRGs or their agonists to upregulate ErbB4 functions.

Supplemental Data

Supplemental Data include two figures and four tables and can be found with this article online at http://www.cell.com/AJHG/.

Consortia

Consortium members of JaCALS include Ryoichi Nakamura, Hazuki Watanabe, Yuishin Izumi, Ryuji Kaji, Mitsuya Morita, Kotaro Ogaki, Akira Taniguchi, Ikuko Aiba, Koichi Mizoguchi, Koichi Okamoto, Kazuko Hasegawa, Masashi Aoki, Akihiro Kawata, Imaharu Nakano, Koji Abe, Masaya Oda, Masaaki Konagaya, Takashi Imai, Masanori Nakagawa, Takuji Fujita, Hidenao Sasaki, and Masatoyo Nishizawa.

Acknowledgments

We thank all the family members for participating in this study. This study was supported in part by KAKENHI (Grants-in-Aid for Scientific Research on Innovative Areas [22129001 and 22129002]) to S.T.; the Global COE Program from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, and a grant-in-aid (H23-Jitsuyoka [Nanbyo]-Ippan-004) from the Ministry of Health, Labour, and Welfare, Japan to S.T. We acknowledge support to R.H.B. from ALS Therapy Alliance, Project ALS, P2ALS, the Angel Fund, the Pierre L. de Bourgknecht ALS Research Foundation, the Al-Athel ALS Research Foundation, the ALS Family Charitable Foundation, and grant 1R01NS050557 from the National Institute of Neurological Disorders and Stroke of the National Institutes of Health and support to G.A.N. from the MND Research Institute of Australia. P.G.-P. was supported by the Alfonso Martin Escudero Foundation (Madrid).

Received: May 12, 2013
Revised: August 26, 2013
Accepted: September 13, 2013
Published: October 10, 2013

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes Project Database, http://www.1000genomes.org/
dbSNP135, http://www.ncbi.nlm.nih.gov/projects/SNP/
ExonPrimer, http://ihg.gsf.de/ihg/ExonPrimer.html
NCBI37/hg19 assembly, http://genome.ucsc.edu/
NHLBI GO Exome Sequencing Project (NHLBI-ESP), https://esp.gs.washington.edu/drupal
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/
Personal genome databases, http://www.sequenceontology.org/resources/10Gen.html
PLINK algorithm, http://pngu.mgh.harvard.edu/purcell/plink/
PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/
Refseq, http://www.ncbi.nlm.nih.gov/projects/RefSeq/
UCSC Human Genome Browser, http://genome.ucsc.edu/

Accession Numbers

The dbSNP accession numbers for the c. 2780G>A and c. 3823C>T mutations reported for *ERBB4* in this paper are ss831884245 and ss831884246, respectively.

References

1. Al-Chalabi, A., Jones, A., Troakes, C., King, A., Al-Sarraj, S., and van den Berg, L.H. (2012). The genetics and neuropathology of amyotrophic lateral sclerosis. Acta Neuropathol. *124*, 339–352.
2. Andersen, P.M., and Al-Chalabi, A. (2011). Clinical genetics of amyotrophic lateral sclerosis: what do we really know? Nat Rev Neurol *7*, 603–615.
3. Takahashi, Y., Seki, N., Ishiura, H., Mitsui, J., Matsukawa, T., Kishino, A., Onodera, O., Aoki, M., Shimozawa, N., Murayama, S., et al. (2008). Development of a high-throughput microarray-based resequencing system for neurological disorders and its application to molecular genetics of amyotrophic lateral sclerosis. Arch. Neurol. *65*, 1326–1332.
4. Ishiura, H., Takahashi, Y., Mitsui, J., Yoshida, S., Kihira, T., Kokubo, Y., Kuzuhara, S., Ranum, L.P., Tamaoki, T., Ichikawa, Y., et al. (2012). C9ORF72 repeat expansion in amyotrophic lateral sclerosis in the Kii peninsula of Japan. Arch. Neurol. *69*, 1154–1158.
5. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.
6. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
7. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.
8. Fukuda, Y., Nakahara, Y., Date, H., Takahashi, Y., Goto, J., Miyashita, A., Kuwano, R., Adachi, H., Nakamura, E., and Tsuji, S. (2009). SNP HiTLink: a high-throughput linkage analysis system employing dense SNP data. BMC Bioinformatics *10*, 121.
9. Gudbjartsson, D.F., Thorvaldsson, T., Kong, A., Gunnarsson, G., and Ingolfsdottir, A. (2005). Allegro version 2. Nat. Genet. *37*, 1015–1016.
10. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

11. Kong, A., Frigge, M.L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S.A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., et al. (2012). Rate of de novo mutations and the importance of father's age to disease risk. Nature *488*, 471–475.
12. Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature *485*, 237–241.
13. Plowman, G.D., Green, J.M., Culouscou, J.M., Carlton, G.W., Rothwell, V.M., and Buckley, S. (1993). Heregulin induces tyrosine phosphorylation of HER4/p180erbB4. Nature *366*, 473–475.
14. Carpenter, G. (2003). ErbB-4: mechanism of action and biology. Exp. Cell Res. *284*, 66–77.
15. Sundvall, M., Korhonen, A., Vaparanta, K., Anckar, J., Halkilahti, K., Salah, Z., Aqeilan, R.I., Palvimo, J.J., Sistonen, L., and Elenius, K. (2012). Protein inhibitor of activated STAT3 (PIAS3) protein promotes SUMOylation and nuclear sequestration of the intracellular domain of ErbB4 protein. J. Biol. Chem. *287*, 23216–23226.
16. Pearson, R.J., Jr., and Carroll, S.L. (2004). ErbB transmembrane tyrosine kinase receptors are expressed by sensory and motor neurons projecting into sciatic nerve. J. Histochem. Cytochem. *52*, 1299–1311.
17. Gassmann, M., Casagranda, F., Orioli, D., Simon, H., Lai, C., Klein, R., and Lemke, G. (1995). Aberrant neural and cardiac development in mice lacking the ErbB4 neuregulin receptor. Nature *378*, 390–394.
18. Golub, M.S., Germann, S.L., and Lloyd, K.C.K. (2004). Behavioral characteristics of a nervous system-specific erbB4 knockout mouse. Behav. Brain Res. *153*, 159–170.
19. Wolpowitz, D., Mason, T.B., Dietrich, P., Mendelsohn, M., Talmage, D.A., and Role, L.W. (2000). Cysteine-rich domain isoforms of the neuregulin-1 gene are required for maintenance of peripheral synapses. Neuron *25*, 79–91.
20. Yang, X., Arber, S., William, C., Li, L., Tanabe, Y., Jessell, T.M., Birchmeier, C., and Burden, S.J. (2001). Patterning of muscle acetylcholine receptor gene expression in the absence of motor innervation. Neuron *30*, 399–410.
21. Song, F., Chiang, P., Wang, J., Ravits, J., and Loeb, J.A. (2012). Aberrant neuregulin 1 signaling in amyotrophic lateral sclerosis. J. Neuropathol. Exp. Neurol. *71*, 104–115.

ORIGINAL ARTICLE

JBMR®

A Recurrent De Novo *FAM111A* Mutation Causes Kenny–Caffey Syndrome Type 2

Tsuyoshi Isojima,[1] Koichiro Doi,[2] Jun Mitsui,[3] Yoichiro Oda,[4] Etsuro Tokuhiro,[5] Akihiro Yasoda,[6] Tohru Yorifuji,[7] Reiko Horikawa,[8] Jun Yoshimura,[2] Hiroyuki Ishiura,[3] Shinichi Morishita,[2] Shoji Tsuji,[3] and Sachiko Kitanaka[1]

[1]Department of Pediatrics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan
[2]Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Japan
[3]Department of Neurology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan
[4]Department of Pediatrics, Ohta Nishinouchi Hospital, Koriyama, Japan
[5]Department of Pediatrics, Odawara City Hospital, Odawara, Japan
[6]Department of Medicine and Clinical Science, Kyoto University Graduate School of Medicine, Kyoto, Japan
[7]Department of Pediatric Endocrinology and Metabolism, Children's Medical Center, Osaka City General Hospital, Osaka, Japan
[8]Division of Endocrinology and Metabolism, National Center for Child Health and Development, Tokyo, Japan

ABSTRACT

Kenny–Caffey syndrome (KCS) is a rare dysmorphologic syndrome characterized by proportionate short stature, cortical thickening and medullary stenosis of tubular bones, delayed closure of anterior fontanelle, eye abnormalities, and hypoparathyroidism. The autosomal dominant form of KCS (KCS type 2 [KCS2]) is distinguished from the autosomal recessive form of KCS (KCS type 1 [KCS1]), which is caused by mutations of the tubulin-folding cofactor E (*TBCE*) gene, by the absence of mental retardation. In this study, we recruited four unrelated Japanese patients with typical sporadic KCS2, and performed exome sequencing in three patients and their parents to elucidate the molecular basis of KCS2. The possible candidate genes were explored by a de novo mutation detection method. A single gene, *FAM111A* (NM_001142519.1), was shared among three families. An identical missense mutation, R569H, was heterozygously detected in all three patients but not in the unaffected family members. This mutation was also found in an additional unrelated patient. These findings are in accordance with those of a recent independent report by a Swiss group that KCS2 is caused by a de novo mutation of *FAM111A*, and R569H is a hot spot mutation for KCS2. Although the function of FAM111A is not known, this study would provide evidence that FAM111A is a key molecule for normal bone development, height gain, and parathyroid hormone development and/or regulation. © 2014 American Society for Bone and Mineral Research.

KEY WORDS: KENNY–CAFFEY SYNDROME; *FAM111A*; PARATHYROID-RELATED DISORDERS; HYPOMAGNESEMIA

Introduction

Kenny–Caffey syndrome (KCS) (OMIM #244460, %127000) is a rare dysmorphologic syndrome characterized by severe proportionate short stature with adult heights of 121 to 149 cm, cortical thickening and medullary stenosis of tubular bones, delayed closure of the anterior fontanelle, eye abnormalities, and hypocalcemia owing to hypoparathyroidism.[1–4] KCS is classified into two types according to its clinical features and inheritance pattern. Classical cases have normal intelligence and are transmitted as an autosomal dominant trait or sporadically and are called KCS type 2 (KCS2) (OMIM %127000).[5] Cases having mental and prenatal growth retardation and transmitted as an autosomal recessive trait are called KCS type 1 (KCS1) (OMIM #244460).[4,6,7]

In 2002, a study of 65 individuals from 34 pedigrees of Middle Eastern origin resulted in the identification of mutations of the tubulin-folding cofactor E (*TBCE*) gene as the cause of KCS1. *TBCE* encodes a molecular chaperone required for heterodimerization of α-tubulin with β-tubulin.[8] KCS2 is extremely rare, with only 5 sporadic cases reported in Japan.[9–12] Because of this rarity, the cause of KCS2 has been unknown until it was recently reported to involve the "family with sequence similarity 111, member A" (*FAM111A*) gene (NM_001142519.1) by a Swiss group in 2013.[13]

In this study, we recruited 4 Japanese patients with typical sporadic KCS2 having normal intelligence and performed whole exome sequencing in 3 unrelated trios to elucidate the molecular basis of KCS2. We hypothesized that KCS2 is caused by de novo mutations and built a de novo mutation detection pipeline to

Received in original form June 10, 2013; revised form August 21, 2013; accepted August 27, 2013. Accepted manuscript online August 31, 2013.
Address correspondence to: Sachiko Kitanaka, MD, PhD, Department of Pediatrics, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8655, Japan. E-mail: sachi-tky@umin.ac.jp
Additional Supporting Information may be found in the online version of this article.

Journal of Bone and Mineral Research, Vol. 29, No. 4, April 2014, pp 992–998
DOI: 10.1002/jbmr.2091

process the raw data from exome sequencing. Using this method, we found an identical de novo mutation in *FAM111A* in all 4 patients. This and the reported independent studies provide evidence that *FAM111A* is the cause of KCS2, and R569H is a hot spot mutation for KCS2.

Materials and Methods

Subjects

Case 1

This 10-year-old girl (Fig. 1, I-1)[9] was born at 40 weeks of gestation to nonconsanguineous, healthy Japanese parents. Polysyndactyly was noticed at birth. At 3 months of age, she was referred to a pediatric endocrinologist because of growth retardation. Her body length, body weight, and head circumferences were 55 cm (−2.5 SD), 5092 g (−1.8 SD), and 37.3 cm (0.2 SD), respectively. She was found to have liver dysfunction with a serum aspartate aminotransferase (AST) level of 227 U/L (reference range 21 to 75) and serum alanine aminotransferase (ALT) level of 227 U/l (reference range 11 to 69). Basal serum insulin-like growth factor (IGF-I), calcium (Ca), and phosphorus (P) levels were within normal limits. At the age of 1 year, hypocalcemia was revealed. Her serum Ca, P, and intact parathyroid hormone (PTH) levels were 1.6 mmol/L (reference range 2.1 to 2.4), 2.6 mmol/L (reference range 0.88 to 1.4), and 11 ng/L (reference range 15 to 50), respectively, with a normal magnesium (Mg) level of 0.86 mmol/L (reference range 0.74 to 0.90). Her serum $1,25(OH)_2D$ level, serum alkaline phosphatase level, and urine Ca/creatinine ratio were within normal ranges. Brain computed tomography (CT) revealed calcification in the basal ganglia (Fig. 2*A*). She was diagnosed with primary hypoparathyroidism and was treated with alfacalcidol [$1\alpha(OH)D_3$]. At 2 years of age, she was diagnosed with KCS2 based on clinical manifestations of proportionate short stature, cortical thickening and medullary stenosis confirmed by radioscopic study (Fig. 2*B*), macrocephaly with delayed closure of the anterior fontanelle, eye abnormalities (hypermetropia and pseudopapilledema), and normal intelligence. Magnesium oxide was administered because of a low serum Mg level (below 0.62 mmol/L) at 3 years of age.

Case 2

This 16-year-old boy (Fig. 1, II-4)[10] was born at 41 weeks of gestation to nonconsanguineous, healthy Japanese parents. When he was 23 days old, he had a generalized convulsion because of hypocalcemia. At this time, his serum Ca, P, Mg, and intact PTH levels were 1.5 mmol/L, 3.1 mmol/L, 0.74 mmol/L, and undetectable, respectively. T-cell subset was normal. He was treated with alfacalcidol on the basis of a diagnosis of primary hypoparathyroidism. Magnesium sulfate was added because of his low serum Mg level at the age of 1 year. He suffered repeated bouts of acute otitis media until the age of two years. His serum IgG level was within the normal range. At 3 years and 1 month, his height, weight, and head circumference were 77.9 cm (−4.4 SD), 9.9 kg (−2.7 SD), and 47.4 cm (−1.5 SD), respectively. He had normal intelligence for his age. He was diagnosed with KCS2 based on clinical findings of proportionate short stature, medullary stenosis revealed by radiography, a widely open anterior fontanelle (Fig. 2*C*, skull radiograph at 9 years), and hypermetropia. He also suffered severe atopic dermatitis after normalization of his serum Ca levels. His growth chart is shown in Fig. 2*D*.

Case 3

This 22-year-old woman (Fig. 1, III-9)[11] was born at 40 weeks of gestation to nonconsanguineous, healthy Japanese parents following an uneventful pregnancy. At 1 month, she had an episode of generalized convulsions because of hypocalcemia. At this episode, her serum Ca, P, Mg, and intact PTH levels were 1.3 mmol/L, 2.9 mmol/L, 0.49 mmol/L, and undetectable, respectively. Oral alfacalcidol administration was started on the basis of a diagnosis of primary hypoparathyroidism. At the age of 5 years 1 month, she was referred to another hospital. Her height was 84.2 cm (−5.3 SD), and her weight was 12.2 kg (−2.2 SD). She had normal intelligence. Brain CT revealed fine calcification in the basal ganglia. Based on clinical manifestations of proportionate short stature, medullary stenosis of the long bones typical of KCS, a 1 × 1-cm opening of her anterior fontanelle, normal intelligence, and hypermetropia, she was diagnosed with KCS2. The patient was started with a combination therapy of vitamin D and magnesium sulphate. Fig. 2*E* shows her radiograph at 14 years of age.

Case 4

This 38-year-old man (Fig. 1, IV-13)[12] was born at 40 weeks of gestation to nonconsanguineous, healthy Japanese parents following an uneventful pregnancy. At 8 days of age, he had a generalized convulsion, and hypocalcemia (0.75 mmol/L) and hypomagnesemia (0.18 mmol/L) were detected. The convulsion was controlled by intravenous administration of Ca gluconate and magnesium sulfate until he was 15 days old. At 4 years of age, he again had an episode of generalized convulsion because of hypocalcemia. At this episode, his serum Ca, P, and intact PTH levels were 1.2 mmol/L, 2.6 mmol/L, and undetectable, respectively. He was diagnosed with primary hypoparathyroidism, and oral alfacalcidol and Ca lactate administration were started. He suffered repeated acute otitis media during infancy and was affected with empyema and bacterial meningitis at 4 years of age. Hypogammaglobulinemia was found, and he was administered gamma globulin intermittently. At 12 years of age, he was referred to another hospital for further investigation. His height was 99 cm (−6.3 SD), and his weight was 16.2 kg (−3.3 SD). He had normal intelligence with an intelligence quotient score of 105. Brain CT revealed fine calcification in the basal ganglia. Based on clinical manifestations of proportionate short stature, medullary stenosis of the long bones, a 4.2 × 1.8-cm opening of his anterior fontanelle, and eye abnormalities (hypermetropia, amblyopia, and pseudopapilledema), he was diagnosed with KCS2. Mg loading and Ca restriction tests revealed that his hypoparathyroidism was secondary to hypomagnesemia. The patient was then changed from vitamin D and Ca lactate to magnesium sulfate treatment, which successfully corrected his serum Ca levels.

We recruited these 4 Japanese patients with clinically diagnosed typical sporadic KCS2 (Fig. 1). Supplemental Table S1{TBL S1} summarizes the clinical characteristics of the 4 patients. We obtained peripheral blood samples from all 4 patients, together with those of 9 unaffected parents or siblings, with informed consent for DNA analysis (Fig. 1). The study was performed with the approval of the Ethics Committee of The University of Tokyo and of each institution where the samples

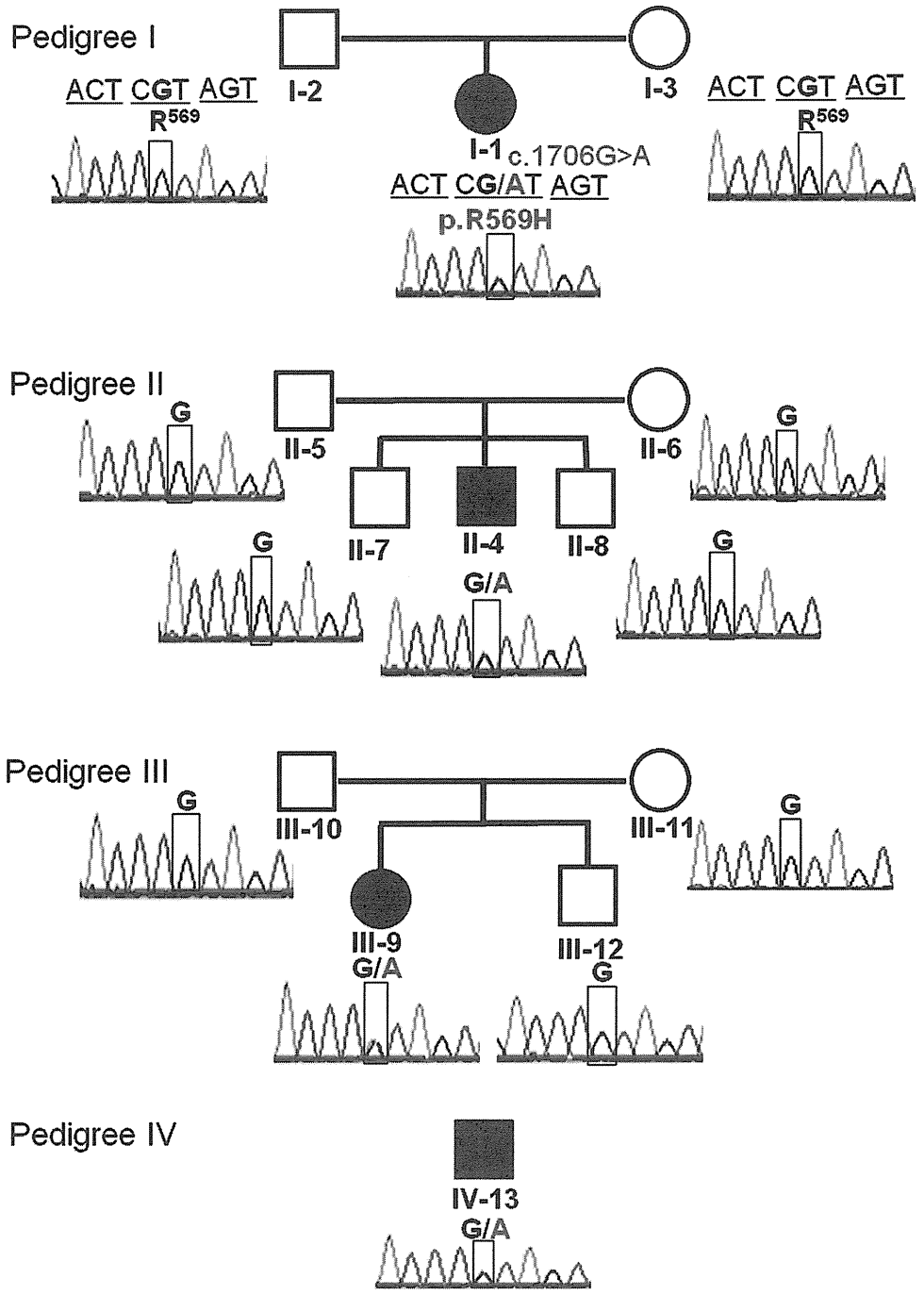

Fig. 1. Four pedigrees analyzed in this study, showing the chromatograms of Sanger sequencing reactions of the *FAM111A* mutation in patients and family members. Data were obtained by Sanger sequencing during the confirmation process. All mutations were checked by bidirectional sequencing. In each pedigree, a black symbol represents the proband, a square indicates a male, and a circle shows a female. In the chromatogram, black letters indicate the wild-type nucleotide sequence. Nucleotides in red indicate mutations. R569H was identified in all probands but not in any of the unaffected family members.

were collected, and conducted in accordance with the Declaration of Helsinki. Genomic DNA was extracted from peripheral white blood cells of the patients and family members using a QIAamp DNA Blood Midi Kit (Qiagen, Hilden, Germany). Healthy Japanese volunteers were recruited, and DNA was extracted with informed consent.

Exome sequencing

Exome sequences were enriched using a TruSeq Exome Enrichment Kit (Illumina, San Diego, CA, USA) from 1 μg of genomic DNA, according to the manufacturer's instructions. The captured DNA samples were subjected to massively parallel

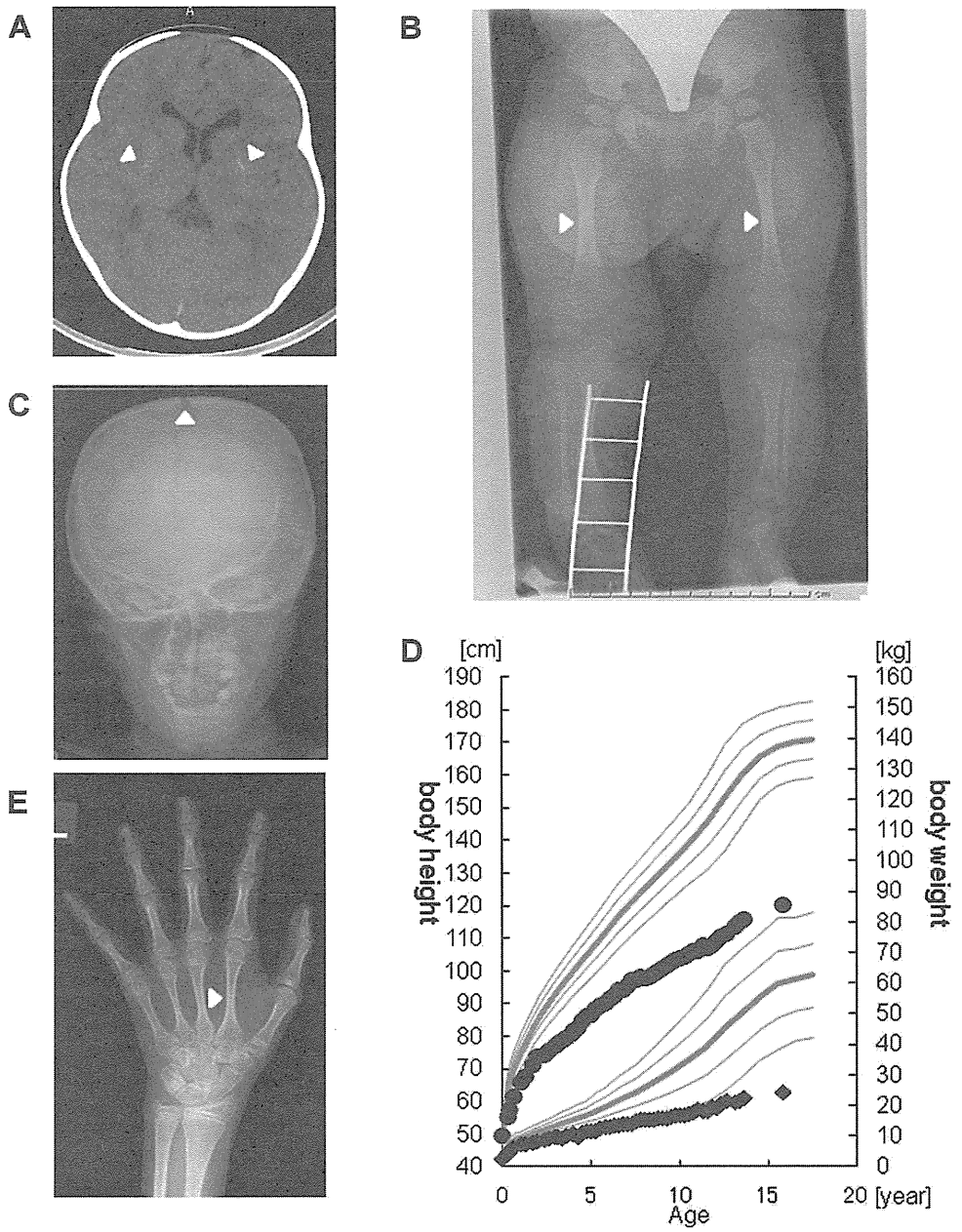

Fig. 2. Radiographic studies and growth charts of probands. (*A*) Brain computed tomography of patient I-1. The arrowheads indicate calcification in the basal ganglia. (*B*) Radiograph of patient I-1 at diagnosis. Cortical thickening and medullary stenosis are evident. The object shown in the right leg is used for fixing a peripheral catheter. (*C*) Radiograph of patient II-4 at age 9 years. It is of note that the anterior fontanelle is open. (*D*) Growth chart of patient II-4 superimposed on the standard growth chart for a Japanese boy. Black circles indicate the patient's height, and black squares indicate his weight. (*E*) Radiograph of patient III-9. Cortical thickening and medullary stenosis can be observed.

sequencing (100-bp paired-end reads) on an Illumina HiSeq2000 sequencing system (Illumina). An average of 95 million reads of the sequence data was obtained for each individual. On an average, 98.50% of the total bases were mapped to the reference genome with a mean coverage of 140.5 ×, which encompassed 91.94% of the targeted regions with coverage >10× (Supplemental Table S2).{TBL S2} The Burrows-Wheeler Aligner (BWA) package[14] and SAMtools[15] were used as default settings for alignment of raw reads and detection of single-nucleotide variants (SNVs) and indels. Subsequently, SNVs and indels were filtered with three trio samples (ie, pedigrees I, II, and III) (Supplemental Fig. S1). We extracted both homo/heterozygous nonsynonymous coding variants, which were called in the proband, and filtered these candidates using the following three steps:

Step 1: Using candidate de novo mutations that are homozygous references in both parents and are supported by 10 or more high-quality reads at the mutated sites for every trio member.

Step 2: Using reliable homozygous references in each parent such that the likelihood of heterozygosis, ${}_nC_i\,(1/2)^i\,(1/2)^{n-i}$, is less than that of homozygosis, ${}_nC_i\,(999/1000)^i(1/1000)^{n-i}$, where the average error rate is assumed to be 1/1000, n represents the number of total reads, and i is the number of reads consistent with the reference. One may have an impression that this condition does not often hold true; however, we often observed cases that violated this condition, especially when the reference base was mutated into one of the other three bases with an almost equal probability.

Step 3: Using reliable de novo mutations of the proband such that the number of alternative allele reads was at least 30% among the total reads, which is the condition proposed in a recent report.[16]

Sanger sequencing

Sanger sequencing was performed to detect *TBCE* (KCS1) and validate the presence of each variant detected by exome sequencing in patients with KCS2 and the absence of each in the genomes of the parents and siblings. The entire coding region and exon–intron boundaries of *TBCE* and *FAM111A* were amplified from genomic DNA by polymerase chain reaction (PCR) using the designed PCR primers (Supplemental Table S3). {TBL S3} Subsequently, PCR products were sequenced using an ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems, Foster City, CA, USA) and the forward and reverse primers used for PCR amplification. Direct sequencing in both directions was performed on an autosequencer (PE Applied Biosystems 3130 × 1, Genetic Analyzer).

FAM111A mRNA expression analysis

Total RNA was prepared using ISOGEN reagent (Nippon Gene, Osaka, Japan), according to the manufacturer's instructions, from peripheral white blood cells of the patients and family members. Total RNA (4 μg) was used to synthesize cDNA with the SuperScript Preamplification System for first-strand cDNA synthesis (Life Technologies, Rockville, MD, USA). mRNA levels were measured using an ECO real-time PCR system (Illumina) and KAPA SYBR Fast qPCR Kit (Kapa Biosystems, Woburn, MA, USA) using the following primer pairs: FAM111Ae5-2F and FAM111Ae5-2R; FAM111Ae5-3F and 5′-CCTCATCACTCATCATTTCTACATCC-3′; GAPDH, 5′-GAAGGTGAAGGTCGGAGTC-3′ (F) and 5′-GAAGATGGTGATGGGATTTC-3′ (R). The relative mRNA level was calculated using an arithmetic formula based on the difference between the threshold cycle of a given target cDNA and that of an endogenous reference cDNA. Direct sequencing of the RT-PCR products was performed by Sanger sequencing as for DNA samples.

Results

We first confirmed by Sanger sequencing that none of the 4 patients had *TBCE* mutations. This finding, together with the fact that all the patients were of normal intelligence, distinguishes these patients from patients with KCS1.

We hypothesized that these sporadic cases may be caused by de novo mutations in novel nonsynonymous coding variants. Whole exome sequencing was performed for 3 patients (I-1, II-4, and III-9; Fig. 1) and their parents (I-2, I-3, II-5, II-6, III-10, and III-11; Fig. 1). Statistical data of exome sequencing experiments are shown in Supplemental Table S2. The candidate variants were selected according to the processes described in Materials and Methods based on the de novo mutation detection pipeline designed in the present study (Supplemental Fig. S1). Supplemental Table S4{TBL S4} summarizes the results of filtering to detect candidate genes for KCS2. To select variants as candidate mutations for KCS2, variations that caused amino acid substitution were extracted, which resulted in 11,024 (pedigree I), 10,828 (pedigree II), and 11,020 (pedigree III) SNVs and indels. After three filtering steps, 5 (pedigree I), 5 (pedigree II), and 6 (pedigree III) SNVs were identified. Among the candidate genes filtered using the three aforementioned filtering steps, only one single gene, *FAM111A* (NM_001142519.1), was shared among all 3 families. Sanger sequence analysis of all exons of *FAM111A* confirmed an identical c.1706G > A heterozygous mutation in exon 5 in all 3 patients (Fig. 1). This mutation is predicted to result in substitution of arginine to histidine in codon 569 (R569H). None of the unaffected family members had this mutation, indicating that R569H was a de novo mutation. This mutation was also found in an additional unrelated patient (IV-13).

R569H is not present in 373 Japanese healthy control subjects of an in-house exome database, and not in another 100 alleles from 50 unrelated healthy Japanese individuals by Sanger sequencing. It was also not found in the Japanese SNP control database established by the National Bioscience Data Base Center that has 1 million genome-wide SNPs of 700 samples (http://gwas.biosciencedbc.jp/snpdb/snp_top.php), nor among 6500 samples listed on the exome variant server (http://evs.gs.washington.edu/EVS/), implying that the minor allele frequency is less than 0.01% in these data. However, one SNP was found in the 1000 Genomes database at R569 (rs184251651), which results in substitution to "cysteine" (minor allele frequency 0.1%).

We assessed the functionality of the R569H mutation using the Sorting Intolerant From Tolerant (SIFT) (http://sift.jcvi.org) and Polymorphism Phenotyping 2 (PolyPhen2) (http://genetics.bwh.harvard.edu/pph2) tools, by homology modeling and threading. These in silico studies predicted R569H as "tolerated" and "benign," respectively.

We analyzed the expression levels of *FAM111A* mRNA in peripheral white blood cells by real-time PCR. *FAM111A* expression levels in the patients were comparable with those in unaffected family members and normal controls (data not shown). We also found that mutant and wild-type *FAM111A* were equivalently expressed in the patients, which were identified by sequencing the reverse-transcribed PCR products.

Discussion

In the present study, we identified *FAM111A* as the gene responsible for KCS2 by applying an exome sequencing strategy, and we identified a heterozygous identical de novo *FAM111A* mutation, R569H, in 4 Japanese patients with KCS2. While preparing this article, another independent research group from Switzerland reported similar findings following whole exome sequencing of the patients.[13] They reported that all 5 clinically diagnosed KCS2 patients had de novo *FAM111A* mutations. Most interestingly, 4 of the 5 patients from different countries had the same R569H mutation as detected in our patients. Our 4 pedigrees are unrelated to each other and live in different areas in Japan. Moreover, the parents of the 3 patients did not have the mutation, suggesting that this recurrent mutation was caused by sporadic mutation. Taken together, these two independent

studies confirm that *FAM111A* is the causative gene for KCS2, and R569H is the hot spot mutation of KCS2.

FAM111A encodes a previously uncharacterized protein consisting of 611 amino acids. The carboxy-terminal half of the protein has homology to trypsin-like peptidases, and the catalytic triad specific to such peptidases is conserved.[17] Transcriptional expression of *FAM111A* is ubiquitous according to the human protein atlas (http://www.proteinatlas.org/ENSG00000166801/normal). It is expressed in the parathyroid gland and bone, but the expression levels are similar to those in other tissues. FAM111A has 35% amino acid homology to FAM111B, a paralog located on 11q12.1 at a distance of only 16 kb from FAM111A. The functions of FAM111A and FAM111B are largely unknown. A recent report showed that FAM111A functions as a host range restriction factor and is required for viral replication and gene expression by specifically interacting with Simian Virus 40 large T antigen (LT).[17] In addition, *FAM111A* mRNA and protein levels have been shown to be regulated in a cell cycle-dependent manner with the lowest expression during the G0 or quiescent phase and peak expression during the G2/M phase.[17] Another recent report revealed that variants in the region including *FAM111A* and *FAM111B* were associated with prostate cancer.[18] However, the clinical course of disease in our 4 patients revealed neither increased viral infections nor carcinogenesis up to early adulthood.

In silico analyses suggested that the de novo mutation (R569H) would not significantly affect the function of FAM111A. We also found that the mutant *FAM111A* mRNA was expressed similarly to the wild type in peripheral blood cells. This raises the question of how this mutation causes KCS2. One hypothesis is that this mutation does not cause loss of function of the protein but rather modulates its peptidase activity for a particular target peptide in a mutant-specific way. Another possibility is that FAM111A functions with some physiological partner(s) and the disease occurs as a result of specific modulation of this putative network. This may fit the observation that FAM111A is regulated in a cell-dependent manner and interacts with the LT C-terminal region.[17] We speculate that one of the candidate partner proteins is TBCE because KCS1 and KCS2 share distinctive phenotypic features: skeletal dysmorphic features and primary hypoparathyroidism.

Some diseases are caused by specific mutations of a single gene. Some mutations may cause a gain-of-function effect, as in achondroplasia or McCune–Albright syndrome,[19,20] whereas others have an unknown function, as in Caffey syndrome caused by mutations in *COL1A1*[21] or in several diseases related to *FGFR3*. In this study, we found that a specific mutation (R569H) of *FAM111A* would lead to KCS2. Intriguingly, one SNP was found in the 1000 Genomes database at R569 (rs184251651), which results in substitution to "cysteine." This SNP has been reported to have minor allele frequency of 0.1% (only one allele) and is not validated. Moreover, the absence of the SNP in 6500 samples in the exome variant server suggests a possibility of sequencing error in the database. Nevertheless, it might be speculated that a specific change to "histidine" may lead to an unidentified function of this protein resulting in KCS2, which is not caused by other amino acids. This hypothesis will be supported by the fact that this amino acid is not well conserved among various species (Fig. 3).

It is reported that 95% and 97% of KCS1 cases had prenatal and postnatal growth retardation, and mental retardation, respectively.[22] In contrast, most of the reported KCS2 patients, including our patients with *FAM111A* mutations, had normal

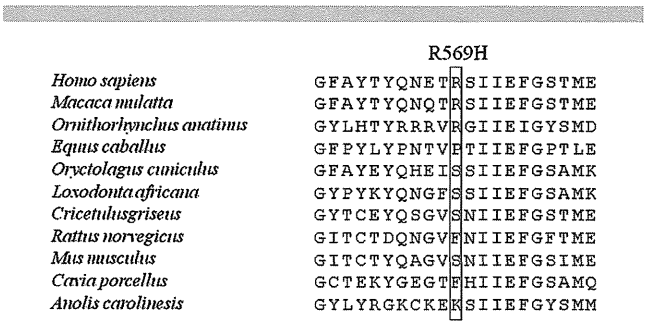

Fig. 3. Homologous comparison of the altered protein. Letters in the rectangular box indicate the human *FAM111A* R569 residue. It is of note that R569 is not well conserved among various species.

birth weight and length and normal intelligence (Supplemental Table S1).[13] These phenotypic differences between KCS1 and KCS2 suggest that the *FAM111A* mutation does not affect bone development and height gain in the fetus but becomes important postnatally. It also suggests that the *FAM111A* mutation does not affect mental development. Now that *FAM111A* has been identified as a causative gene for KCS2, further studies on the physiological function of FAM111A and TBCE should be performed to uncover the phenotypic differences between these two types.

There are several human diseases, as well as mouse models of hypoparathyroidism, caused by aberrations in the cascade of genes indispensable for the development and regulation of the parathyroid gland.[23,24] To date, FAM111A is not known to relate to any of these genes. There have been only a few reports describing the pathophysiology of hypoparathyroidism in KCS. Absence of the parathyroid glands has been reported in some patients with KCS2 and KCS1.[25,26] In contrast, some patients do not have hypoparathyroidism from early infancy, suggesting the presence of some parathyroid gland as in our patient I-1.[4,27] Furthermore, hypoparathyroidism may be secondary to hypomagnesemia as in our patient IV-13. Considering the fact that all of our 4 patients as well as another reported KCS2 case had hypomagnesemia,[4] FAM111A might be involved in magnesium homeostasis. Although further investigation is necessary to reveal the cause of hypoparathyroidism in KCS2, this study shows that a new gene, *FAM111A*, is indispensable for PTH development and/or regulation.

In conclusion, our finding that all 4 Japanese KCS2 patients we tested have the same de novo mutation (R569H) of *FAM111A* indicates that KCS2 is caused by a heterozygous mutation of *FAM111A*, and R569H is the hot spot mutation in patients with KCS2. Although the function of FAM111A is largely unknown, this study provides evidence that FAM111A is a key molecule for normal bone development, height gain, and PTH development and/or regulation. Our finding further creates a new research area in the fields associated with shared phenotypic features in KCS and different phenotypes between KCS1 and KCS2.

Disclosures

TI has received research grants and speaker's fees from Novo Nordisk and has received speaker's fees from Eli Lilly. SK has received research grants and speaker's fees from Novo Nordisk,

Pfizer, Eli Lilly, and JCR Pharmaceuticals and has received speaker's fees from Chugai Pharmaceutical. All other authors state that they have no conflicts of interest.

Acknowledgments

We thank the patients and their family members who participated in this study. We also thank Minako Takaki and Reiko Onai for technical support, and Mitsuko Itoh for kind assistance in English usage. This study was supported by a Grant-in-Aid from the Ministry of Education, Science, Sports, and Culture of Japan.

Authors' roles: Study design: TI and SK. Patients' sample and clinical data collection: YO, ET, AY, TY, and RH. Study conduct and data collection: TI, JM, and HI. Exome sequencing data analysis: JM, KD, JY, HI, SM, and ST. Data interpretation: TI, JM, KD, SM, ST, and SK. Study support and intellectual input: SM, ST, and SK. Drafting manuscript: TI. Revising manuscript content: SM, ST, and SK. Approving final version of manuscript: all authors. TI, SM, and SK take responsibility for the integrity of the data analysis.

References

1. Kenny FM, Linarelli L. Dwarfism and cortical thickening of tubular bones. Transient hypocalcemia in a mother and son. Am J Dis Child. 1966;111(2):201–7.
2. Caffey J. Congenital stenosis of medullary spaces in tubular bones and calvaria in two proportionate dwarfs—mother and son; coupled with transitory hypocalcemic tetany. Am J Roentgenol Radium Ther Nucl Med. 1967;100(1):1–11.
3. Fanconi S, Fischer JA, Wieland P, et al. Kenny syndrome: evidence for idiopathic hypoparathyroidism in two patients and for abnormal parathyroid hormone in one. J Pediatr. 1986;109(3):469–75.
4. Lee WK, Vargas A, Barnes J, Root AW. The Kenny-Caffey syndrome: growth retardation and hypocalcemia in a young boy. Am J Med Genet. 1983;14(4):773–82.
5. Franceschini P, Testa A, Bogetti G, et al. Kenny-Caffey syndrome in two sibs born to consanguineous parents: evidence for an autosomal recessive variant. Am J Med Genet. 1992;42(1):112–6.
6. Sabry MA, Zaki M, Abul Hassan SJ, et al. Kenny-Caffey syndrome is part of the CATCH 22 haploinsufficiency cluster. J Med Genet. 1998;35(1):31–6.
7. Sabry MA, Farag TI, Shaltout AA, et al. Kenny-Caffey syndrome: an Arab variant? Clin Genet. 1999;55(1):44–9.
8. Parvari R, Hershkovitz E, Grossman N, et al. Mutation of TBCE causes hypoparathyroidism-retardation-dysmorphism and autosomal recessive Kenny-Caffey syndrome. Nat Genet. 2002;32(3):448–52.
9. Tadaki H, Tokuhiro E, Shiga K, Kikuchi N, Mukai N, Fujieda K. A case of a 2-year-old girl with Kenny-Caffey Syndrome Type 2. Clin Pediatr Endocrinol. 2005;14(2 Suppl):22.
10. Oda Y, Ono R, Hiwatari M, Iwasaki H, Namai Y, Iimori Y. A case report: three-year-old boy of Kenny-Caffey Syndrome Type 2. Clin Pediatr Endocrinol. 2000;9(2):140.
11. Yorifuji T, Muroi J, Uematsu A. Kenny-Caffey syndrome without the CATCH 22 deletion. J Med Genet. 1998;35(12):1054.
12. Izumi Y, Tanae A, Kuratsuji T, et al. A case of 12-year-old boy with Kenny syndrome associated with hypomagnesemia and humoral immunodeficiency [in Japanese]. Shoninaika. 1987;19(10):1503–2.
13. Unger S, Go´rna MW, Le Béchec A, et al. FAM111A mutations result in hypoparathyroidism and impaired skeletal development. Am J Hum Genet. 2013 May 14. [Epub ahead of print].
14. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.
15. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
16. Kong A, Frigge ML, Masson G, et al. Rate of de novo mutations and the importance of father's age to disease risk. Nature. 2012;488(7412):471–5.
17. Fine DA, Rozenblatt-Rosen O, Padi M, et al. Identification of FAM111A as an SV40 host range restriction and adenovirus helper factor. PLoS Pathog. 2012;8(10):e1002949.
18. Akamatsu S, Takata R, Haiman CA, et al. Common variants at 11q12, 10q26 and 3p11.2 are associated with prostate cancer susceptibility in Japanese. Nat Genet. 2012;44(4): 426–9, S421.
19. Shiang R, Thompson LM, Zhu YZ, et al. Mutations in the transmembrane domain of FGFR3 cause the most common genetic form of dwarfism, achondroplasia. Cell. 1994;78(2):335–42.
20. Weinstein LS, Shenker A, Gejman PV, Merino MJ, Friedman E, Spiegel AM. Activating mutations of the stimulatory G protein in the McCune-Albright syndrome. N Engl J Med. 1991;325(24):1688–95.
21. Gensure RC, Mäkitie O, Barclay C, Chan C, Depalma SR, Bastepe M, Abuzahra H, Couper R, Mundlos S, Sillence D, Ala Kokko L, Seidman JG, Cole WG, Jüppner H. A novel COL1A1 mutation in infantile cortical hyperostosis (Caffey disease) expands the spectrum of collagen-related disorders. J Clin Invest. 2005;115(5):1250–7.
22. Courtens W, Wuyts W, Poot M, Szuhai K, Wauters J, Reyniers E, Eleveld M, Diaz G, Nöthen MM, Parvari R. Hypoparathyroidism-retardation-dysmorphism syndrome in a girl: A new variant not caused by a TBCE mutation--clinical report and review. Am J Med Genet A. 2006;140(6):611–7.
23. Shoback D, Clinical practice. Hypoparathyroidism. N Engl J Med. 2008;359(4):391–403.
24. Zajac JD, Danks JA. The development of the parathyroid gland: from fish to human. Curr Opin Nephrol Hypertens. 2008;17(4):353–6.
25. Boynton JR, Pheasant TR, Johnson BL, Levin DB, Streeten BW. Ocular findings in Kenny's syndrome. Arch Ophthalmol. 1979;97(5):896–00.
26. Parvari R, Diaz GA, Hershkovitz E. Parathyroid development and the role of tubulin chaperone E. Horm Res. 2007;67(1):12–21.
27. Wilson MG, Maronde RF, Mikity VG, Shinno NW. Dwarfism and congenital medullary stenosis (Kenny syndrome). Birth Defects Orig Artic Ser. 1974;10(12):128–32.

ORIGINAL ARTICLE

Mutations in *COQ2* in Familial and Sporadic Multiple-System Atrophy

The Multiple-System Atrophy Research Collaboration

ABSTRACT

BACKGROUND

Multiple-system atrophy is an intractable neurodegenerative disease characterized by autonomic failure in addition to various combinations of parkinsonism, cerebellar ataxia, and pyramidal dysfunction. Although multiple-system atrophy is widely considered to be a nongenetic disorder, we previously identified multiplex families with this disease, which indicates the involvement of genetic components.

METHODS

In combination with linkage analysis, we performed whole-genome sequencing of a sample obtained from a member of a multiplex family in whom multiple-system atrophy had been diagnosed on autopsy. We also performed mutational analysis of samples from members of five other multiplex families and from a Japanese series (363 patients and two sets of controls, one of 520 persons and one of 2383 persons), a European series (223 patients and 315 controls), and a North American series (172 patients and 294 controls). On the basis of these analyses, we used a yeast complementation assay and measured enzyme activity of parahydroxybenzoate-polyprenyl transferase. This enzyme is encoded by the gene *COQ2* and is essential for the biosynthesis of coenzyme Q_{10}. Levels of coenzyme Q_{10} in lymphoblastoid cells and brain tissue were measured on high-performance liquid chromatography.

RESULTS

We identified a homozygous mutation (M78V-V343A/M78V-V343A) and compound heterozygous mutations (R337X/V343A) in *COQ2* in two multiplex families. Furthermore, we found that a common variant (V343A) and multiple rare variants in *COQ2*, all of which are functionally impaired, are associated with sporadic multiple-system atrophy. The V343A variant was exclusively observed in the Japanese population.

CONCLUSIONS

Functionally impaired variants of *COQ2* were associated with an increased risk of multiple-system atrophy in multiplex families and patients with sporadic disease, providing evidence of a role of impaired COQ2 activities in the pathogenesis of this disease. (Funded by the Japan Society for the Promotion of Science and others.)

The members of the Multiple-System Atrophy Research Collaboration are listed in the Appendix. Address reprint requests to Dr. Shoji Tsuji, Department of Neurology, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan, or at tsuji@m.u-tokyo.ac.jp.

This article was published on June 12, 2013, at NEJM.org.

N Engl J Med 2013;369:233-44.
DOI: 10.1056/NEJMoa1212115

MULTIPLE-SYSTEM ATROPHY IS A PROgressive neurodegenerative disease that is clinically characterized by autonomic failure in addition to various combinations of parkinsonism, cerebellar ataxia, and pyramidal dysfunction. The term multiple-system atrophy was introduced in 1969 to encompass the disease entities of olivopontocerebellar ataxia, striatonigral degeneration, and the Shy–Drager syndrome, on the basis of neuropathological findings in these disorders.[1] Multiple-system atrophy is characterized by the development of cytoplasmic aggregates of α-synuclein, primarily in oligodendroglia.[2-7] However, the pathogenic mechanisms underlying this disease remain unknown, making it difficult to develop effective therapies.

The disorder is classified into two subtypes: subtype C, characterized predominantly by cerebellar ataxia, and subtype P, characterized predominantly by parkinsonism.[8] Among patients with multiple-system atrophy, subtype C has been reported to be more prevalent than subtype P in the Japanese population (65 to 67% vs. 33 to 35%),[9,10] whereas subtype P has been reported to be more prevalent than subtype C in Europe (63% vs. 34%)[11] and North America (60% vs. 13%, with 27% of cases unclassified).[12] Although multiple-system atrophy has been defined as a nongenetic disorder until recently, several multiplex families with the disease have been described, indicating that strong genetic factors confer susceptibility to the disease.[13-15]

METHODS

PATIENTS AND MULTIPLEX FAMILIES

Patients with multiple-system atrophy were enrolled in the study on the basis of research protocols that were approved by the institutional review board at each participating center. Written informed consent was obtained from all participants.

The diagnosis of multiple-system atrophy was made on the basis of the current consensus criteria for the disease.[8] Four Japanese families (Families 1 through 4, whose members have been described previously[13]) and two additional Japanese families (Family 8 and Family 12) were enrolled in this study (Fig. 1). In Family 1, the parents were first-degree cousins, which is consistent with autosomal recessive inheritance. The clinical features of these families are summarized in Table S1 in the Supplementary Appendix, available with the full text of this article at NEJM.org.

Autopsy findings for Participants II-4[13] and II-8 in Family 1 and Participant II-6 in Family 8 showed widespread and abundant cytoplasmic aggregates of α-synuclein, primarily in oligodendroglia, in association with neurodegeneration in striatonigral and olivopontocerebellar structures. These findings confirmed the diagnosis of multiple-system atrophy.

PATIENTS WITH SPORADIC DISEASE AND CONTROLS

As with the multiplex families, the diagnosis of sporadic multiple-system atrophy was made on the basis of the current consensus criteria.[8] A total of 363 patients with multiple-system atrophy and 520 controls were included in the Japanese series, 223 patients and 315 controls in the European series, and 172 patients and 294 controls in the North American series (persons of European or Hispanic descent living in North America) (Text S2 and Table S2 in the Supplementary Appendix). Ancestry was determined by self-report on a multiple-choice questionnaire. We also enrolled an independent series of 2383 Japanese controls.

ASSOCIATION WITH OTHER NEURODEGENERATIVE DISEASES

To determine the specificity of the association between variants in candidate genes and multiple-system atrophy, we enrolled 2728 Japanese patients with Alzheimer's disease, 659 with Parkinson's disease, and 634 with amyotrophic lateral sclerosis (ALS). Their demographic characteristics are provided in Text S2 in the Supplementary Appendix.

LINKAGE ANALYSIS AND WHOLE-GENOME SEQUENCING

We performed parametric and nonparametric linkage analyses using Affymetrix SNP 6.0 arrays and software for linkage analysis.[16,17] The genomic DNA from Participant II-4 in Family 1 was subjected to four runs in an Illumina Genome Analyzer IIx (100-bp-long paired ends). We used BWA software[18] and SAMtools sequence-alignment mapping[19] with the default settings for alignment and variation detection against the human reference genome (National Center for Biotechnology Information build 36 [also known as hg18]).

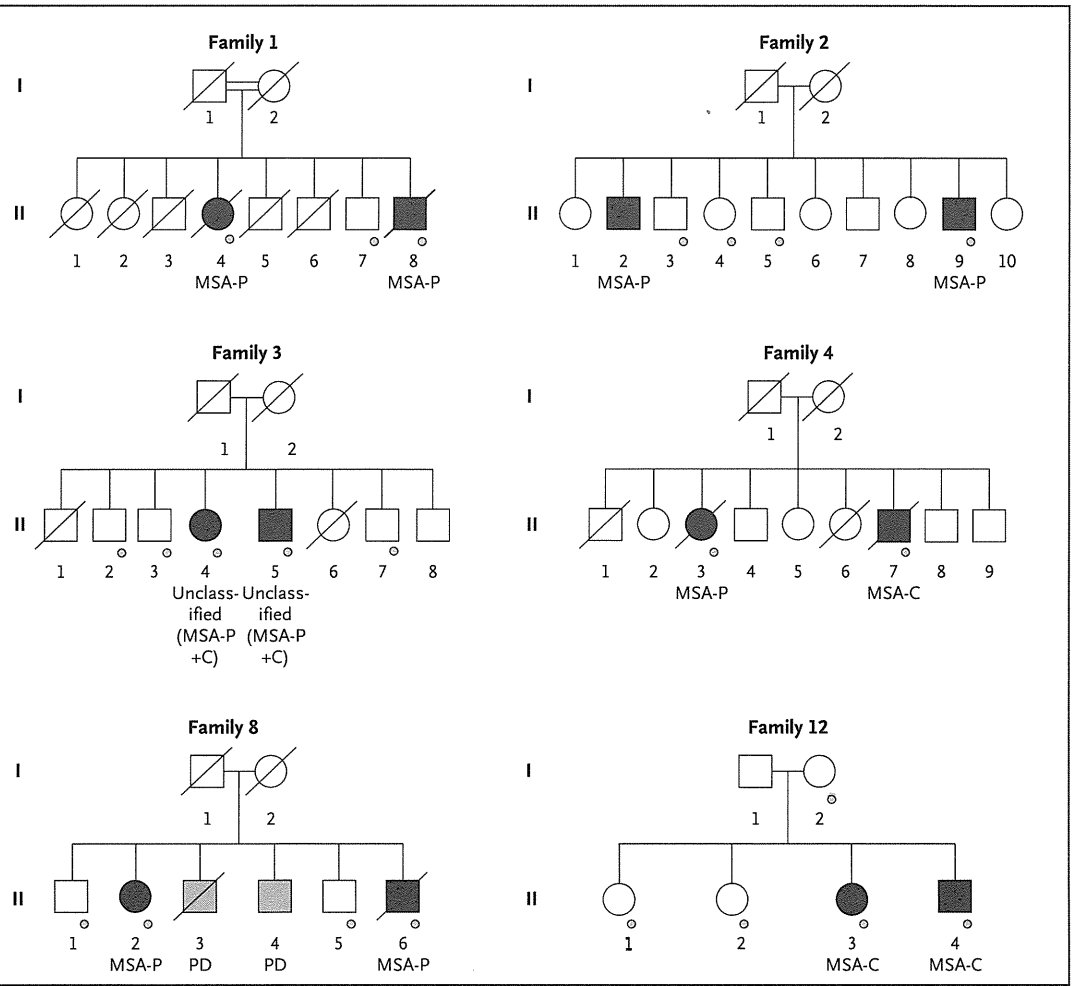

Figure 1. Pedigrees of Six Multiplex Families with Multiple-System Atrophy.

The affected siblings in Family 1 were born to consanguineous parents (first cousins).[13] In this family, the two patients with multiple-system atrophy (Participants II-4 and II-8) also had retinitis pigmentosa, which was not present in the other siblings. The diagnosis of definite multiple-system atrophy in three patients (Participants II-4 and II-8 in Family 1 and II-6 in Family 8) was confirmed at autopsy. In Family 8, two siblings (Participants II-3 and II-4) of the affected family members had Parkinson's disease (PD). In Family 1, in which homozygous M78V-V343A mutations in *COQ2* were identified, the parents (Participants I-1 and I-2), who were obligate carriers of the mutation, showed no overt signs of parkinsonism, cerebellar ataxia, or autonomic dysfunction, according to family report. In Family 12, in whom compound heterozygous R337X/V343A mutations were identified, Participants I-1 and I-2 (obligate carriers of the mutations) and the heterozygous carrier (Participant II-2) showed no overt signs of parkinsonism, cerebellar ataxia, or autonomic dysfunction on examination by a neurologist. Squares represent male family members, circles female family members, black symbols family members with multiple-system atrophy, gray symbols family members with Parkinson's disease, open symbols unaffected family members, slashes deceased family members, and small circles family members for whom genomic DNA samples were available. MSA-C denotes multiple-system atrophy of the cerebellar type, MSA-P multiple-system atrophy with predominant parkinsonism, and unclassified MSA-P+C similarly predominant parkinsonian and cerebellar signs.

ANALYSIS OF *COQ2* AND OTHER GENES ASSOCIATED WITH COENZYME Q_{10}

On the basis of linkage analysis and whole-genome sequencing, we sequenced *COQ2* and the other 11 genes involved in the biosynthetic pathway for coenzyme Q_{10} (*PDSS1*, *PDSS2*, *COQ3*, *COQ4*, *COQ5*, *COQ6*, *COQ7*, *ADCK3*, *COQ9*, *COQ10A*, and *COQ10B*), using the Sanger method (Table S3 in the Supplementary Appendix).

We prepared samples of mutant human *COQ2* complementary DNA (cDNA) by means of site-directed mutagenesis (Table S4 in the Supple-