Biallelic variants in *LIG3* cause a novel mitochondrial neurogastrointestinal encephalomyopathy

Elena Bonora,[1,†] Sanjiban Chakrabarty,[2,†] Georgios Kellaris,[3,†] Makiko Tsutsumi,[4,†] Francesca Bianco,[1,‡] Christian Bergamini,[5,‡] Farid Ullah,[3] Federica Isidori,[1] Irene Liparulo,[5] Chiara Diquigiovanni,[1] Luca Masin,[5] Nicola Rizzardi,[5] Mariapia Giuditta Cratere,[1,6] Elisa Boschetti,[1] Valentina Papa,[7] Alessandra Maresca,[8] Giovanna Cenacchi,[7] Rita Casadio,[9] Pierluigi Martelli,[9] Ivana Matera,[10] Isabella Ceccherini,[10] Romana Fato,[5] Giuseppe Raiola,[11] Serena Arrigo,[10] Sara Signa,[10] Angela Rita Sementa,[10] Mariasavina Severino,[10] Pasquale Striano,[10] Chiara Fiorillo,[10] Tsuyoshi Goto,[12] Shumpei Uchino,[13,14] Yoshinobu Oyazato,[15] Hisayoshi Nakamura,[16] Sushil K. Mishra,[17] Yu-Sheng Yeh,[12] Takema Kato,[4] Kandai Nozu,[18] Jantima Tanboon,[16] Ichiro Morioka,[19] Ichizo Nishino,[16] Tatsushi Toda,[20] Yu-ichi Goto,[21] Akira Ohtake,[22] Kenjiro Kosaki,[23] Yoshiki Yamaguchi,[24] Ikuya Nonaka,[16] Kazumoto Iijima,[18] Masakazu Mimaki,[13] Hiroki Kurahashi,[4] Anja Raams,[2] Alyson MacInnes,[25] Mariel Alders,[26] Marc Engelen,[27] Gabor Linthorst,[25] Tom de Koning,[28] Wilfred den Dunnen,[29] Gerard Dijkstra,[30] Karin van Spaendonck,[26] Dik C. van Gent,[2] Eleonora M. Aronica,[31] Paolo Picco,[10] Valerio Carelli,[7,8,§] Marco Seri,[1,§] Nicholas Katsanis,[3,§] Floor A. M. Duijkers,[26,§] Mariko Taniguchi-Ikeda[4,18,32,§] and Roberto De Giorgio[33,§]

[†,‡,§]These authors contributed equally to this work.

Abnormal gut motility is a feature of several mitochondrial encephalomyopathies, and mutations in genes such as *TYMP* and *POLG*, have been linked to these rare diseases. The human genome encodes three DNA ligases, of which only one, ligase III (*LIG3*), has a mitochondrial splice variant and is crucial for mitochondrial health.

We investigated the effect of reduced LIG3 activity and resulting mitochondrial dysfunction in seven patients from three independent families, who showed the common occurrence of gut dysmotility and neurological manifestations reminiscent of mitochondrial neurogastrointestinal encephalomyopathy. DNA from these patients was subjected to whole exome sequencing. In all patients, compound heterozygous variants in a new disease gene, *LIG3*, were identified. All variants were predicted to have a damaging effect on the protein. The *LIG3* gene encodes the only mitochondrial DNA (mtDNA) ligase and therefore plays a pivotal role in mtDNA repair and replication. *In vitro* assays in patient-derived cells showed a decrease in LIG3 protein levels and ligase activity. We demonstrated that the *LIG3* gene defects affect mtDNA maintenance, leading to mtDNA depletion without the accumulation of multiple deletions as observed in other mitochondrial disorders. This mitochondrial dysfunction is likely to cause the phenotypes observed in these patients. The most prominent and consistent clinical signs were severe gut dysmotility and neurological abnormalities, including leukoencephalopathy, epilepsy, migraine, stroke-like episodes, and neurogenic bladder. A decrease in the number of myenteric neurons, and increased fibrosis and elastin levels were the most prominent changes in the gut. Cytochrome *c* oxidase (COX) deficient fibres in skeletal muscle were also observed. Disruption of *lig3* in zebrafish reproduced the brain alterations and impaired gut transit *in vivo*.

In conclusion, we identified variants in the *LIG3* gene that result in a mitochondrial disease characterized by predominant gut dysmotility, encephalopathy, and neuromuscular abnormalities.

1   Department of Medical and Surgical Sciences, St. Orsola-Malpighi Hospital, University of Bologna, Bologna, 40138, Italy
2   Department of Molecular Genetics, Erasmus MC, Rotterdam, 3000 CA, The Netherlands
3   Center for Human Disease Modeling, Duke University, Durham, NC 27710, USA
4   Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, Aichi, 470-1192, Japan
5   Department of Pharmacy and Biotechnology, University of Bologna, Bologna, 40126, Italy
6   Division of Genetics and Cell Biology, San Raffaele Scientific Institute, Milan, 20132, Italy
7   Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, 40123, Italy
8   IRCCS Istituto delle Scienze Neurologiche di Bologna, Programma di Neurogenetica, Bologna, 40139, Italy
9   Biocomputing Group, Department of Biological, Geological, Environmental Sciences, University of Bologna, Bologna, 40126, Italy
10 IRCCS Istituto Giannina Gaslini, Genova, 16128, Italy
11 Department of Paediatrics, Pugliese-Ciaccio Hospital, Catanzaro, 88100, Italy
12 Laboratory of Molecular Function of Food, Division of Food Science and Biotechnology, Graduate School of Agriculture, Kyoto University, Uji, 611-0011, Japan
13 Department of Pediatrics, Teikyo University School of Medicine, Tokyo, 173-8605, Japan
14 Department of Pediatrics, Graduate School of Medicine, The University of Tokyo, Tokyo, 113-0033, Japan
15 Department of Pediatrics, Kakogawa Central City Hospital, Kakogawa, Hyogo, 675-8611, Japan
16 Department of Neuromuscular Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Tokyo, 187-8502, Japan
17 Glycoscience Group, National University of Ireland, Galway, H91 CF50, Ireland
18 Department of Pediatrics, Kobe University Graduate School of Medicine, Hyogo, 650-0017, Japan
19 Department of Pediatrics and Child Health, Nihon University School of Medicine, Tokyo, 173-8610, Japan
20 Department of Neurology, Graduate School of Medicine, The University of Tokyo, Tokyo, 113-0033, Japan
21 Department of Mental Retardation and Birth Defect Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Tokyo, 187-8502, Japan
22 Department of Pediatrics & Clinical Genomics, Faculty of Medicine, Saitama Medical University, Saitama, 350-0495, Japan
23 Center for Medical Genetics, Keio University School of Medicine, Tokyo, 160-8582, Japan
24 Laboratory of Pharmaceutical Physical Chemistry, Tohoku Medical and Pharmaceutical University, Miyagi, 981-8558, Japan
25 Department of Metabolic Diseases, Amsterdam UMC, University of Amsterdam, Amsterdam, 1100 DD, The Netherlands
26 Department of Clinical Genetics, Amsterdam UMC, University of Amsterdam, Amsterdam, 1100 DD, The Netherlands
27 Department of Neurology, Amsterdam UMC, University of Amsterdam, Amsterdam, 1100 DD, The Netherlands
28 Department of Metabolic Diseases, UMCG, Groningen, 9700 RB, The Netherlands
29 Department of Pathology, UMCG, Groningen, 9700 RB, The Netherlands
30 Department of Gastroenterology, UMCG, Groningen, 9700 RB, The Netherlands
31 Department of Pathology, Amsterdam UMC, University of Amsterdam, Amsterdam, 1100 DD, The Netherlands
32 Department of Clinical Genetics, Fujita Health University Hospital, Aichi, 470-1192, Japan
33 Department of Morphology, Surgery and Experimental Medicine, St. Anna Hospital, University of Ferrara, Ferrara, 44124, Italy

Correspondence to: Professor Roberto De Giorgio
Department of Morphology, Surgery and Experimental Medicine, University of Ferrara, 44124 Ferrara, Italy
E-mail: roberto.degiorgio@unife.it

Correspondence may also be addressed to: Dr Floor A. Duikers
Department of Clinical Genetics. Amsterdam University Medical Centers, University of Amsterdam, 1105 AZ,
Amsterdam, The Netherlands
E-mail: f.a.duijkers@amsterdamumc.nl

Associate professor Mariko Taniguchi-Ikeda
Department of Clinical Genetics, Fujita Health University Hospital, 1-98, Dengakugakubo, Kutsukake-cho
Toyoake, Aichi 470-1192, Japan
E-mail: mtani@fujita-hu.ac.jp or taniguchi_mariko@me.com

Introduction

The human genome encodes three DNA ligases (I, III and IV). All DNA ligases are expressed in the nucleus, but only ligase III (*LIG3*) has a mitochondrial splice variant. Nuclear LIG3 interacts with X-ray repair cross-complementing protein 1 (XRCC1) for DNA maintenance by base excision repair, but other nuclear ligases can compensate for defects in the *LIG3* gene, which encodes the LIG3 protein.[1-5] However, LIG3 is crucial in mitochondria, as it is the only ligase responsible for mitochondrial DNA (mtDNA) replication and maintenance, working along with mtDNA polymerase gamma (POLG) and other mtDNA replisome factors.[6] The lethality of a *LIG3* null mutation can be alleviated by targeting another DNA ligase to mitochondria, whereas overexpression of the LIG3 protein appears to increase the resistance to oxidative damage in mitochondria.[7] Therefore, reduced LIG3 activity is expected to affect mitochondrial health and would lead to diseases resulting from mitochondrial dysfunction.

A key clinical feature of several mitochondrial diseases is severely abnormal gut motility, such as in patients with chronic intestinal pseudo-obstruction (CIPO). This is an impairment of gut propulsion, mimicking a mechanical obstruction without detectable anatomical causes.[8-10] Mitochondrial encephalomyopathies may also be characterized by prevalent leukoencephalopathy caused by mitochondrial dysfunction.[6] Mitochondrial neurogastrointestinal encephalopathy (MNGIE), the 'tip of the ice-berg' of such rare diseases, is caused by mutations in the *TYMP* gene, which encodes thymidine phosphorylase.[11] Similar phenotypes are also caused by mutations in *POLG*, or mutations in the mtDNA itself, as observed in patients with mitochondrial encephalomyopathy with lactic acidosis and stroke-like episodes (MELAS).[12]

We here report a novel mitochondrial gastrointestinal encephalomyopathy caused by biallelic variants in *LIG3*,[1,3,13,14] leading to a syndrome predominantly characterized by severe gut dysmotility (i.e. CIPO) and encephalomyopathy.

Patients and methods

In the present study, analyses of the patients were performed in three different institutes, which led to some differences in the methods used per family, as detailed below.

Ethical approval

Data from patients and controls were handled in accordance with the local ethics committees (St. Orsola Hospital Ethics Committee; Institutional Review Board of Kobe University School of Medicine and Medical Ethical Board of Amsterdam University) and analyses were performed after obtaining written informed consent from the patients or patients' parents, according to the declaration of Helsinki.

Next-generation sequencing analysis

Families were recruited from different institutions as follows: Italy, Family 1, Patients 1-1, 1-2, and 1-3; The Netherlands, Family 2, Patients 2-1 and 2-2; and Japan, Family 3, Patients 3-1 and 3-2. Details of patients are provided in the Supplementary material.

Sample collection and immunohistochemistry

Muscle, gut, skin and urothelial sediment cells (USCs) were collected and histologically analysed as described in the Supplementary material. Procedures for western blot analyses are also provided in the Supplementary material.

Immunofluorescence: Family 1

Immunofluorescence analysis was performed as described in the Supplementary material. The following primary antibodies were used: rabbit anti-LIG3 (26583-1-AP; 1:50, ProteinTech). Secondary antibodies used were the following: Alexa Fluor® 488 goat anti-rabbit IgG, Alexa Fluor® 555 goat anti-rabbit (all from Abcam; 1:800).

Analysis of mtDNA content by real-time quantitative PCR

Family 1

Total DNA was extracted from fibroblasts using the Qiagen Mini kit and from muscle biopsies by standard phenol-chloroform extraction. Real-time quantitative PCR was used to assess mtDNA content according to previously validated methods.[15]

Family 2

The mtDNA/nuclear DNA (nDNA) ratio was measured in duplicate on freshly obtained muscle tissue in a diagnostic setting (Radboud UMC Laboratory, The Netherlands) and compared to age-matched controls (above the age of 21 years) from reports in the literature.[16]

Family 3

Total DNA was extracted from fibroblasts or myoblasts using DNeasy Blood and Tissue Kit (Qiagen). The Human mtDNA Monitoring Primer Set (Takara) was used for amplification of mtDNA and nDNA. Real-time PCR was performed using SYBR Premix Ex Taq II (Takara) and 7300 Real-Time PCR System (Thermo Fisher) according to previously published protocols.[17,18]

DNA ligase assay: Patient 2-1

Mitochondrial protein extracts were prepared from primary fibroblasts as previously described.[19] Ligation of linearized pUC18 plasmid using mitochondrial extracts was performed as previously reported.[19] Briefly, PstI-linearized pUC18 was incubated at 16°C for 16 h with mitochondrial extracts

from control and mutant samples. Plasmid DNA was subsequently deproteinized, purified and separated on 1.2% analytical agarose gels. Reactions were supplemented by the addition of ATP; T4 DNA ligase (New England Biolabs) was used as a positive control. Each ligation assay was performed twice.

Quantitative PCR-based analysis of mtDNA repair activity

Wild-type and *LIG3*-mutant fibroblasts (10^6) from Patient 2-1 were seeded in 6-cm dishes 16–18 h prior to the experiments. Cells were washed once with Ham's F-10 medium without supplements, and the conditioned medium was used subsequently. Cells were exposed to 200 μM H_2O_2 for 15 min and harvested immediately or cultured with conditioned medium for 6 h. High-molecular weight genomic DNA was extracted from untreated or H_2O_2-treated wild-type and *LIG3*-mutant cell lines using Wizard® SV Genomic DNA Purification System (Promega). The quantitative PCR reaction was performed with the LongAmp® Taq PCR kit (New England Biolabs) as follows: 15 ng total DNA was added to a reaction mix of 50 μl with 100 ng/μl bovine serum albumin, 200 μM dNTPs each, 1–3 mM MgO(Ac)$_2$, and 20 pmol each of the two primers. PCR was performed with primer pairs to amplify an 8.9-kb fragment of mtDNA (long), and in a separate reaction to amplify a 221-bp fragment of mtDNA (short), to calculate mtDNA damage and copy number in untreated and H_2O_2-treated samples, respectively.[17] The primer sequences used were as previously described.[17] PCR products for both long and short mitochondria amplicons were quantified using the PicoGreen™ dsDNA quantification assay (Thermo Fisher). Fluorescent values obtained from the short PCR products of each sample were used to normalize the results from the long PCR products. These values were used to estimate the average number of lesions per 10 kb mitochondrial genome using a Poisson distribution.[17] Relative amplification was calculated by comparing treated samples with untreated samples for both wild-type and mutant cells and the unpaired *t*-test was performed for statistical analysis.

Ethidium bromide treatment: Patient 3-2

Fibroblasts were cultured in the presence of 50 ng/ml ethidium bromide (EtBr), 1 mM pyruvate, and 50 μg/ml uridine for 9 days followed by culturing without EtBr for 7 days. Subcultures were performed every 2 to 3 days to keep the confluency of the cells between 30% and 80%.

Mitochondrial oxygen consumption

Patient 1-1

To measure mitochondrial oxygen consumption 1.5×10^6 cells for each cell type (wild-type and Patient 1-1-derived fibroblasts) were harvested, washed in PBS, resuspended in complete medium and assayed for oxygen consumption at 37°C using a thermostatically regulated oxygraph chamber (Instech

Mod.203). Basal respiration was measured in Dulbecco's modified Eagle medium and compared with data obtained after the injection of oligomycin (1 μM) and carbonyl cyanide 4-(trifluoromethoxy) phenylhydrazone (FCCP) (1–6 μM). Antimycin A (5 μM) was added at the end of the experiments to completely block the mitochondrial respiration.[20] Respiratory rates were expressed in nmol O_2/min/mg of total protein, which was determined by the Lowry assay.

Patient 3-2

In USCs, basal and maximal oxygen consumption rates (OCR) were analysed using Seahorse Bioscience XF-24 extracellular flux bioanalyzer (Agilent) and Seahorse XF Cell Mito Stress Test. USCs were seeded onto XF 24-well microplates in growth medium at 1.5×10^5/well (*n* = 5). The next day, growth medium was replaced with assay medium supplemented with glucose, pyruvate and L-glutamine, and the pH was adjusted to 7.4. Cells were equilibrated prior to Mito Stress testing in the analysis medium for 30 min at 37°C in a CO_2-free incubator. ATP-linked respiration was determined by oligomycin (1 μM) and maximal respiration was induced using FCCP (1 μM). Nonmitochondrial respiration was determined after injection of rotenone and antimycin A (1 μM each). Data were analysed using XF Cell Mito Stress Test Report Generator. After assay completion, cells were rinsed with PBS and frozen. After thawing, double stranded DNA of each well was measured using DNA Quantity kit (Cosmo BIO) using a microplate reader.

ATP determination: Patient 1-1

Nucleotides were extracted and detected using a Kinetex C18 column (250 × 4.6 mm, 100 Å, 5 μm; Phenomenex) with a two pump Agilent 1100 series system (Agilent).[21] Absorbance (260 nm) was monitored with a photodiode array detector (Agilent 1100 series system). Nucleotide peaks were identified by comparison and coelution with known standards and quantification by peak area measurement compared with standard curves.

Zebrafish husbandry

Embryos were obtained by natural mating of wild-type adults (TU/AB strain, Aquatica BioTech). Adults were maintained on a 14-h/10-h day/night cycle. Larvae were nurtured in embryo media (0.3 g/l NaCl, 75 mg/l $CaSO_4$, 37.5 mg/l $NaHCO_3$ and 0.003% methylene blue) at 28°C until phenotypic analysis at 3 days post fertilization (dpf), 5 dpf and 8 dpf, respectively. All zebrafish studies were approved by the Duke University Institutional Animal Care and Use Committee.

Gene suppression and complementation in zebrafish embryos

A splice-blocking morpholino (MO) was designed to target the *Danio rerio lig3* exon 11 splice donor (e11i11) and was

synthesized by Gene Tools LLC. To identify the optimal dose for *in vivo* complementation, a dose curve was generated by injecting 2, 4 and 6 ng (1 nl MO per embryo; one- to four-cell stage) into TU/AB embryos. Total RNA was extracted from 1 dpf embryos (15–20 per condition) using TRIzol® (Thermo Fisher) and was reverse transcribed using the SuperScript® III Reverse Transcriptase kit (Thermo Fisher). We used cDNA as the template for RT-PCR with primers flanking the *lig3* MO target site. PCR products were purified with QIAquick gel extraction kit (Qiagen) and cloned in TOPO-TA cloning vector (Thermo Fisher). Plasmids were purified and Sanger sequenced according to standard protocols. For rescue experiments, wild-type human *LIG3* open reading frame (Ultimate ORF collection; Clone ID: IOH40893) was subcloned into the pCS2 + vector by LR Clonase II®-mediated recombination (Thermo Fisher). To produce constructs containing the sequences of patient-associated variants, site-directed mutagenesis was performed.[22] Constructs were linearized with *NotI*, and the resulting template was transcribed with the mMessage mMachine® SP6 transcription kit (Thermo Fisher). Unless otherwise noted, 6 ng *lig3* MO was used either alone or in combination with 200 pg *LIG3* mRNA for *in vivo* complementation studies.

CRISPR/Cas9 genome editing of zebrafish embryos

Guide (g) RNAs targeting the *lig3*-coding region were designed with CHOPCHOP. A GeneArt precision gRNA synthesis kit (Thermo Fisher) was used to *in vitro* transcribe gRNAs, followed by injection of 1 nl of injection cocktail containing 100 pg/nl gRNA and 200 pg/nl Cas9 protein (PNA Bio) directly into the cell of TU/AB embryos (one-cell stage). Heteroduplex assays were used to determine targeting efficiency in founder (F0) mutants.[23,24] Genomic DNA was extracted from 2 dpf embryos. The region flanking the gRNA target site was amplified by PCR; PCR products were denatured (95°C for 2 min), reannealed (–2°C/s to 85°C and –0.1°C to 25°C), separated on a 15% TBE 1.0-mm poly-acrylamide gel, stained with EtBr and imaged using the Chemi-Doc system (Bio-Rad). To estimate the percentage of mosaicism of *lig3* F0 mutants (*n* = 5), PCR products were gel-purified with QIAquick Gel extraction kit (Qiagen), cloned into the TOPO-TA vector (Thermo Fisher) and plasmids were isolated from individual colonies (*n* = 10–12 colonies/embryo) and Sanger sequenced according to standard procedures.

Phenotypic analysis of zebrafish larvae

Whole-mount immunostaining with anti-acetylated tubulin (T7451, Sigma-Aldrich) and anti-HuC/D (A-21271, Thermo Fisher) antibodies was performed to analyse the brain and enteric neurons along the gastrointestinal tract, respectively.

Smooth muscle morphology of the gastrointestinal tract was assessed using the anti-phospho-myosin light chain 2 antibody (Ser19; 9970, CST). To quantify neurons in the cerebellar area, whole-mount acetylated tubulin immunostaining was performed.[22,25] Dorsal images were acquired manually with an AxioZoomV.16 microscope (Zeiss) and Axiocam 503 monochromatic camera (Zeiss). Cerebellar structures of interest were measured using ImageJ software. Total cerebellar area was measured blindly by two experts on acetylated tubulin-stained larvae by outlining the structures with a fluorescent signal. Quantitative evaluation of enteric neurons was performed via HuC/D staining in injected embryos at 5 dpf. For phospho-myosin staining, injected embryos were analysed at 8 dpf. Fluorescent cell counts were performed with ImageJ software. To assess peristalsis of the zebrafish gut, embryos at 8 dpf were anesthetized using 1× tricaine embryo media. Gastrointestinal motility was recorded with 2-min time lapse videos.

Statistical analysis

Chi-squared (χ^2), ANOVA and two-tailed parametric *t*-tests were performed as reported in the corresponding 'Results' sections and figure legends, using GraphPad Prism software v.7.00 (GraphPad). *P*-values < 0.05 were considered to indicate a statistically significant difference between two groups. For statistical analysis of cerebellum measurements in zebrafish, a non-parametric one-way ANOVA followed by the Tukey multiple comparison test was performed. For the gut peristalsis assay, the χ^2 test was used to perform pairwise statistical comparisons across experimental conditions.

The analysis of mtDNA deletions by droplet digital PCR, RNA extraction and transcript analysis, ultrastructural analysis, mitochondrial network analysis, reactive oxygen species quantification, and measurement of transmembrane potential are described in the Supplementary material.

Data availability

The authors confirm that the data supporting the findings of this study are available within the article and its Supplementary material.

Results

Clinical features of the patients

In seven affected individuals from three separate families compound heterozygous variants in the *LIG3* gene were identified by whole exome sequencing (WES) (Fig. 1A–C). These patients showed a largely overlapping phenotype and an overview of their clinical features is presented in Table 1. The clinical histories of each patient can be found in the Supplementary material. The clinical features of these patients for many aspects resembled the mitochondrial

disease MNGIE. Severe dysmotility of the gut was present in all patients and most patients fulfilled the diagnostic criteria for CIPO.[8-10,26] Brain MRI of all patients showed leukoencephalopathy and/or progressive cortical atrophy, whereas cerebellar atrophy was observed only in patients from Family 3 (Fig. 1D–F). In addition to the MNGIE-like features, all patients had other neurological features including epilepsy, stroke-like episodes, migraine and developmental delay, reminiscent of MELAS. Further clinical manifestations included neurogenic bladder and macular degeneration, which were present in most patients. Hearing loss was only observed in one patient (Table 1). Increased lactate/pyruvate ratio in the CSF was found in the patients of Family 3. Although the clinical phenotype was quite consistent across the three families, the age of onset and disease severity differed considerably ranging from paediatric severe disease with premature death to adult cases. In particular, the two patients of Family 3 manifested as infants with severe gut dysmotility, severe developmental delay and epilepsy, interpreted as West syndrome.[27] One of the two children in this family died at the age of 2 years. Conversely, the disease onset occurred much later in the three patients from Family 1 (between the ages of 9 and 10 years), with a rapid disease progression. The presentation in Family 2 was even later, with some features detectable at paediatric age, but the most severe manifestations only occurring at adult age with a slower disease progression compared with the other two families.

Evidence of mitochondrial dysfunction in patient tissue

Skeletal muscle biopsies of patients from all three families (Patients 1-1, 2-2 and 3-2) showed alterations that are observed in patients with mitochondrial myopathies, e.g. succinate dehydrogenase (SDH) hyper-reactivity, reduced cytochrome *c* oxidase (COX)-staining (Supplementary Fig. 1A–C) and altered myofibre structure (Patient 3-2) on electron microscopy (Supplementary Fig. 1D). Nevertheless, these muscle alterations were variable/subtle, as compared with the severe clinical phenotype observed in the gut and CNS of the patients.

Identification of *LIG3* mutations

The recurrent phenotype of CIPO and neuromuscular involvement in all sibs, along with unaffected parents, suggested a recessive disorder, with a possible mitochondrial aetiology. In all three families we identified compound heterozygous variants in *LIG3*, shared by the corresponding affected sibs and inherited from the corresponding heterozygous healthy parents (Fig. 2A and Supplementary Fig. 2A–C). The variants in Family 1 (p.K537N and p.G964R missense variants) were not present in public databases (Database of Single Nucleotide Polymorphisms, dbSNP; Genome Aggregation Consortium, gnomAD; ClinVar; 1000Genomes) or in an inhouse database (650 Italian

exomes). In Family 2 the missense variant p.C999Y was novel and the premature stop codon p.R267Ter was present in dbSNP with a very low minor allele frequency (MAF) (Supplementary Table 1). In Family 3 we identified the missense variant p.P609L and the premature stop codon p.R811Ter. Both variants were present in dbSNP and gnomAD with a very low MAF (Supplementary Table 1). No deleterious variants in other genes compatible with a recessive or *de novo* inheritance were detected. All variants mapped to the conserved LIG3 domains[28,29] and *in silico* analyses predicted their pathogenicity (Fig. 2A, Supplementary Fig. 2D and Supplementary Tables 1 and 2).[30,31] No causative point mutations or deletions in mtDNA were identified in Patients 1-1, 2-2 and 3-2.

LIG3 mRNA and protein analysis

Detailed analyses on the consequences of the *LIG3* mutations were conducted using primary skin fibroblasts derived from Patients 1-1, 2-1 and 3-2. The missense p.K537N variant identified in Family 1, mapping to the most 3′ nucleotide of exon 9, was predicted to alter splice-site recognition (Human Splicing Finder v3.0, HSF; Supplementary Fig. 3A). RT-PCR with the primers in exon 8 and exon 10 showed only one band of the expected size for exon 9 inclusion in control cDNA, but the cDNA from patient cells also gave rise to a shorter transcript lacking exon 9 (Fig. 2B). Exon 9 skipping results in a protein lacking 52 amino acids (495–537 in-frame deletion) in the conserved adenylation-nucleotidyltransferase (NTase) domain of the ATP- and POLG-binding regions (Fig. 2C).[13,14,28] Exon skipping in this mutant protein results in the loss of ∼5 kDa of the LIG3 protein. However, we detected a severe reduction in the total amount of LIG3 protein without the appearance of a shorter LIG3 species on western blot analysis of control and patient fibroblast lysates (Fig. 2D) (a severe reduction in expression is also shown in Supplementary Fig. 3B–D). Concordantly, transient transfection of the cDNA encoding for human *LIG3* carrying the exon 9 deletion in HEK293 cells did not show any band smaller than that of the wild-type LIG3 protein (Supplementary Fig. 3E), suggesting that the protein is early degraded. Although in Patient 1-1 fibroblasts the exon 9 variant gave rise to exon skipping, we also evaluated its potential effect as missense change p.K537N in transient transfection of HEK293: the protein product was present (Supplementary Fig. 3F), but the co-immunoprecipitation assay using an antibody against LIG3, followed by western blotting for POLG, showed a very low affinity for POLG compared to wild-type LIG3 (Supplementary Fig. 3G). Based on western blot results in fibroblasts, we assumed that, in the Patient 1-1 fibroblasts, only the other human LIG3 allele (p.G964R) was expressed at the protein level and that this amino acid change caused reduced stability. In line with these observations, transient transfection of the cDNA encoding for human LIG3 with the missense change p.G964R in HEK293 revealed that LIG3 expression was comparable to that of the empty vector, at difference with
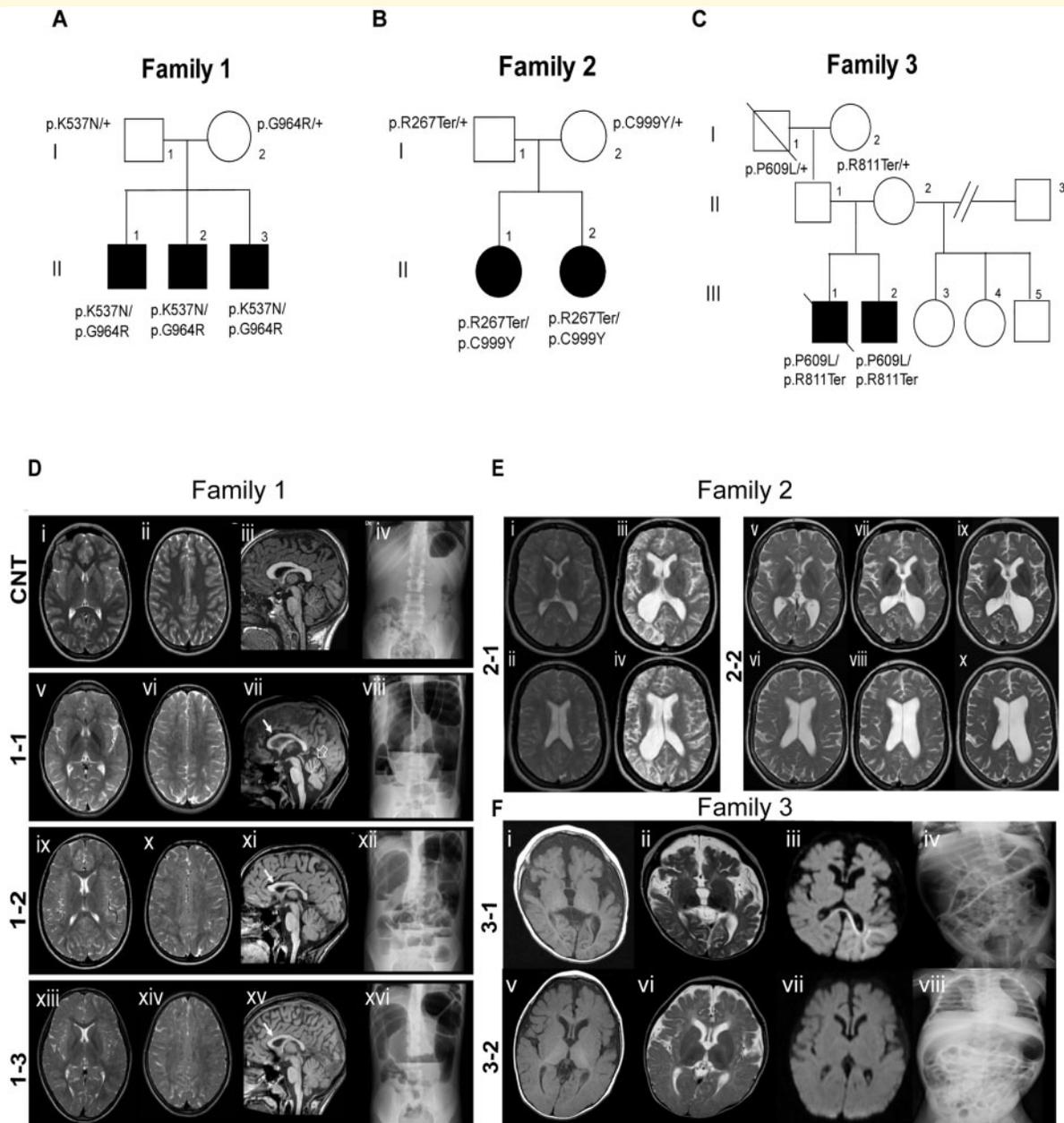
Figure 1 Identification of *LIG3* mutations. (**A–C**) Pedigrees of the three families with *LIG3* mutations. Filled and open symbols denote affected and healthy individuals, respectively; diagonal bars indicate deceased members. (**A**) Family 1: inheritance of paternal p.K537N and maternal p.G964R variants in the three affected brothers. (**B**) Family 2: inheritance of paternal p.R267Ter and maternal p.C999Y variants in the two affected sisters. (**C**) Family 3: inheritance of paternal p.P609L and maternal p.R811Ter variants in the two affected brothers. (**D**) Brain MRI and abdominal X-ray images of the patients in Family 1. (**ii**) Axial T_2-weighted, and (**iii**) sagittal T_1-weighted images of a control subject (CNT) demonstrating normal supratentorial white matter, corpus callosum and cerebellar vermis. Corresponding images for Patients 1-1 (**v–vii**), 1-2 (**ix–xi**) and 1-3 (**xiii–xv**) demonstrate leukoencephalopathy (involving periventricular and subcortical white matter, corpus callosum and internal capsules) characterized by diffuse but irregular T_2 hyperintensities with a peculiar 'leopard-skin' pattern due to the sparing of small white matter islets. Arrows (**vii**, **xi** and **xv**) indicate mild callosal hypoplasia. Patient 1-1 shows additional signs of cerebellar atrophy (**vii**, open arrow). (**viii–xvi**) Abdominal X-ray images displaying bowel dilatation with air-fluid levels compared with the control subject. (**E**) Brain MRI of the patients of Family 2. Axial T_2-weighted images of Patients 2-1 (*left*) and 2-2 (*right*) at different time points (**i**, t = 0 years and **iii**, t = + 3 years; **v** and **vi**, t = 0; **vii** and **viii**, t = + 2 years, **ix** and **x**, t = + 4 years). In Patient 2-1 there is diffuse hyperintensity of the entire cerebral white matter. After several stroke-like episodes, progressive atrophy with asymmetrical 'ex vacuo' dilatation of the ventricular system was noted. Patient 2-2 displays a similar pattern and evolution over time. (**F**) Brain MRI and abdominal X-ray images of patients in Family 3. In Patient 3-1 axial (**i**) T_1-weighted, (**ii**) T_2-weighted and (**iii**) diffusion-weighted images demonstrate brain atrophy and restricted diffusion in the left temporo-occipital, corticosubcortical regions extending to the corpus callosum. In Patient 3-2, brain MRI demonstrates generalized brain atrophy (**v–vii**). Abdominal X-ray images show marked bowel distension (**iv** and **viii**).

Table 1 Clinical features of the affected sibs with LIG3 compound heterozygous variants

	Patient 1-1	Patient 1-2	Patient 1-3	Patient 2-1	Patient 2-2	Patient 3-1	Patient 3-2
Current age, years	17	13	11	40	39	2 (deceased)	4
Age of onset, years	9	10	10	18	18	1	2
GI symptoms and surgery	Subacute and/or recurrent episodes of intestinal obstruction without an occluding lesion, malnutrition (parenteral nutrition), diarrhoea, ileostomy (at 16 years)	Subacute and/or recurrent episodes of intestinal obstruction without an occluding lesion, malnutrition (parenteral nutrition), abdominal distension and constipation, ileostomy (at 11 years)	Subacute and/or recurrent episodes of intestinal obstruction without an occluding lesion, diarrhoea/constipation, ileostomy (at 10 years)	Chronic intestinal pseudo-obstruction (18 years), TPN (from 33 years), enlarged liver, cholecystolithiasis	Chronic intestinal pseudo-obstruction (at 18 years), TPN (at 37 years).	Severe abdominal distension, hepatomegaly, NG tube (at 1 month)	Severe abdominal distension, hepatomegaly, NG tube (at 1 month), IVH (at 2 years)
Brain imaging	Diffuse leukoencephalopathy, cerebellar atrophy	Diffuse leukoencephalopathy, thin corpus callosum	Diffuse leukoencephalopathy	After a stroke-like event at 33 years, increasing degree of atrophy and gliosis over time, most prominent in parietal-occipital right hemisphere	Progressive diffuse leukoencephalopathy after stroke-like episodes	Leukoencephalopathy, cerebellar atrophy	Leukoencephalopathy, cerebellar atrophy
Neurological signs	Pyramidal signs, mild ataxia, dysmetria	Pyramidal signs	Pyramidal signs	Unilateral neglect and cognitive deterioration/decline after stroke-like episode at 33 years	Nystagmus and diminished feeling in the legs and cognitive deterioration at 36 years	Involuntary movements, jerks	Involuntary movements, jerks
Headache	Yes, complicated with motor deficits	Yes, complicated with motor deficits	Yes, complicated with motor deficits	Yes, migraine with aura, starting from teenage years	Yes, migraine with aura, starting from paediatric years	N/A	N/A
Psychiatric features	Repetitive pattern of behaviours and interests	Repetitive pattern of behaviours	None	None	None	N/A	N/A
Urogenital abnormalities	Neurogenic bladder, right mild pyelectasia	Neurogenic bladder	None	Possible voiding dysfunction, frequent retention bladder	Unilateral duplicated collecting system, frequent retention bladder	N/A	Neurogenic bladder
Additional features	Recurrent arthralgias, stomatitis, macular degeneration	Macular degeneration	None	Macular degeneration	Macular degeneration, frequent infections (recent years)	Apnoea, recurrent infections, pneumonia	Cataracts, hearing loss, dysphagia

GI = gastrointestinal; IVH = intravenous hydration; N/A = not assessed; NG = nasogastric; TPN = total parenteral nutrition.

the HEK293 cells transfected with wild-type LIG3 (Supplementary Fig. 3F). Molecular modelling showed that p.G964 interacts with p.R558 of murine XRCC1 (a residue conserved between human and murine proteins). The LIG3 variant p.G964R was predicted to hamper the interaction between LIG3 and XRCC1 because of the close proximity (0.44 nm) of two positively charged amino acids in the interacting site (p.R558 in XRCC1 and mutated p.R964; Supplementary Fig. 3H). This impairment prevented LIG3 translocation into the nucleus (Supplementary Fig. 3I).[1-3]

The compound heterozygous mutations in Family 2 (p.R267Ter and p.C999Y) were also analysed. RT-PCR on Patient 2-1 fibroblasts and controls showed that the c.799C>T allele (p.R267Ter variant) was not expressed, and only the c.2996G>A allele (p.C999Y variant) was present in the final transcript of the patient's fibroblasts (Fig. 2E), suggesting that mRNA carrying the c.799C>T variant (premature stop codon p.R267Ter) was degraded by nonsense-mediated decay. Transient transfection of HEK293 cells with the plasmid encoding the LIG3 p.C999Y variant did not show an increase in expression, compared to cells transfected with the empty vector, as seen for the vector with wild-type LIG3 (Supplementary Fig. 3F). Molecular modelling showed that the variant p.C999Y is predicted to severely destabilize LIG3 structure ($\Delta\Delta G$ = –2.02 kcal/mol; Supplementary Fig. 3H).[28-31] Western blot analysis in patient's fibroblasts indeed resulted in a marked decrease in LIG3 protein levels (Fig. 2D and Supplementary Fig. 3C).

The p.R811Ter variant in Family 3 was not expressed and only the c.1826C>T allele (p.P609L variant) was present in the final transcript (Fig. 2F), suggesting that mRNA carrying the c.2431C>T variant (premature stop codon p.R811Ter) was degraded by nonsense-mediated decay, as shown for the premature stop-codon found in patients of Family 2. The p.P609L variant, mapping to the end of an alpha-helix in the ATP-binding domain, was predicted to destabilize the protein ($\Delta\Delta G$ = –0.74 kcal/mol; Fig. 2C).[30,31]

Compared to controls, fibroblasts from Patient 3-2 showed very low amounts of LIG3 protein (Fig. 2D and Supplementary Fig. 3D), in agreement with the transient transfection of the construct carrying the p.P609L variant in HEK293 cells (Supplementary Fig. 3F). We also investigated whether protein instability could be rescued by the proteasome inhibitor MG132 which, however, evoked only a marginal increase of LIG3 levels in mutant cells (Supplementary Fig. 3D).

LIG3 mutations result in impaired ligase activity and mtDNA depletion

As LIG3 is the only mtDNA ligase, we investigated whether the ligase activity of mitochondrial extracts from patient cells was affected. The mitochondrial extracts from control and Patient 2-1 fibroblasts were incubated with linearized plasmid DNA (Fig. 3A). Indeed, mutant extracts were unable to religate the linearized plasmid DNA, whereas mitochondrial extracts from control cells showed clear ligation activity. Subsequently, the mtDNA ligation and repair capacity was analysed in control and patient fibroblasts, according to Furda et al.[17] Control and mutant cells were exposed to oxidative stress (H_2O_2 for 15 min) and then allowed to repair the DNA damage for 6 h (Fig. 3B). Normal fibroblasts showed a reduced amount of PCR amplification product directly after the H_2O_2 exposure and a significant recovery after 6 h ($P < 0.05$; Fig. 3B). However, mutant cells showed little or no recovery of intact mtDNA, consistent with a severe defect in mtDNA repair (Fig. 3B).

We also investigated whether mtDNA replication was affected. The replication capacity of mtDNA after EtBr treatment was impaired in fibroblasts from Patient 3-2. Control and patient cells lost their mtDNA content at a similar rate and after 9 days of EtBr treatment all cells contained only low levels of mtDNA (controls between 0.5% and 1.2%; patient 0.9% of the starting level). However, the controls showed a clear increase in mtDNA content 6 days after withdrawal of EtBr, whereas this recovery was not observed in the patient samples. This experiment demonstrates that the patient cells have a clearly decreased capacity of mtDNA replication, in addition to the previously shown impairment in mtDNA repair (Fig. 3C).

We subsequently investigated whether the *LIG3* defects also resulted in reduced mtDNA content. Quantitative assessment of mtDNA copy number in a skeletal muscle biopsy sample from Patient 1-1 demonstrated significant mtDNA depletion compared with controls ($P < 0.01$; Fig. 3D). In Patient 2-2, mtDNA content of the skeletal muscle biopsy was 66% of that in age-matched healthy controls (Supplementary Fig. 4A).[16] A significant decrease in mtDNA copy number was also observed in myoblasts derived from Patient 3-2 compared to controls ($P < 0.05$; Fig. 3E).

Patient 1-1-derived fibroblasts showed significant mtDNA depletion compared to controls ($P < 0.01$; Fig. 3F). Conversely, in skin-derived fibroblasts of Patient 3-2 mtDNA copy number was similar to control cells (Supplementary Fig. 4B).

A quantitative assessment of mtDNA deletions was performed as reported in the Supplementary material, but deletions were not detected in both fibroblasts and skeletal muscle of Patient 1-1 (Supplementary Fig. 4C and D).

Mitochondrial dysfunction caused by LIG3 mutations

We next investigated how partial mtDNA depletion affected mitochondrial shape and function. Mitochondria were visualized with MitoTracker Green and analysed using confocal fluorescence microscopy. Overall, the mitochondria of mutant fibroblasts (Patient 1-1) displayed a fragmented network compared with control cells (Fig. 4A and Supplementary Fig. 5A–E). Compared to controls,
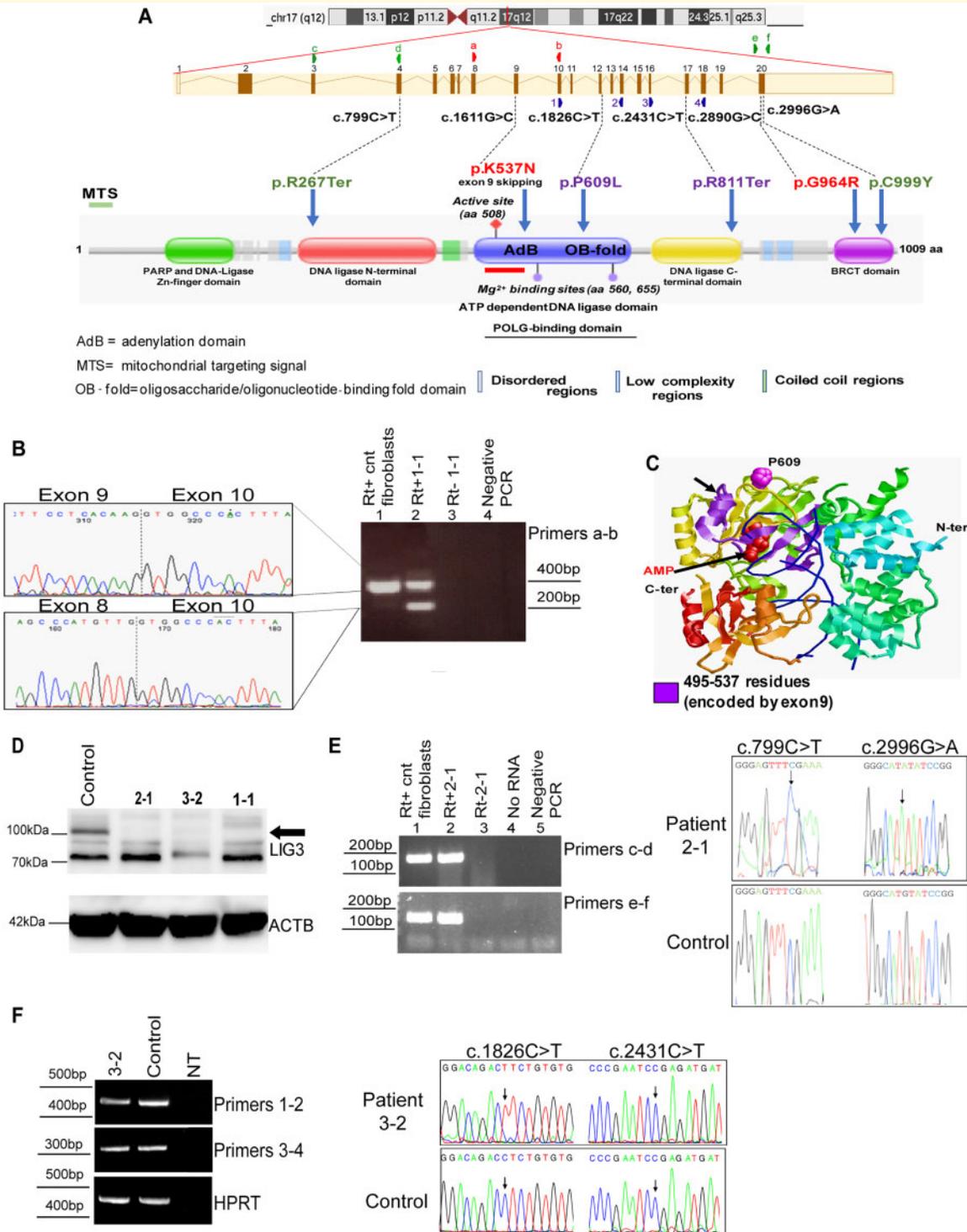
Figure 2 Identification of the molecular defects of *LIG3*. (**A**) Mutations mapped on the Pfam database (https://pfam.xfam.org/) domains of LIG3 [mitochondrial transport signal (MTS), ATP-binding domain, POLG and XRCC1-binding sites]. Variants in Family 1 are indicated in red, variants in Family 2 are indicated in green, and variants in Family 3 are indicated in violet. Arrowheads indicate the primer pairs used for subsequent RT-PCR analysis. aa = amino acids. (**B**) Splicing alteration induced by the p.K537N variant. Control fibroblasts (lane 1, cnt) show a band of the expected size for the transcript containing exon 9. Fibroblasts from Patient 1-1 (lane 2) show two bands: the higher band corresponds to the transcript with exon 9, and the lower band corresponds to a transcript without exon 9, as shown by the corresponding electropherograms. A cropped image of the gel electrophoresis is reported, without the marker lane (relevant sizes indicated; bp = base pairs). Lane 3: no RT = RT-PCR carried out without reverse transcriptase on Patient 1-1 mRNA; lane 4: negative control of RT-PCR (no template). (**C**) Structure of the ligase domain of human LIG3 (residues 257-833) in complex with DNA and AMP. The region encoded by exon 9 is in violet. The position p.P609 is represented in pink. (**D**) LIG3 western blot analysis in total cell lysates derived from human control and patient fibroblasts, showing a severely

(continued)

ultrastructural analysis in fibroblasts of Patient 1-1 showed qualitative and quantitative mitochondrial defects such as abnormal cristae and swollen mitochondria and a greater average mitochondrial area (0.011854 versus 0.00644 μm^2, respectively) (Supplementary Fig. 5F).

Analysis of the respiration rate confirmed the mitochondrial dysfunction in patient cells. We observed a significantly decreased uncoupled oxygen consumption (Fig. 4B) and lower ATP content (Fig. 4C) in Patient 1-1 fibroblasts. In addition, USCs which have high mitochondrial content,[32] were established from Patient 3-2 and tested for oxygen consumption rate. The results showed a decrease in oxygen consumption rate in USCs of Patient 3-2 compared with control USCs (Supplementary Fig. 5G). Although no differences in mitochondrial membrane potential were detected in intact control and Patient 1-1 fibroblasts (Supplementary Fig. 5H), in digitonin-permeabilized cells *LIG3*-mutant fibroblasts were less responsive to different respiratory substrates (glutamate malate, ADP, and succinate), inhibitors (oligomycin A and rotenone) and uncoupler (FCCP) compared to controls ($P = 0.0012$; Fig. 4D). MitoSOX[TM] Red staining (a selective probe for mitochondrial superoxide) detected a significantly increased production of mitochondrial reactive oxygen species in mutant cells compared to control cells ($P = 0.0393$; Fig. 4E), confirming the mitochondrial dysfunction.[33]

As the addition of glutamine (6 mM) has been shown to increase the survival of cells carrying mitochondrial defects, but not control cells,[34] the glutamine concentration was increased in the culture medium from the standard 2 mM to 6 mM. In control fibroblasts a higher glutamine concentration resulted in a decreased growth rate (Supplementary Table 3), in agreement with previous findings,[34-36] indicating that in the absence of mitochondrial defects, excess glutamine may exert a detrimental effect. Conversely, fibroblasts from Patient 1-1 grew more efficiently when exposed to 6 mM than 2 mM glutamine (Fig. 4F and Supplementary Table 3; $P = 0.0248$ at 96 h culture).

In vivo modelling of *LIG3* mutations

As a final line of evidence that the *LIG3* mutations can indeed cause the phenotype observed in the patients, we investigated the intestinal and neuronal phenotypes in a zebrafish model. We evaluated the consequences of *LIG3* mutations on the cerebellar structure[25,37] and the developing digestive system.[38] Reciprocal BLAST with the human LIG3 protein sequence against the zebrafish genome (Zfin v.10) identified a single *lig3* orthologue on chromosome 5, encoding a single transcript for which the encoded protein (NP_001025345) had 69.7% identity to human LIG3 (NP_039269).

We engineered F0 mutants using CRISPR/Cas9 genome editing (Supplementary Fig. 6A and B). F0 clutches with a *lig3* gene disruption (at 3 dpf) showed a significant decrease in the overall cerebellar area compared with both uninjected embryos and embryos injected with gRNA alone (Fig. 5A and B). The observed cerebellar phenotype was consistent with published data from *Lig3*[−/−] mice.[2]

We also applied a splice-blocking MO antisense oligonucleotide targeting the splice donor site of *lig3* exon 11 (e11i11; Supplementary Fig. 6A). The MO resulted in the skipping of exon 11 and the induction of a frameshift soon thereafter (Supplementary Fig. 6C). Injection of this MO resulted in a dose-dependent decrease in cerebellar area phenocopying the F0 CRISPR mutants ($P < 0.0001$; Supplementary Fig. 7A and B). The cerebellar phenotype was rescued by co-injecting human wild-type *LIG3* mRNA (Supplementary Fig. 7B), but not by mRNA harbouring patient-derived mutant variants (p.K537N and p.G964R) ($P < 0.0001$; Fig. 5C).

We also tested whether *lig3* suppression (MO) or ablation (CRISPR) perturbed the morphology of the gut and the function of the gastrointestinal tract. Quantification of HuC/D-positive neurons showed no significant differences between MO-treated or F0 mutant zebrafish compared with controls (Supplementary Fig. 7C and D). Anti-phosphomyosin staining in the zebrafish gut (8 dpf) did not show any myopathic changes (Supplementary Fig. 7E). We assayed gastrointestinal function at 8 dpf (at which age the gastrointestinal system is developed) by analysing the pattern of peristalsis using high-speed video microscopy. Qualitative scoring by investigators blind to the injection cocktail showed a significant number of morphants and knockout zebrafish with abnormal gut peristalsis [Fig. 5D,

Figure 2 Continued
decreased LIG3 protein levels (black arrow) in patient fibroblasts compared with controls. ACTB = beta-actin loading control. Immunoblotting for LIG3 and for ACTB (endogenous control) proteins were performed on the same blot. Cropped images are reported. (**E**) *LIG3* transcript analysis in Patient 2-1 and control cDNAs. *Left*: Agarose gel electrophoresis image of the RT-PCR products. Cropped images of gel electrophoresis are reported, without the marker lane (relevant sizes indicated). *Right*: Sanger sequencing analysis of the products of RT-PCR with primers c–d and primers e–f. Arrows indicate the c.799C > T and c.2996G > A variants in *LIG3* transcript. The electropherogram shows that only the c.2996G > A (p.C999Y) variant is present in the *LIG3* transcript of the patient, and not the c.799C > T (p.R267Ter) variant. (**F**) *LIG3* transcript analysis in Patient 3-2 and control cDNAs. *Left*: Agarose gel electrophoresis image of the RT-PCR products. HPRT = internal control. Cropped images of gel electrophoresis are reported, without the marker lane (relevant sizes indicated). *Right*: Sanger sequencing analysis of the products of RT-PCR with primers 1–2 and primers 3–4. Arrows indicate the c.1826C > T and c.2431C > T variants in *LIG3* transcript. The electropherogram shows that only the c.1826C > T (p.P609L) variant is present in the *LIG3* transcript of the patient, and not the c.2431C > T (p.R811Ter) variant.
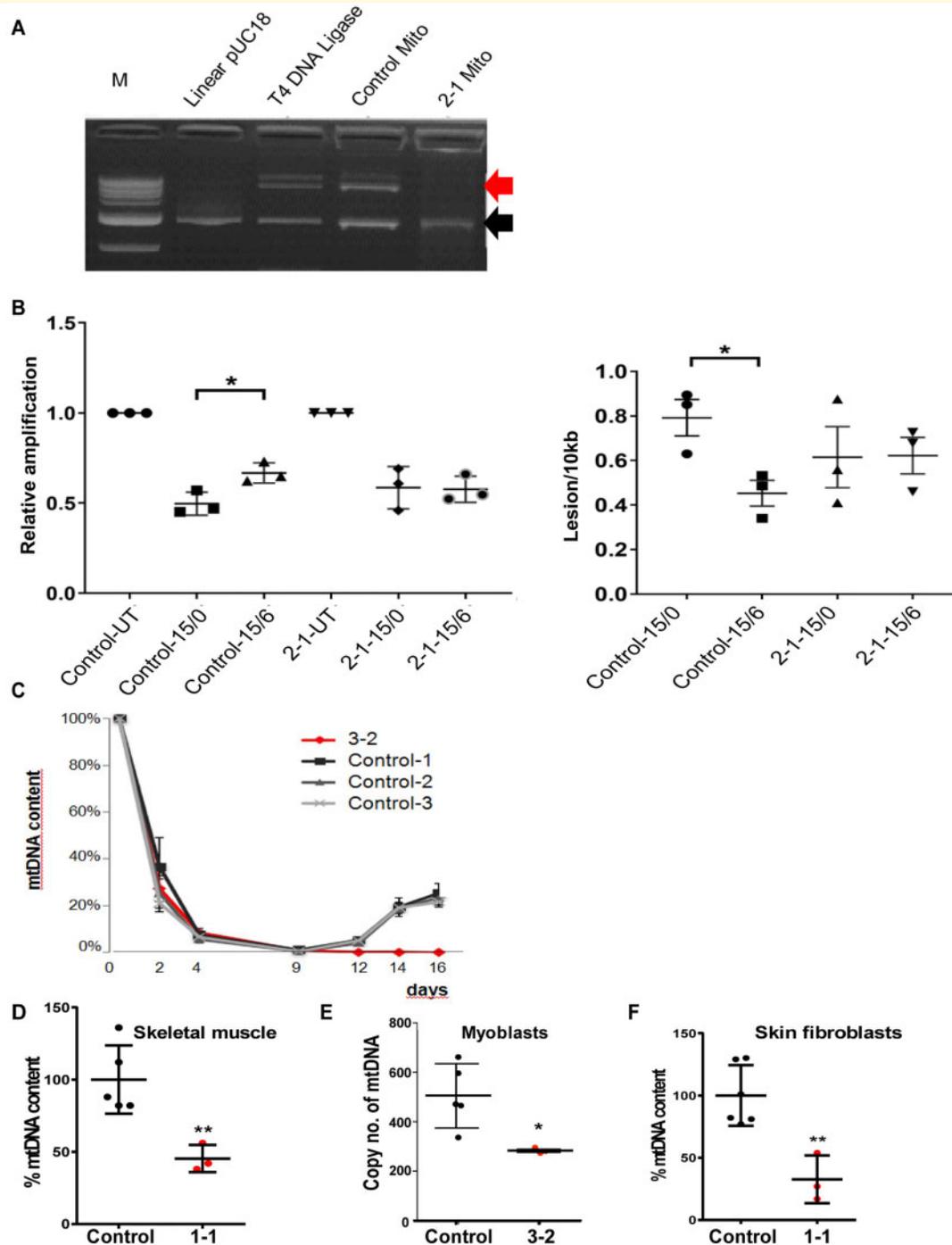
Figure 3 Defects in mtDNA maintenance due to *LIG3* mutations. (**A**) Comparison of ligase activity on linearized pUC18 plasmid DNA (length ~2.6 kb) in mitochondrial extracts between Patient 2-1 and control fibroblasts. T4 DNA ligase = positive control. Black arrow = linearized plasmid; red arrow = expected ligated product. M = molecular weight marker. (**B**) Reduced mtDNA repair kinetics after oxidative damage in control and Patient 2-1-derived fibroblasts. Values were normalized to mtDNA copy number ($n = 3$ experiments). UT = untreated; 15/0 = cells treated for 15 min with H_2O_2 and directly analysed for mtDNA repair kinetics; 15/6 = cells treated 15 min with H_2O_2, recovered for 6 h, then analysed for PCR amplification of a long mtDNA fragment relative to a short PCR fragment (*left*) and the calculated number of lesions per 10 kb of mitochondrial genome (*right*). *$P < 0.05$, ANOVA multiple testing. (**C**) Mitochondrial DNA depletion experiments. Control and Patient 3-2-derived fibroblasts were treated with EtBr for 9 days ($n = 3$ experiments; mean ± standard deviation, SD). (**D**) Quantification of mtDNA in the skeletal muscle of Patient 1-1 (red circles) and controls (black circles). Data shown for $n = 5$ controls and three replicates of the Patient 1-1 muscle biopsy. **$P < 0.01$, Student *t*-test. (**E**) Mitochondrial DNA quantification in myoblasts from Patient 3-2 (in red; three replicates of Patient 3-2 samples) and controls (in black; data shown for five controls). *$P < 0.05$, Student *t*-test. (**F**) Quantification of mtDNA in skin-derived fibroblasts ({atient 1-1); data shown for five samples for control (black circles) and Patient 1-1 (red circles). **$P < 0.01$, Student's *t*-test.

Supplementary Videos 1 and 2 (normal versus abnormal peristalsis) and Supplementary Fig. 7F).

Haematoxylin and eosin staining in the zebrafish gut (8 dpf) showed a reduction of goblet cells in morphants (e11i11) and F0 CRISPR mutants compared with controls (Supplementary Fig. 8A and B, $P < 0.01$ and $P < 0.05$, respectively). We then assayed swim bladder formation at 4 dpf (at which age the swim bladder is fully formed) by quantifying the swim bladder area, in morphants and F0 CRISPR mutants. Swim bladder area was reduced in morphants and F0 CRISPR mutants compared with controls (Supplementary Fig. 9A and B, $P < 0.0001$). Next, we quantified the expression of mitochondrial markers (*mt-nd1*, *mt-co1*) in zebrafish embryos at 4dpf. We observed a significant reduction for *mt-nd1* in morphants (e11i11) ($P < 0.01$; Supplementary Fig. 9C) with a similar trend in F0 CRISPR mutants, and a significant reduction for *mt-co1* in morphants (e11i11) ($P < 0.05$; Supplementary Fig. 9C) and F0 CRISPR mutants ($P < 0.05$; Supplementary Fig. 9C). We also tested whether *lig3* suppression (MO) or ablation (CRISPR) perturbed the morphology of the eye. Haematoxylin and eosin staining in the zebrafish eye (5 dpf), revealed a significant reduction of the outer nuclear layer in morphants (e11i11) and F0 CRISPR mutants, compared with controls (Supplementary Fig. 10A and B; $P < 0.0001$).

Discussion

In this study we described a novel recessive syndrome caused by mutations in the *LIG3* gene. The affected individuals in three independent families showed neurogastrointestinal encephalomyopathy characterized by CIPO, neurogenic bladder, myopathic changes, and neurological impairment with stroke-like episodes, epilepsy and leukoencephalopathy. The underlying cause is a defect in mtDNA ligase activity, leading to decreased mtDNA repair capacity, reduced mtDNA content and impaired mitochondrial function.

It was previously reported that the inactivation of *Lig3* led to loss of cell viability and early embryonic lethality in mice, with profound mitochondrial dysfunction due to reduced mtDNA in the nervous system.[2,5] At the single cell level, viability could be restored by mitochondrial targeting of another DNA ligase, showing that the pathogenic effect was caused by the inability to ligate mtDNA. Neuron-targeted conditional *Lig3* knockout mice showed a reduced brain size and cerebellar abnormalities with increased apoptosis in the granular layer, consistent with our observation of the zebrafish phenotype and brain defects as manifested by affected members of Family 3.[5] Gut peristalsis was not investigated in the *Lig3* mutant mice, so we link for the first time this phenotype to *LIG3* mutations. However, this is not unexpected as *POLG* mutations also cause compromised mtDNA maintenance and gut dysmotility.[39,40]

Given the functional redundancy of different ligases in the nucleus, our study suggests that primary mitochondrial dysfunction is the major, if not the only, contributor to the clinical phenotype, as documented by the downstream disruption of mtDNA maintenance and repair, with the depletion of mitochondrial genomes in skeletal muscle and fibroblasts from patients. This explains the impaired mitochondrial function in tissues with high-energy requirements, such as neurons and muscle cells. Disruption of *lig3* in the zebrafish recapitulated the cerebellar phenotype (a hallmark of *Lig3*$^{-/-}$ mice) and eye defects, as well as the severe impairment of gut propulsion that was observed in all patients of the three investigated families. Disruption of *lig3* also led to a significant decrease in the expression of mitochondrial markers in the zebrafish.

The features of our patients are reminiscent of known mitochondrial disorders, such as MNGIE, MELAS, and *POLG*-associated phenotypes, which all involve post-mitotic, high-energy dependent tissues, i.e. brain, skeletal muscle, and smooth muscle.[41] Notably, mtDNA depletion occurs in both MNGIE and *POLG*-mutant patients. The latter patients may additionally show migraine and epilepsy with stroke-like episodes, which are features also observed in MELAS as well as in our *LIG3*-mutant patients. Interestingly, multiple deletions in mtDNA, which are frequently observed in both MNGIE and *POLG*-associated phenotypes, were not found in patients with *LIG3* mutations. The overall CNS involvement in patients with *LIG3* mutations was more severe than in MNGIE patients with *TYMP* mutations, who usually have asymptomatic leukoencephalopathy.

The age of onset of major symptoms in patients with *LIG3* mutations was in infancy for Family 3, paediatric age for Family 1 and adult age for Family 2. The rate of progression of disease also seemed faster in Family 3 and slower in Family 2. We hypothesize that the difference in severity is caused by differences in residual protein levels and activity, but this has not been addressed directly. It is possible that the residual level of mitochondrial ligase function is higher in point mutants disrupting the interaction with XRCC1 (Families 1 and 2), which may lead to decreased protein levels but relatively normal function, whereas the mutations in Family 3 may lead to severely reduced protein levels and function. Further identification of patients with additional *LIG3* mutations will increase our understanding of the phenotypes of this new disorder, as well as provide a comparison with other mitochondrial diseases. Moreover, *LIG3* polymorphisms have been associated with a variety of pathological conditions such as tumours, neurodegenerative disorders (i.e. Alzheimer's disease), recurrent depression and neural tube defects,[42-47] suggesting that a better understanding of LIG3 function may have implication also in these diseases.

There is still a limited understanding of the metabolic consequences of mitochondrial defects.[48,49] Glutamine is crucially involved in metabolic fluxes, by replenishing the intermediates of mitochondrial respiration.[50,51] A recent
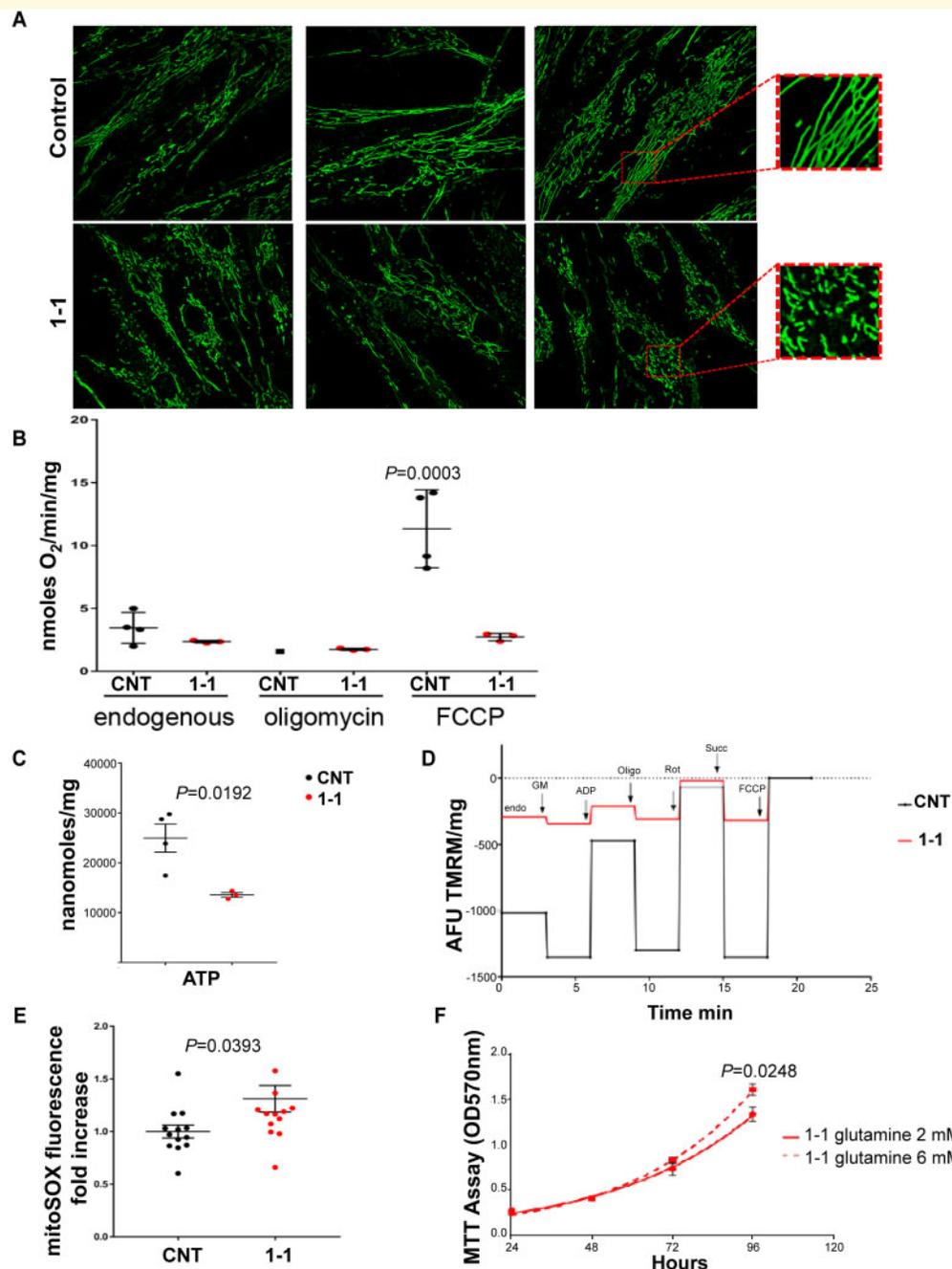
Figure 4 Functional impairment of mitochondria caused by *LIG3* mutations. (**A**) MitoTracker Green staining demonstrated a fragmented mitochondrial network in Patient 1-1 compared to control fibroblasts. Squares on the right are magnified images of the *insets*. (**B**) Mitochondrial oxygen consumption in endogenous and uncoupled conditions. Compared to control (CNT, black circles) cells, fibroblasts from Patient 1-1 (red circles) showed a significantly reduced respiration in the presence of FCCP ($P = 0.0003$; $n = 3$ independent experiments; ANOVA for multiple comparisons). (**C**) ATP content, normalized for protein content, was significantly decreased in patient fibroblasts (red circles) compared with control fibroblasts (black circles; $P = 0.0192$; $n = 3$ independent experiments; Student t-test). (**D**) Analysis of mitochondrial membrane potential in digitonin-permeabilized cells. A representative trace (spectrofluorimetric acquisition) of the pattern measured by the sequential addition of specific substrates (ADP = adenosine diphosphate; GM = glutamate-malate; Succ = succinate), inhibitors (Oligo = oligomycin A; Rot = rotenone) and an uncoupler (FCCP) as indicated. Compared to controls (black line), fibroblasts from Patient 1-1 (red line) were less responsive to different respiratory substrates and inhibitors ($P = 0.0012$; $n = 2$ independent experiments; Student t-test). AFU = arbitrary fluorescence units. (**E**) Analysis of mitochondrial reactive oxygen species production with MitoSOX shows a significant increase in Patient 1-1 fibroblasts (red circles) compared with control fibroblasts (black circles; $P = 0.0393$; Student t-test). (**F**) Growth curve (MTT assay) of Patient 1-1-derived fibroblasts grown in 2 mM (solid line) or 6 mM glutamine (dashed line). Patient cells cultured in 6 mM glutamine show a significantly increased growth after 96 h compared with cells grown in 2 mM glutamine ($P = 0.0248$; ANOVA test; five replicates/each time point).
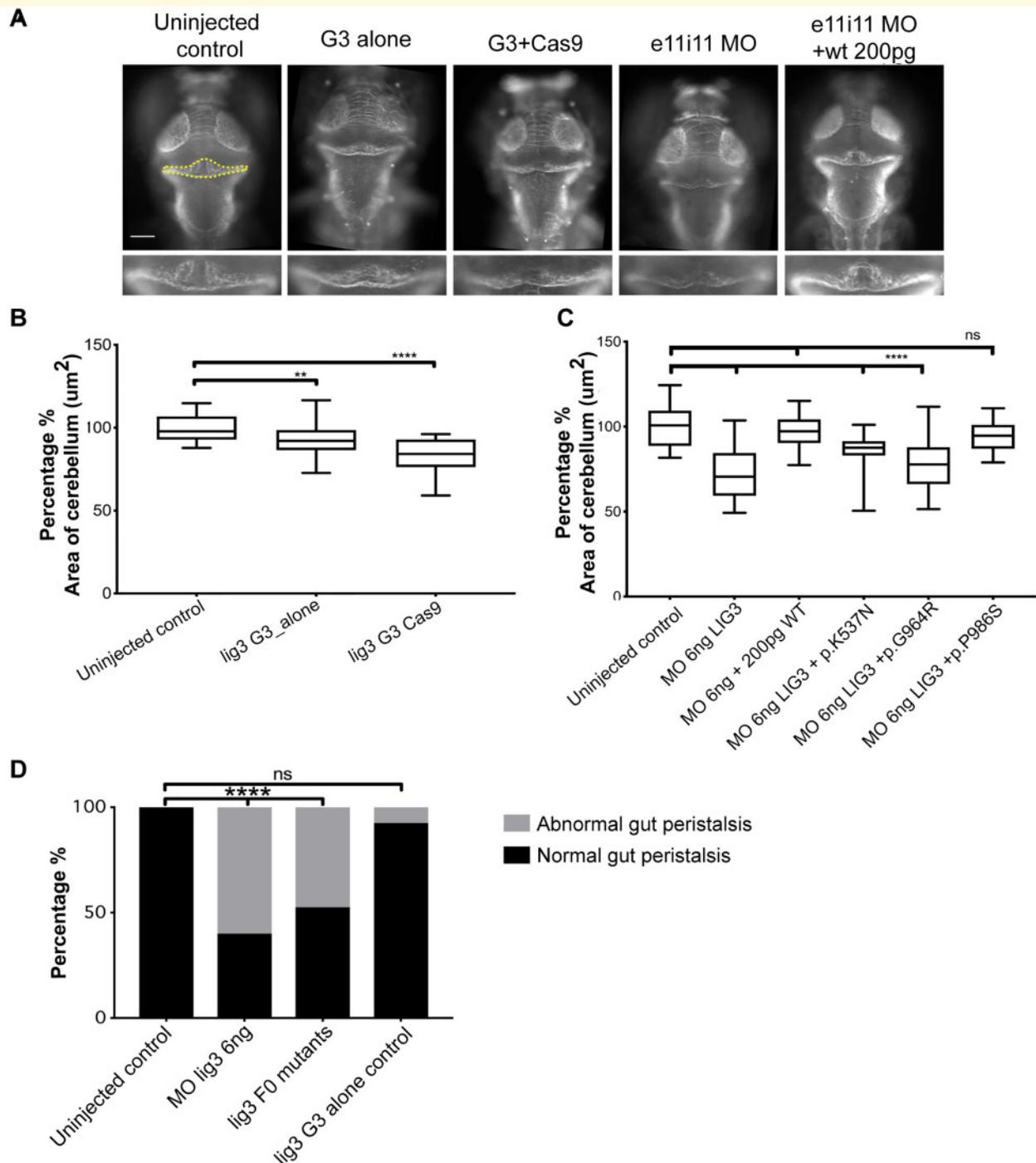
Figure 5 Suppression of *lig3* causes cerebellar hypoplasia and intestinal pseudo-obstruction in the zebrafish. (**A**) Representative dorsal images of larvae (3 dpf) immunostained for acetylated tubulin. To quantify neurons in the cerebellar area, whole-mount acetylated tubulin immunostaining was performed. Cerebellar structures of interest were measured using ImageJ software. Total cerebellar area was measured on acetylated tubulin-stained larvae by outlining structures with a fluorescent signal. Disruption of *lig3* causes cerebellar defects in both F0 CRISPR/Cas9-mutated (G3 + Cas9) zebrafish and morphants (e11i11 MO); e11i11 MO + 200 pg: complementation of *lig3* MO (e11i11 MO) with wild-type (WT) human *LIG3* mRNA (200 pg). Statistical analyses were performed using a nonparametric one-way ANOVA, with the Tukey multiple comparison test. ns = not significant; yellow dashed outline = area of the cerebellum that was measured; G3 = guide alone. (**B**) Quantification of cerebellar area in CRISPR/Cas9 genome edited larvae is shown for a guide targeting exon 8 of *lig3*. Comparison of F0 CRISPR/Cas9 clutches (*lig3* G3 Cas9) with either uninjected embryos or embryos injected with gRNA alone, but no enzyme (*lig3* G3_alone), showed a significant decrease in the overall cerebellar area in F0 mutants ($P < 0.0001$ for each comparison, replicated, scored blind). (**C**) Complementation of *lig3* MO with wild-type (WT) or mutated human mRNA. In contrast to embryos injected with wild-type *LIG3* mRNA or with a benign variant, i.e. not affecting ligase activity (p.P986S, rs498673), which were indistinguishable from the uninjected ones, injection with mRNA encoding p.K537N and p. G964R did not show rescue of the cerebellar phenotype. (**D**) Ablation of *lig3* causes abnormal gut peristalsis in F0 CRISPR/Cas9 mutants and morphants ($P < 0.0001$, n = 40 larvae/batch, replicated, χ^2 test). *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$.

study showed that increased glutamine concentration improved the survival of cells carrying mtDNA defects.[34] In our study, glutamine added to *LIG3*-mutant fibroblasts led to an increase in cell growth. These results suggest that the bioenergetic impairment induced by *LIG3* mutations may be ameliorated by boosting the glutamine anaplerotic pathway although the evidence is still limited.[34]

In conclusion, we described a new mitochondrial recessive disorder caused by biallelic variants in *LIG3*, with a clinical phenotype characterized by CIPO and neurologic abnormalities, including brain abnormalities similar to those observed in other mitochondrial diseases, epilepsy and stroke-like episodes, which are clearly associated with mitochondrial dysfunction caused by defective mtDNA maintenance.

Web resources

CHOPCHOP: https://chopchop.rc.fas.harvard.edu/
ClustalOmega: https://www.ebi.ac.uk/Tools/msa/clustalo/
The Single Nucleotide Polymorphism Database (dbSNP): https://www.ncbi.nlm.nih.gov/snp
Genome Aggregation database (gnomAD): http://gnomad.broadinstitute.org/
GraphPad Prism: https://www.graphpad.com/scientific-software/prism/
Genotype-Tissue Expression (GTEx): https://commonfund.nih.gov/gtex
Human Splicing Finder (HSF) v.3.1: http://www.umd.be/HSF3/index.html
The Human Protein Atlas: https://www.proteinatlas.org/
ImageJ (National Institutes of Health, Bethesda, MD): http://rsbweb.nih.gov/ij
Mendelian Clinically Applicable Pathogenicity Score (M-CAP): http://bejerano.stanford.edu/mcap/
Pfam database: https://pfam.xfam.org/
Prediction of functional effects of human nsSNPs (PolyPhen-2): http://genetics.bwh.harvard.edu/pph2/
Primer3: http://bioinfo.ut.ee/primer3/
Protein Data Bank (PDB): http://www.rcsb.org/
Protein Variation Effect Analyzer (Provean): http://provean.jcvi.org/index.php
Zebrafish Information Network (ZFIN): https://zfin.org/

Acknowledgements

Funding

Competing interests

The authors report no competing interests.

Supplementary material

Supplementary material is available at *Brain* online.

References

1. Cuneo MJ, Gabel SA, Krahn JM, Ricker MA, London RE. The structural basis for partitioning of the XRCC1/DNA ligase 3-α BRCT-mediated dimer complexes. *Nucleic Acids Res.* 2011;39:7816-7827.
2. Gao Y, Katyal S, Lee Y, et al. DNA ligase 3 is critical for mtDNA integrity but not Xrcc1-mediated nuclear DNA repair. *Nature.* 2011;471:240-244.
3. Odell ID, Barbour JE, Murphy DL, et al. Nucleosome disruption by DNA ligase 3-XRCC1 promotes efficient base excision repair. *Mol Cell Biol.* 2011;31:4623-4632.
4. Shokolenko IN, Fayzulin RZ, Katyal S, McKinnon PJ, Wilson GL, Alexeyev MF. Mitochondrial DNA ligase is dispensable for the viability of cultured cells but essential for mtDNA maintenance. *J Biol Chem.* 2013;288:26594-26605.
5. Simsek D, Furda A, Gao Y, et al. Crucial role for DNA ligase 3 in mitochondria but not in Xrcc1-dependent repair. *Nature.* 2011;471:245-258.
6. DiMauro S, Schon EA, Carelli V, Hirano M. The clinical maze of mitochondrial neurology. [Review]. *Nat Rev Neurol.* 2013;9:429-444.
7. Akbari M, Keijzers G, Maynard S, et al. Overexpression of DNA ligase 3 in mitochondria protects cells against oxidative stress and improves mitochondrial DNA base excision repair. *DNA Repair (Amst).* 2014;16:44-53.
8. Di Nardo G, Di Lorenzo C, Lauro A, et al. Chronic intestinal pseudo-obstruction in children and adults: diagnosis and therapeutic options [Review]. *Neurogastroenterol Motil.* 2017;e12945.
9. Downes TJ, Cheruvu MS, Karunaratne TB, De Giorgio R, Farmer AD. Pathophysiology, diagnosis, and management of chronic intestinal pseudo-obstruction [Review]. *J Clin Gastroenterol.* 2018;52:477-489.
10. Goldstein AM, Thapar N, Karunaratne TB, De Giorgio R. Clinical aspects of neurointestinal disease: pathophysiology, diagnosis, and treatment [Review]. *Dev Biol.* 2016;417:217-228.

11. Nishino I, Spinazzola A, Hirano M. Thymidine phosphorylase gene mutations in MNGIE, a human mitochondrial disorder. *Science.* 1999;283:689-692.

12. Carelli V, La Morgia C. Clinical syndromes associated with mtDNA mutations: where we stand after 30 years [Review]. *Essays Biochem.* 2018;62:235-254.

13. Lakshmipathy U, Campbell C. The human DNA ligase 3 gene encodes nuclear and mitochondrial proteins. *Mol Cell Biol.* 1999;19:3869-3876.

14. Tomkinson AE, Sallmyr A. Structure and function of the DNA ligases encoded by the mammalian LIG3 gene. *Gene.* 2013;531:150-157.

15. Giordano C, Iommarini L, Giordano L, et al. Efficient mitochondrial biogenesis drives incomplete penetrance in Leber's hereditary optic neuropathy. *Brain.* 2014;137:335-353.

16. Dimmock D, Tang LY, Schmitt ES, Wong LJ. Quantitative evaluation of the mitochondrial DNA depletion syndrome. *Clin Chem.* 2010;56:1119-1127.

17. Furda A, Santos JH, Meyer JN, Van Houten B. Quantitative PCR-based measurement of nuclear and mitochondrial DNA damage and repair in mammalian cells. *Methods Mol Biol.* 2014;1105:419-437.

18. Kornblum C, Nicholls TJ, Haack TB, et al. Loss-of-function mutations in MGME1 impair mtDNA replication and cause multisystemic mitochondrial disease. *Nat Genet.* 2013;45:214-219.

19. Lakshmipathy U, Campbell C. Double strand break rejoining by mammalian mitochondrial extracts. *Nucleic Acids Res.* 1999;27:1198-1204.

20. Diquigiovanni C, Bergamini C, Evangelisti C, et al. Mutant MYO1F alters the mitochondrial network and induces tumor proliferation in thyroid cancer. *Int J Cancer.* 2018;143:1706-1719.

21. Bergamini C, Moruzzi N, Sblendido A, Lenaz G, Fato R. A water soluble CoQ$_{10}$ formulation improves intracellular distribution and promotes mitochondrial respiration in cultured cells. *PLoS One.* 2012;7:e33712.

22. Niederriter AR, Davis EE, Golzio C, Oh EC, Tsai IC, Katsanis N. In vivo modeling of the morbid human genome using Danio rerio. *J Vis Exp.* 2013;24:e50338.

23. Marjoram L, Alvers A, Deerhake ME, et al. Epigenetic control of intestinal barrier function and inflammation in zebrafish. *Proc Natl Acad Sci USA.* 2015;112:2770-2775.

24. Sanna-Cherchi S, Khan K, Westland R, et al. Exome-wide Association Study Identifies GREB1L mutations in congenital kidney malformations. *Am J Hum Genet.* 2017;101:789-802.

25. Loviglio MN, Arbogast T, Jønch AE, et al. The Iimmune signaling adaptor LAT contributes to the neuroanatomical phenotype of 16p11-2 BP2-BP3 CNVs. *Am J Hum Genet.* 2017;101:564-577.

26. Ohkubo H, Iida H, Takahashi H, et al. An epidemiologic survey of chronic intestinal pseudo-obstruction and evaluation of the newly proposed diagnostic criteria. *Digestion.* 2012;86:12-19.

27. Canafoglia L, Franceschetti S, Antozzi C, et al. Epileptic phenotypes associated with mitochondrial disorders. *Neurology.* 2001;56:1340-1346.

28. El-Gebali S, Mistry J, Bateman A, et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* 2019;47:D427-D432.

29. Madeira F, Park YM, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 2019;47:W636-W641.

30. Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum Mutat.* 2009;30:1237-1244.

31. Savojardo C, Fariselli P, Martelli PL, Casadio R. INPS-MD: a web server to predict stability of protein variants from sequence and structure. *Bioinformatics.* 2016;32:2542-2544.

32. Lazzeri E, Ronconi E, Angelotti ML, et al. Human urine-derived renal progenitors for personalized modeling of genetic kidney disorders. *JASN.* 2015;26:1961-1974.

33. Diquigiovanni C, Bergamini C, Diaz R, et al. A novel mutation in SPART gene causes a severe neurodevelopmental delay due to mitochondrial dysfunction with complex I impairments and altered pyruvate metabolism. *FASEB J.* 2019;33:11284-11302.

34. Chen Q, Kirk K, Shurubor YI, et al. Rewiring of glutamine metabolism is a bioenergetic adaptation of human cells with mitochondrial DNA mutations. *Cell Metab.* 2018;27:1007-1025.

35. Pichili VB, Rao KV, Jayakumar AR, Norenberg MD. Inhibition of glutamine transport into mitochondria protects astrocytes from ammonia toxicity. *Glia.* 2007;55:801-809.

36. Rama Rao KV, Jayakumar AR, Norenberg MD. Induction of the mitochondrial permeability transition in cultured astrocytes by glutamine. *Neurochem Int.* 2003;43:517-523.

37. Guissart C, Latypova X, Rollier P, et al. Dual Molecular Effects of dominant RORA mutations cause two variants of syndromic intellectual disability with either autism or cerebellar ataxia. *Am J Hum Genet.* 2018;102:744-759.

38. Bonora E, Bianco F, Cordeddu L, et al. Mutations in RAD21 disrupt regulation of APOB in patients with chronic intestinal pseudo-obstruction. *Gastroenterology.* 2015;148:771-782.

39. Filosto M, Mancuso M, Nishigaki Y, et al. Clinical and genetic heterogeneity in progressive external ophthalmoplegia due to mutations in polymerase gamma. *Arch Neurol.* 2003;60:1279-1284.

40. Van Goethem G, Schwartz M, Löfgren A, Dermaut B, Van Broeckhoven C, Vissing J. Novel *POLG* mutations in progressive external ophthalmoplegia mimicking mitochondrial neurogastrointestinal encephalomyopathy. *Eur J Hum Genet.* 2003;11:547-549.

41. Kanungo S, Morton J, Neelakantan M, Ching K, Saeedian J, Goldstein A. Mitochondrial disorders. [Review]. *Ann Transl Med.* 2018;6:475.

42. Corral R, Lewinger JP, Joshi AD, et al. Genetic variation in the base excision repair pathway, environmental risk factors, and colorectal adenoma risk. *PLoS One.* 2013;8:e71211.

43. Czarny P, Kwiatkowski D, Toma M, et al. Impact of single nucleotide polymorphisms of base excision repair genes on DNA damage and efficiency of DNA repair in recurrent depression disorder. *Mol Neurobiol.* 2017;54:4150-4159.

44. Kwiatkowski D, Czarny P, Toma M, et al. Association between single-nucleotide polymorphisms of the *hOGG1, NEIL1, APEX1, FEN1, LIG1,* and *LIG3* genes and Alzheimer's disease risk. *Neuropsychobiology.* 2016;73:98-107.

45. Li D, Suzuki H, Liu B, et al. DNA repair gene polymorphisms and risk of pancreatic cancer. *Clin Cancer Res.* 2009;15:740-746.

46. Li G, Wang X, Wang X, et al. Polymorphism rs1052536 in base excision repair gene is a risk factor in a high-risk area of neural tube defects in China. *Med Sci Monit.* 2018;24:5015-5026.

47. Liao YH, Ren JT, Zhang W, et al. Polymorphisms in homologous recombination repair genes and the risk and survival of breast cancer. *Gene Med.* 2017;19:9-10.

48. Eales KL, Hollinshead KE, Tennant DA. Hypoxia and metabolic adaptation of cancer cells. [Review]. *Oncogenesis.* 2016;5:e190.

49. Smeitink JA, Zeviani M, Turnbull DM, Jacobs HT. Mitochondrial medicine: a metabolic perspective on the pathology of oxidative phosphorylation disorders [Review]. *Cell Metab.* 2006;3:9-13.

50. Daye D, Wellen KE. Metabolic reprogramming in cancer: unraveling the role of glutamine in tumorigenesis [Review]. *Semin Cell Dev Biol.* 2012;23:362-369.

51. Sullivan LB, Gui DY, Hosios AM, Bush LN, Freinkman E, Vander Heiden MG. Supporting aspartate biosynthesis is an essential function of respiration in proliferating cells. *Cell.* 2015;162:552-563.

ORIGINAL INVESTIGATION

The involvement of U-type dicentric chromosomes in the formation of terminal deletions with or without adjacent inverted duplications

Takema Kato[1] · Hidehito Inagaki[1] · Syunsuke Miyai[1] · Fumihiko Suzuki[1] · Yuki Naru[1] · Yasuko Shinkai[1] · Asuka Kato[1] · Kazuo Kanyama[1] · Seiji Mizuno[2] · Yukako Muramatsu[3] · Toshiyuki Yamamoto[4] · Mitsuhisa Shinya[5,6] · Yukiko Tazaki[5,6] · Sayuri Hiwatashi[5,6] · Toshiro Ikeda[5,6] · Mamoru Ozaki[7] · Hiroki Kurahashi[1,2]

Abstract

An inverted duplication with a terminal deletion (inv-dup-del) is one of the complex constitutional structural rearrangements that can occur in a chromosome. Although breakages of dicentric chromosome have been suggested, the precise mechanism of this is yet to be fully understood. In our present study, we investigated the genomic structure of 10 inv-dup-del cases to elucidate this mechanism. Two recurrent 8p inv-dup-del cases harbored a large copy-number-neutral region between the duplication and deletion in common. Although the other non-recurrent cases did not appear to have this copy-number-neutral region, refined sequencing analysis identified that they contained a small intervening region at the junction between the inverted and non-inverted segment. The size of this small intervening region ranged from 1741 to 3728 bp. Combined with a presence of microhomology at the junction, a resolution of the replication fork stalling through template switching within the same replication fork is suggested. We further observed two cases with mosaicism of the dicentric chromosome and various structural rearrangements related to the dicentric chromosome. Refined analysis allowed us to identify different breakpoints on the same chromosome in the same case, implicating multiple rounds of U-type formation and its breakage. From these results, we propose that a replication-based mechanism generates unstable dicentric chromosomes and that their breakage leads to the formation of inv-dup-dels and other related derivative chromosomes.

Introduction

An inverted duplication with a concomitant terminal deletion (inv-dup-del) is one of the common complex chromosomal rearrangements (CCRs) that arises in humans (Weckselblatt and Rudd 2015). Inv-dup-del rearrangements develop mainly in two successive steps. The first of these is the formation of a symmetric dicentric chromosome, which is unstable during mitosis because two functional centromeres can act bidirectionally. The formation of this unstable intermediate leads to anaphase bridging, followed by asymmetric breakage between the two centromeres in a subsequent mitotic division. Finally, it forms a stable large monocentric chromosome with an inverted duplication contiguous to a distal deletion and a small monocentric chromosome with a deletion (Zuffardi et al. 2009).

Two possible mechanisms are currently proposed for the formation of a dicentric chromosomal structure, the intermediate product of the inv-dup-del rearrangement (Zuffardi et al. 2009). One is a non-allelic homologous recombination (NAHR)-based mechanism. NAHR between segmental duplications with an inverted orientation produces a dicentric chromosome, and it induces breakage between the two centromeres during mitotic division. One of the representative NAHR-mediated inv-dup-dels is chromosome 8p inv-dup-del, which is mediated by homologous recombination between olfactory receptor-gene clusters (Shimokawa et al. 2004; Yu et al. 2010; García-Santiago et al. 2015). Since many 8p inv-dup-dels have been reported in unrelated families, those that are NAHR-mediated are thought to arise recurrently. In addition, NAHR between inverted repeats within the olfactory receptor-gene clusters induces benign polymorphic pericentric inversions. Meiotic recombination within the inverted segment in heterozygotes of polymorphic pericentric inversions produces a dicentric chromosomal structure that serves as an intermediate of the inv-dup-dels (Giglio et al. 2001).

✉ Hiroki Kurahashi
  kura@fujita-hu.ac.jp

Extended author information available on the last page of the article

Another proposed mechanism for the generation of dicentric chromosomal structures is a U-type exchange for the formation of a sporadic inv-dup-del (Ballif et al. 2003; Rowe et al. 2009; Yu and Graf 2010). This is based on a replication-based mechanism involving template-switching between sister chromatids. When a DNA replication fork encounters a replication block caused by different types of DNA damage, DNA synthesis may stall and a DNA end will then emerge. Break-induced replication then initiates at the site of microhomology on the opposite strand forming a U-type exchange between sister chromatids. Single stranded DNA nearby can be a potential template for strand switching and is usually present during lagging strand synthesis; it appears in the uncoupling of helicase and polymerase even during leading strand synthesis. Strand switching can facilitate a DNA synthesis restart, which continues to the next fork or as far as the telomere leading to the formation of dicentric chromosomes. This series of events leading to an inv-dup-del rearrangement is consistent with the concepts of replication fork stalling and template switching (FoSTeS) and microhomology-mediated break-induced replication (MMBIR) (Zhang et al. 2009a, b). This pathway can lead to a non-recurrent inv-dup-del rearrangement in a sequence-independent manner.

In our present study, we further elucidated the mechanism of dicentric chromosome formation, and the secondary structural rearrangements that consequently arise, by analyzing the breakpoint junction sequence of 10 cases of inv-dup-del representing 9 different chromosome regions. Breakpoint sequences found in common among the non-recurrent cases suggested the involvement of the replication-based mechanism in the formation of the dicentric chromosome. In addition, we obtained evidence of dynamic structural changes from the dicentric chromosome to the terminal deletion with or without adjacent inverted duplication during cell culture of samples from miscarriage fetuses. Taken together, we hypothesized from our present data that the replication stall model for the formation of a dicentric chromosome followed by a breakage-fusion-bridge (BFB) cycle is required for the generation of an inv-dup-del rearrangement.

Materials and methods

Ethical statement

This study was approved by the Ethical Review Board for Human Genome Studies at Fujita Health University. We obtained informed consent from all of the participating patients. All genetic experiments were carried out in accordance with the relevant guidelines and regulations.

Subjects

We analyzed 10 patients harboring an inv-dup-del that had been identified by initial standard chromosome banding and subsequent cytogenetic microarray. The karyotypes of the inv-dup-dels are summarized in Table 1. The breakpoints of the inv-dup-dels in 2 patients (cases 9 and 10) were almost identical and located at 8p, suggesting that they are recurrent. Two patients had a mosaic ring shaped chromosome (cases 3 and 8). Case 5 harbored an inv-dup-del with an additional triplication. Case 10 showed an inv-dup-del on 8p with an additional duplication of 8q. In addition, to better understand the development of inv-dup-del rearrangement from an intermediate dicentric chromosome, we analyzed 2 miscarriage fetuses (cases 11 and 12), whose karyotype showed mosaic isodicentric chromosomes.

Copy-number-analysis by cytogenetic microarray analyses

Genomic DNA samples were isolated from peripheral blood, cultured cells or chorionic villi using a Gentra Puregene Tissue Kit (Qiagen, Hilden, Germany). To identify the extent of the copy-number-alterations of the inv-dup-del, we employed a CytoScan HD, 750 k Array (Affymetrix, Santa Clara, CA) or Human Genome CGH Microarray (Agilent, Santa Clara, CA) in accordance with the manufacturer's instructions. Regions of copy-number-alteration were visualized using Chromosome Analysis Suite 3.2 (Affymetrix) or Genomics Workbench 7.0 (Agilent).

Copy-number-analysis by next generation sequencing (NGS)

In cases 11 and 12, we performed copy-number-analysis by NGS. The isolation and purification of genomic DNA from cultured cells or chorionic villi was carried out using a Gentra Puregene Tissue Kit (QIAGEN). Genomic DNA samples were diluted and amplified by whole-genome amplification using SurePlex (Illumina, San Diego, CA). One nanogram of each whole-genome amplified DNA sample was prepared for NGS analysis. Shallow whole-genome sequencing was performed for comprehensive copy-number-analysis using the VeriSeq PGS Kit MiSeq (Illumina). The sequencing data were analyzed using BlueFuse Multi Software (Illumina).

Breakpoint characterization by NGS

To identify the breakpoint junctions in our sporadic inv-dup-del study cases, we carried out NGS analysis of each patient's genomic DNA. Mate-pair whole-genome

Table 1 Patients information with inv-dup-dels

Sample no.	Chromosome	Karyotype	Genomic coordinates from microarray analysis	Copy-number-neutral region	Size (bp)	Microhomology (bp)
1	4p	–	arr[GRCh37] 4p16.3(71552_664410)X1,4p16.3p14(703668_36230853)X3	chr4:689,203–690,943	1741	3
2	4q	der(4)(pter→q3?4.3::q?34.3→31.3)	arr[GRCh37] 4q31.23q34.2(151058587_177382551)X3,4q34.2q35.2(177383079_190957473)X1	chr4:177,383,829–177,387,559	3728	2
3	9p	r(9)(::::::p24→q34.3::)[27]/45,XY,−9[3]	arr[GRCh37] 9p24.3p24.1(203861_8216831)X1,9p24.1p21.1(8216910_31620128)X3	chr9:8,215,737–8,218,461	2725	0
4	9p	add(9)(p24)	arr[GRCh37] 9p24.3p23(214367_11573590)X1,9p23p13.1(12048553_39156954)X3	chr9:11,712,793–11,715,125	2333	−4
5	10q	add(10)(q25.1)	arr[GRCh37] 10q26.12q26.3(122328262_132082817)X3~4,10q26.3(132942513_13540 4523)X1	chr10:132,878,450–132,880,484, chr10:122,306,763–122,307,064	2035, 303	2
6	11q	–	arr[GRCh37] 11q23.1q24.1(111612172_121578675)X3,11q24.1q25(121829034_134868407)X1	chr11:121,765,738–121,767,908	2167	3
7	18p	add(18)(p11.2)	arr[GRCh37] 18p11.32p11.31(14,316–3,458,388)X1,18p11.31p11.21(3,572,316–14,733,870)X3	chr18:3,492,546–3,494,355	1810	3
8	21q	r(21)[27]/45,XX,−21[2]/46,XX,?dic r(21)[1]	arr[GRCh37] 21q21.1q22.13(16832706_44518076)X3,21q22.3(44526182_48097372)X1	chr21:44,524,195–44,526,196	2001	0
9	8p	der(8)(:p12→p23.1::p23.1→qter)	arr[GRCh37] 8p23.3p23.1(158048_6981988)X1,8p23.1p11.22(11945855_39451389)X3			
10	8p	der(8)(8qter→8q24.2::8p11.2→8p23.1::8p23.1→8qter)	arr[GRCh37] 8p23.3p23.1(158048_6999114)X18p23.1p11.22(12560781_39078328)X3,8q24.21q24.3(129285242_146295771)X3			

sequencing was performed to determine the breakpoints of the inverted duplication in the dicentric chromosome. Since the breakpoint junctions can have palindromic (inverted repeated) characteristic sequences, we generated mate-pair long libraries of 9 kb to minimize the loss of palindromic DNA molecules at the breakpoint junction during library preparation. These 9-kb mate-pair libraries were prepared using a Nextera Mate Pair Library Preparation Kit (Illumina) in accordance with the manufacturer's protocol. The libraries were sequenced with 101 bp-paired-end reads on a HiSeq 1500 platform (Illumina). Sequence reads were trimmed adapter sequences using NxTrim and mapped with BWA 0.7.10 against hg19 (Li and Durbin 2010). Subsequently, discordantly mapped paired-reads were extracted to detect chromosomal structural rearrangements using BreakDancer (Fan et al. 2014). Putative breakpoints produced by inverted duplications, triplications and terminal deletions were confirmed by visual inspection of NGS data using Integrative Genomics Viewer (IGV) (Thorvaldsdottir et al. 2013). The nucleotide sequences of the breakpoint junctions of the inverted duplications and triplications were determined by breakpoint-specific PCR and Sanger sequencing using an ABI3130xl sequencer (Life Technologies, Foster City, CA). The length of the inverted sequence homology at the between copy-number-gain and -loss was searched using a YASS sequence similarity search (Noe and Kucherov 2005). Default parameter setting was used for this analysis.

Product of conception karyotyping

Products of conception (POC) were analyzed for chromosome abnormalities using standard cell culture followed by Giemsa banding.

Results

Copy-number variation features in recurrent and non-recurrent inv-dup-dels

We employed a cytogenetic microarray to characterize the 10 inv-dup-del rearrangements in our study cohort. The copy-number-profile detected by SNP microarray showed a clear difference between the cases involving 8p23.1 (cases 9 and 10) and others (Fig. 1). The two 8p23.1 cases harbored a large copy-number-neutral region between the copy-number-gain and -loss regions (~ 5 Mb). The endpoints of these copy-number-neutral regions were found to be almost identical between the two subjects and located within the olfactory receptor (OR) gene clusters. In contrast, none of the eight sporadic inv-dup-del cases showed
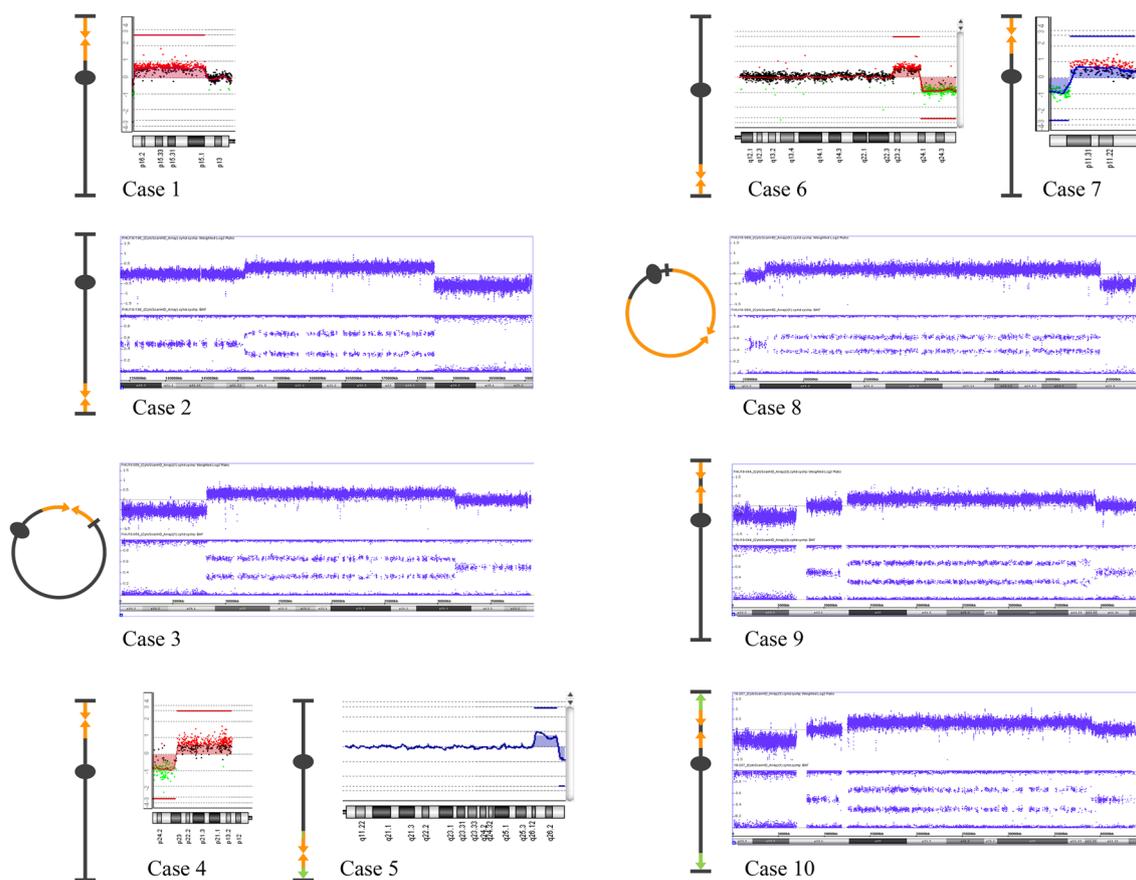
Fig. 1 Copy-number-profile of the inv-dup-dels determined by microarray. Copy-numbers were determined using an Affymetrix CytoScan microarray and visualized with ChAS software (Case 2, 3, 8, 9, 10). The upper blue plots show the signal for the weighted log 2 ratio. The lower plots indicate the B allele frequency. Copy-numbers were also determined using an Agilent CGH microarray and visual-ized with Genomics Workbench (Case 1, 4, 5, 6, 7). The log 2 ratio of the genomic copy-number is also plotted and indicated by horizontal lines. Illustrative diagrams of chromosomal structure are shown at the left. Inverted duplications are shown by arrows with orange color. Other rearrangements are shown with green color

any detectable copy-number-neutral regions between the copy-number-gain and -loss (cases 1–8; Fig. 1).

Breakpoint analysis of the non-recurrent inv-dup-dels

Whole genome sequencing of the 9-kb mate-pair library was performed to further analyze the breakpoints in our eight study cases with non-recurrent inv-dup-del structures. The inverted duplication structures were confirmed in each case by the presence of a junction for the inverted orientation of the two copy-number-gain regions. The DNA fragment incorporating the junction manifested a characteristic volcano-like pattern in the mapping of the discordant reads (Fig. 2). Some cases had telomeric repeat sequences at the proximal end of the inverted duplication, i.e. mapped at the proximal end of the copy-number-gain regions, reflecting telomere healing of the breakage (data not shown). In the case of ring chromosomes, the proximal end of the inverted duplication possibly contained a 9q subtelomeric repeat sequence (case 3) or 21q repetitive sequence (case 8), suggesting that ring formation was another healing pathway for the breakage (data not shown). Based on the fact that inverted and non-inverted segment was fused in head-to-head orientation and the chromosome end-like sequence appeared at the other end of the inverted segment, the inv-dup-del structures were confirmed in these cases.

Although high-resolution microarray did not detect an intervening copy-number-neutral region between the inverted and non-inverted segment, refined analysis of the breakpoint junction sequences identified a small copy-number-neutral region at the junction in all of the cases with a non-recurrent inv-dup-del (Fig. 2, red arrows). The size of the copy-number-neutral regions ranged from 1741 to 3728 bp (Fig. 3, Table 1). Microhomologies from 1 to 3 bp were found at breakpoint junctions between inverted and
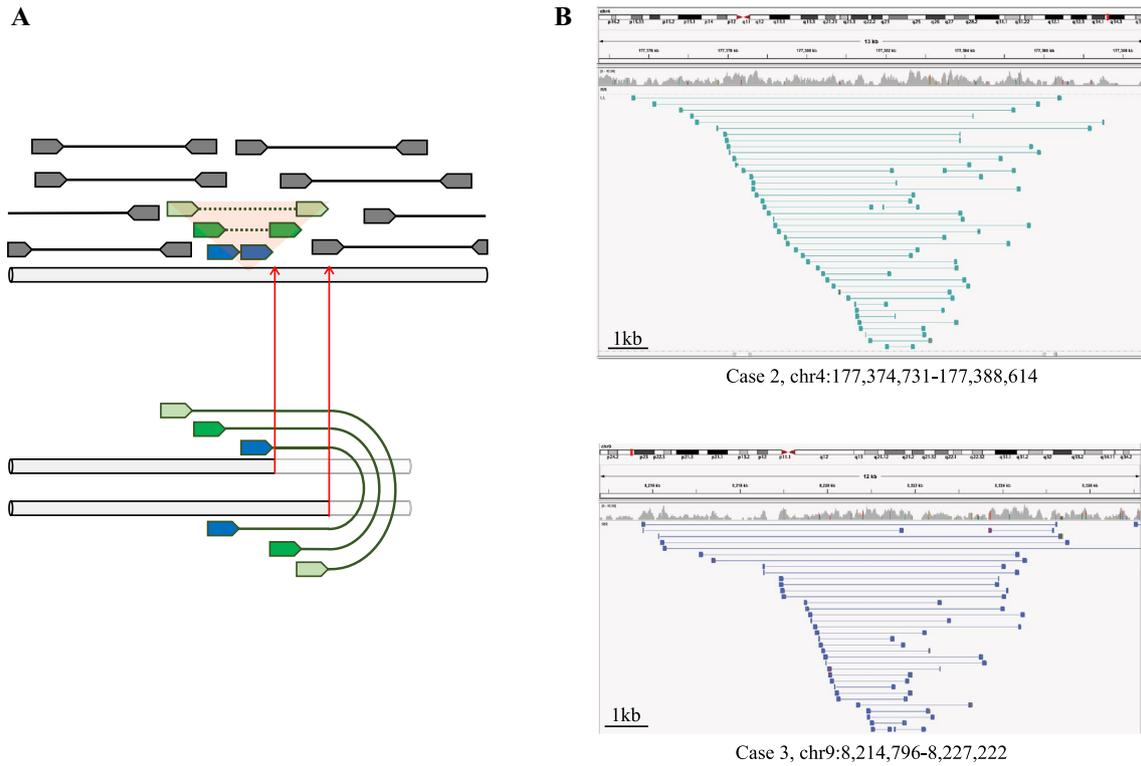
A

B

Case 2, chr4:177,374,731-177,388,614

Case 3, chr9:8,214,796-8,227,222

Fig. 2 Discordant reads at the junction of the dicentric chromosome. **a** Strategy for determining mate-pair sequences. The upper panel provides a schematic of the reference human genome structure. The lower panel depicts the assumed structure of the inverted duplication. Gray reads show normal paired-end reads, which were mapped in the forward and reverse direction to the reference genome. Green colored reads indicate discordant reads, which were mapped in the same orientation. Red arrows indicate the putative breakpoints. **b** Mapping of mate-pair sequences on the reference human genome sequence. The region around the breakpoint of dicentric chromosome was shown using IGV browser. A characteristic volcano-like pattern could be observed

Fig. 3 Distribution of the lengths of the intervening copy-number-neutral regions. Bars indicate the number of analyzed cases in studies published from 2003 to 2014 (black bar) (Ballif et al. 2003; Bonaglia et al. 2008; Hermetz et al. 2014; Rowe et al. 2009) and in this current study (hatched bar)
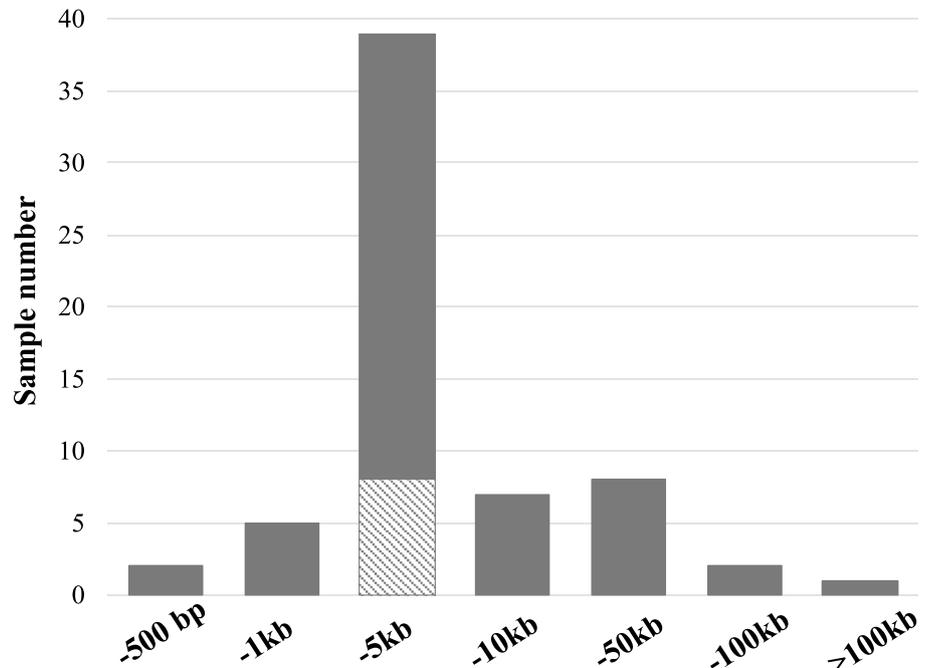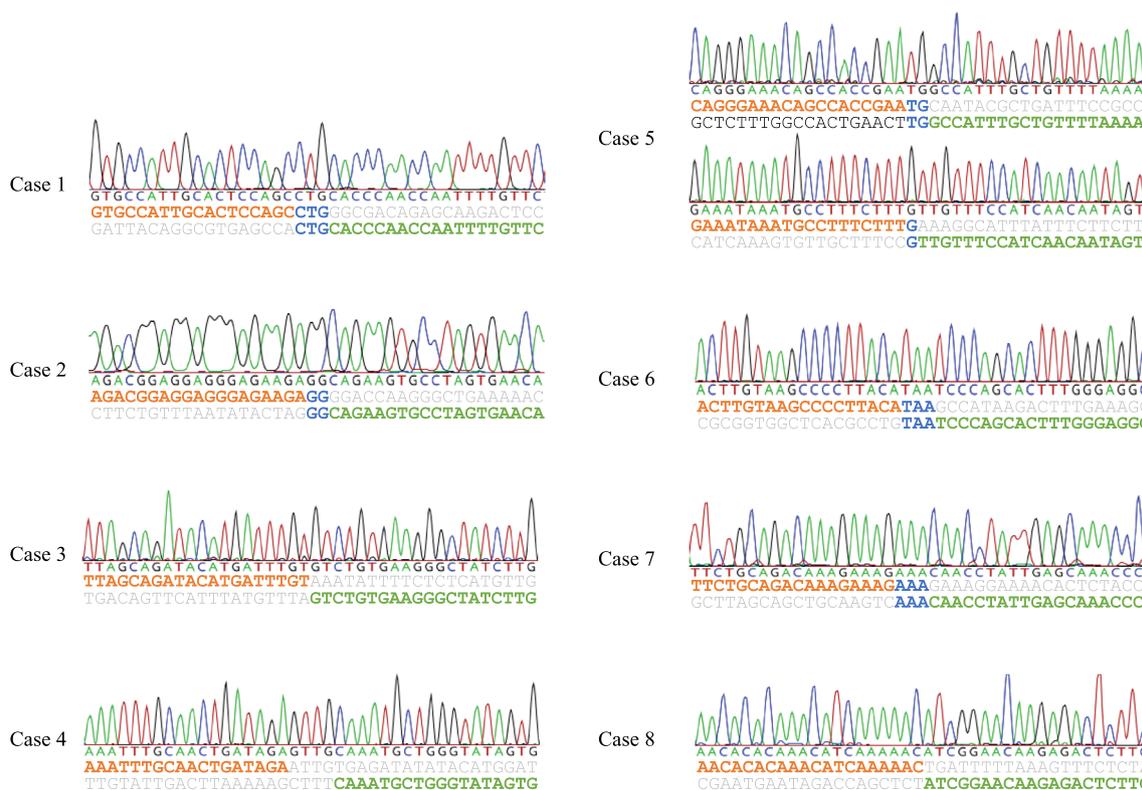
Fig. 4 Junction sequences of fold-back chromosomes. The orange and green characters indicate the plus and minus strands, respectively. Blue characters indicate homologous nucleotides. Gray characters indicate unknown nucleotide insertions

non-inverted segment in 5 cases, but none were evident in 3 cases with the remaining subject carrying a 4 bp microinsertion (Fig. 4, Table 1).

To investigate whether the initial formation of the dicentric chromosome in our study subjects was caused by an inverted repeat sequence at the breakpoint, we conducted sequence similarity analysis around the breakpoint region. Sequences with high similarity were found as expected in the inverted orientation at the proximal and distal end of the copy-neutral-region in the two recurrent 8p inv-dup-del cases (case 9 and 10), suggesting that the breakpoints were located within the olfactory receptor-gene clusters (Fig. 5). In contrast, significant inverted repeat sequences were not observed in any of the other inv-dup-del cases (cases 1–8). It is likely that the inverted duplications were generated in these eight individuals through repeat-independent, microhomology-based mechanisms, such as FoSTeS/MMBIR or microhomology-mediated end joining (MMEJ).

Breakage of the dicentric chromosome to form an inv-dup-del rearrangement

We additionally analyzed two fetal loss cases for which molecular cytogenetic analysis revealed evidence of

dicentric chromosomes leading to inv-dup-del formation. Case 11 underwent a miscarriage after 8 gestational weeks. Karyotyping of her POC sample revealed that all clones harbored add(13)(q22). Furthermore, derivative chromosome 13 showed a mirror structure. NGS chromosome analysis further revealed a copy-number-gain adjacent to a deletion on chromosome 13 (Fig. 6a). Based on these results, we concluded that the derivative chromosome was an isodicentric chromosome 13 (idic(13)). When we used uncultured POC cells for NGS copy-number-analysis, we identified an inv-dup-del pattern on the chromosome 13 which was likely to have derived from a breakage of idic(13). Intriguingly, the endpoint of the deletion of the two samples did not show concordance. If the inv-dup-del was a breakage product of idic(13), the deletion endpoint should be concordant, but the position of the deletion endpoint in inv-dup-del was more proximal (see "Discussion").

Case 12 (Fig. 6b) underwent a stillbirth at gestational week 34 due to multiple fetal anomalies. We analyzed three independent cultures of her POC sample. Karyotyping from one of the cultures revealed a 46,XX,del(15)(q26.1) [46]/46,XX,psu idic(15)(q26.3) mosaicism [8]. To analyze for a possible breakage of the idic(15) chromosome, we performed an SNP array of the same sample. We observed
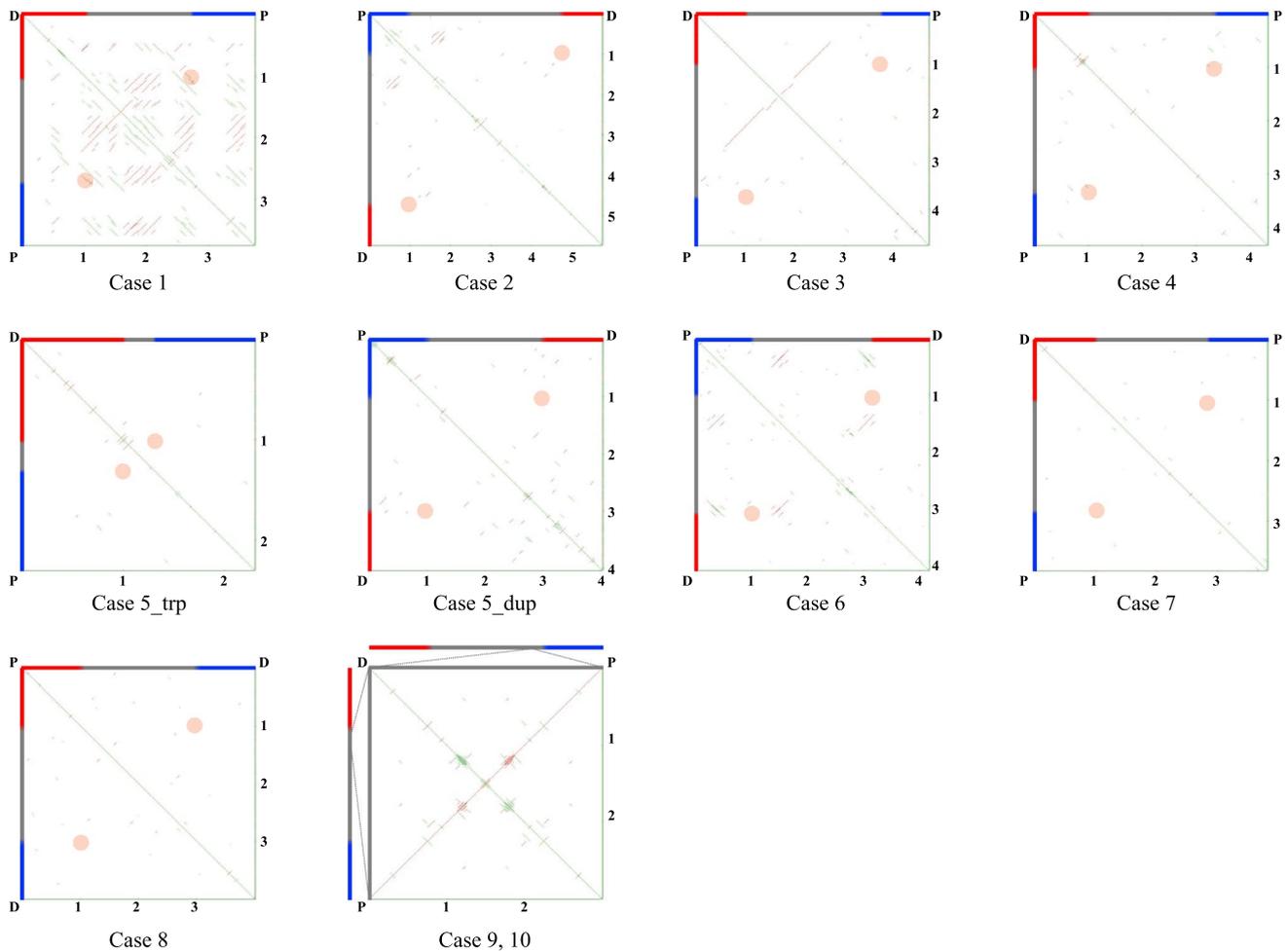
Case 1 Case 2 Case 3 Case 4

Case 5_trp Case 5_dup Case 6 Case 7

Case 8 Case 9, 10

Fig. 5 Sequence similarity analysis of the small intervening regions in our case subjects. Intervening regions which covered ±1 kb around the copy-number-neutral region are schematically represented on both axes. Black bars indicate copy-number-neutral regions, while blue and red bars indicate copy-number-gain and -loss, respectively. We defined the default parameter values for this analysis. Green line indicates a direct repeat and pink line denotes an inverted repeat. Pink circles indicate sequence similarities between the proximal and distal breakpoints. With regard to cases 9 and 10, parts of the two olfactory receptor-gene cluster regions (REPD[hg19]chr8:7,466,506-7,468,005 and REPP[hg19]chr8:12,466,006-12,467,505) were merged and used for this analysis. P, proximal side; D, distal side. The scale represents size in 1 kb

a stepwise copy-number-loss on chromosome 15q. It was likely that the distal breakpoint corresponded to the idic(15) breakpoint whilst the proximal breakpoint corresponded to a terminal deletion endpoint or inv-dup-del endpoint. We performed karyotype analysis of another culture which revealed 46,XX,add(15)(q26.1)[24]/46,XX,del(15)(q26.1) [19]/46,XX,psu idic(15)(q26.3)[14]. As was the case with the first culture, an SNP array indicated a stepwise copy-number-loss on chromosome 15q. The distal breakpoint was found to be identical to that of the first culture, whilst the position of the proximal breakpoint differed (see "Discussion"). In addition, a copy-number-gain of chromosome 17q was observed, suggesting that the add(15)(q26.1) was der(15)t(15;17)(q26;q21). We also performed SNP array analyses of a frozen uncultured sample for case 12, which

indicated a copy-number-gain on chromosome 9q in addition to a stepwise copy-number-loss of chromosome 15q, which might be derived from a line with 46,XX,der(15)t(9;15) (q34;q26). In summary, the unstable idic(15) chromosome may be susceptible to DNA breakage, leading not only to inv-dup-del rearrangement, but also to a terminal deletion such as del(15)(q26.1), or other unbalanced translocations like der(15)t(15;17)(q26;q21) or der(15)t(9;15)(q34;q26).

Discussion

We conducted our current molecular cytogenetic study to clarify how an inv-dup-del rearrangement is generated. We carried out the breakpoint junction analysis of dicentric
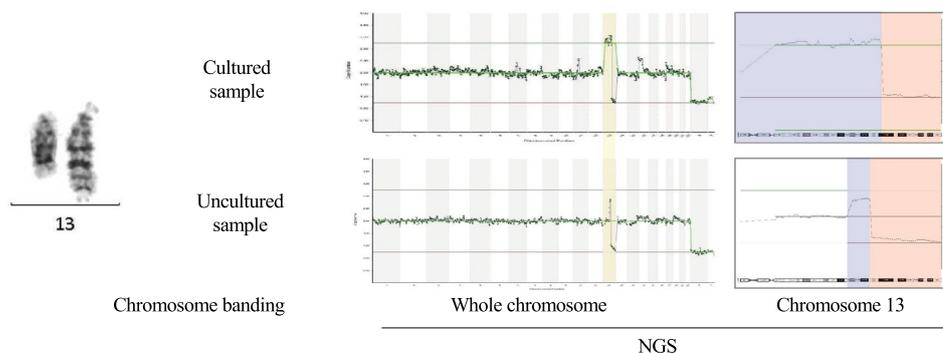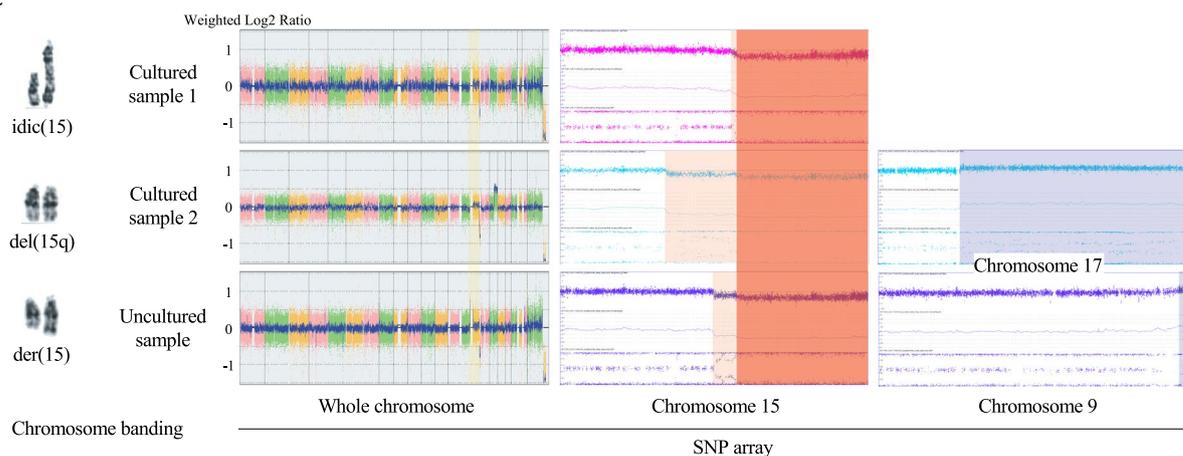
Fig. 6 Chromosome conformation changes during cell culture. Various chromosome analyses of cases 11 and 12 indicating that dicentric chromosomes in both subjects led to secondary structural rearrangements. **a** Case 11 harbors a chromosomal abnormality on chromosome 13. The upper panel shows a cultured sample and the lower panel shows an uncultured sample. Chromosome 13 was painted with a yellow signal. Copy-number-alterations are indicated in blue (copy-number-gain) and red (copy-number-loss). **b** Case 12 harbors chromosomal abnormalities on chromosomes 9, 15 and 17. The upper and middle panels show cultures of sample 1 and 2, respectively. The lower panel shows an uncultured sample. Chromosome 15 was painted with a yellow signal. Copy-number-alterations are indicated in blue (copy-number-gain) or red (copy-number-loss). In cultured sample 2, chromosome banding revealed three different karyotypes

chromosomes and a follow-up survey of secondary rearrangements that arose due to these dicentric chromosomes.

Involvement of a replication-based mechanism in the formation of a U-type dicentric chromosome

SNP microarray data showed a significant difference between recurrent and non-recurrent inv-dup-dels. Recurrent inv-dup-dels harbor large copy-number neutral region between the copy-number-gain and -loss. In contrast, non-recurrent sporadic inv-dup-dels do not have copy-number neutral region. This implicates mechanism leading to the formation of a dicentric chromosome might be different. The endpoints of copy-number-neutral regions were located within the OR gene clusters. This suggested that the repetitive nature of OR gene clusters was involved in the development of the recurrent inv-dup-dels in cases 9 and 10. NAHR between the OR genes may possibly underlie the most common recurrent inv-dup-del (Giglio et al. 2001; García-Santiago et al. 2015). Although we used high-resolution microarray in our current study, any sporadic inv-dup-del cases were not able to detect copy-number-neutral regions between the copy-number-gain and -loss. It was previously proposed that a symmetrical U-type exchange is the underlying mechanism leading to non-recurrent dicentric chromosome formation (Rowe et al. 2009). In our current study however, when we analyzed junction sequences at a nucleotide resolution, asymmetrical dicentric structures with an intervening sequence were observed. The presence of microhomologies or microinsertions at the fusion points of inverted and non-inverted segments (Fig. 4) indicated that the inverted duplications were generated through repeat-independent, microhomology-based mechanisms, such as FoSTeS/MMBIR or microhomology-mediated end joining.

In a previous report, Hermetz et al., proposed a fold-back model for the formation of dicentric chromosomes via a double-strand break (DSB) repair pathway (Hermetz et al. 2014). In this model, a DSB undergoes resection, which generates a long 3′ single-stranded DNA (ssDNA) overhang. This in turn promotes intrastrand pairing to the short inverted repeat via a homology-based mechanism leading to the formation of a dicentric chromosome. Notably however, we did not observe any homology between both ends of the intervening copy-number-neutral region at the boundary of the inverted and non-inverted segments in our current cases (Fig. 5). Further, the size of the intervening sequence showed a unique distribution, ranging from 1 to 5 kb in our analyzed cases. When we increased the number of inv-dup-del cases using data from previous studies in which where the breakpoints of inv-dup cases were analyzed at a nucleotide resolution (Ballif et al. 2003; Bonaglia et al. 2008; Rowe et al. 2009; Hermetz et al. 2014), the size of the intervening sequence peaked at 1–5 kb (Fig. 3).

Uncoupling of the helicase and polymerase leads to a single-stranded DNA in the leading strand, which when stalled can produce an uncoupled nascent lagging strand under certain conditions. Such single-stranded DNA can be up to 3–5 kb in length, depending on the nature of the replication fork block (Lopes et al. 2006). It is likely that the tight distribution of these sizes may reflect a resolution of the replication fork block through template switching within the same replication fork. Interestingly, when we increased our current sample number by including prior data from the literature, another low-pitched distribution curve appeared up to 100 kb (Fig. 3). This possibly reflects a resolution of the stalled replication fork by an adjacent fork which may be distant in a genomic sense but in close proximity within the nucleus. We are thus proposing a replication stall model as a mechanism of formation for U-type dicentric chromosomes.

Dicentric chromosome instability results in secondary rearrangements

In general, dicentric chromosome breakages would produce two reciprocal products: inv-dup-del chromosomes and terminal deletion chromosomes. We observed in our present series that two POC samples had a mosaicism of cells with dicentric chromosomes and inv-dup-del or other rearrangements associated with the dicentric chromosome. This suggested that the formation of the dicentric chromosome and other related rearrangements including the inv-dup-del were sequential events. The breakage of an unstable dicentric chromosome leads to the formation of not only an inv-dup-del chromosome and terminal deletion chromosome after stabilization through telomere healing of the broken ends, but also results in various structural rearrangements via stabilization by telomere capture.

Interestingly, our SNP microarray analysis of case 11 indicates that the endpoint of the deletion of the possible inv-dup-del was more proximal to that of the idic(13) chromosome. This suggested that the inv-dup-del chromosome might have originated from a terminal deletion chromosome reciprocal of the inv-dup-del chromosome derived from a breakage of idic(13). The terminal deletion chromosome might form a U-type dicentric chromosome leading to the second inv-dup-del chromosome having the proximal deletion endpoint. Likewise, SNP microarray data for case 12 indicated a more proximal endpoint of der(15) to that of the idic(15). This also suggested that the der(15) chromosome might have been generated in the second round of breakage of the dicentric chromosome. Dicentric chromosome must be stable and segregate to daughter cells correctly since one of the centromeres is functionally silenced (Stimpson et al. 2010). However, just after the formation of a U-type dicentric chromosome, both of the centromeres might be still active and the dicentric chromosomes are therefore unstable, leading to secondary rearrangements into various derivative chromosomes (Soler et al. 2003; Chabchoub et al. 2007; Schlade-Bartusiak et al. 2013; Pedurupillay et al. 2014). Hence, we can observe multiple clones with different derivative chromosomes.

Finally, in case 12, two types of translocation chromosomes related to the idic(15) chromosome were observed. In general, an unbalanced translocation would originate from segregation of the one of the translocation chromosomes from a balanced translocation carrier. However, our current results suggest that if the unbalanced translocation arises as de novo, it might originate from breakage of dicentric chromosome followed by telomere capture using the other chromosomal end.

In summary, a dicentric chromosome is so unstable that it leads to the generation of an inv-dup-del chromosome, as well as a reciprocal terminal deletion through telomere healing and also an unbalanced translocation via telomere capture.

the design of the study. YM—participated in the design of the study. TY—participated in the design of the study. MS—participated in design of the study. YT—participated in the design of the study, carried out the molecular biology work. SH—participated in the design of the study. TI—participated in the design of the study. MO—participated in the design of the study. HK—Coordinated and conceived the study, being involved in the critical revision of the manuscript for important intellectual content.

Data availability All data analyzed during this study are included in this published article.

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

Ethical approval This study was approved by the Ethical Review Board for Human Genome Studies at Fujita Health University. The written informed consent was obtained from patients. All experiments were carried out in accordance with the relevant guidelines and regulations.

Informed consent We have obtained consent to participate in the study. We have obtained consent to publication in the study.

References

Ballif BC, Yu W, Shaw CA et al (2003) Monosomy 1p36 breakpoint junctions suggest pre-meiotic breakage-fusion-bridge cycles are involved in generating terminal deletions. Hum Mol Genet 12:2153–2165. https://doi.org/10.1093/hmg/ddg231

Bonaglia MC, Giorda R, Massagli A et al (2008) A familial inverted duplication/deletion of 2p25.1–25.3 provides new clues on the genesis of inverted duplications. Eur J Hum Genet 17:179–186. https://doi.org/10.1038/ejhg.2008.160

Chabchoub E, Rodriguez L, Galan E et al (2007) Molecular characterisation of a mosaicism with a complex chromosome rearrangement: evidence for coincident chromosome healing by telomere capture and neo-telomere formation. J Med Genet 44:250–256. https://doi.org/10.1136/jmg.2006.045476

Fan X, Abbott TE, Larson D, Chen K (2014) BreakDancer: Identification of genomic structural variation from paired-end read mapping. Curr Protoc Bioinformatics 45:15.6.1–11. https://doi.org/10.1002/0471250953.bi1506s45

García-Santiago FA, Martínez-Glez V, Santos F et al (2015) Analysis of invdupdel(8p) rearrangement: clinical, cytogenetic and molecular characterization. Am J Med Genet 167A:1018–1025. https://doi.org/10.1002/ajmg.a.36879

Giglio S, Broman KW, Matsumoto N et al (2001) Olfactory receptor-gene clusters, genomic-inversion polymorphisms, and common chromosome rearrangements. Am J Hum Genet 68:874–883. https://doi.org/10.1086/319506

Hermetz KE, Newman S, Conneely KN et al (2014) Large inverted duplications in the human genome form via a fold-back mechanism. PLoS Genet 10:e1004139–e1004214. https://doi.org/10.1371/journal.pgen.1004139

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595. https://doi.org/10.1093/bioinformatics/btp698

Lopes M, Foiani M, Sogo JM (2006) Multiple mechanisms control chromosome integrity after replication fork uncoupling and restart at irreparable UV lesions. Mol Cell 21:15–27. https://doi.org/10.1016/j.molcel.2005.11.015

Noe L, Kucherov G (2005) YASS: enhancing the sensitivity of DNA similarity search. Nucleic Acids Res 33:W540–W543. https://doi.org/10.1093/nar/gki478

Pedurupillay CRJ, Misceo D, Gamage TH et al (2014) Post-zygotic breakage of a dicentric chromosome results in mosaicism for a telocentric 9p marker chromosome in a boy with developmental delay. Gene 533:403–410. https://doi.org/10.1016/j.gene.2013.09.090

Rowe LR, Lee J-Y, Rector L et al (2009) U-type exchange is the most frequent mechanism for inverted duplication with terminal deletion rearrangements. J Med Genet 46:694–702. https://doi.org/10.1136/jmg.2008.065052

Schlade-Bartusiak K, Tucker T, Safavi H et al (2013) Independent post-zygotic breaks of a dicentric chromosome result in mosaicism for an inverted duplication deletion 9p and terminal deletion 9p. Eur J Med Genet 56:229–235. https://doi.org/10.1016/j.ejmg.2013.01.013

Shimokawa O, Kurosawa K, Ida T et al (2004) Molecular characterization of inv dup del(8p): analysis of five cases. Am J Med Genet 128A:133–137. https://doi.org/10.1002/ajmg.a.30063

Soler A, Sánchez A, Carrió A et al (2003) Fetoplacental discrepancy involving structural abnormalities of chromosome 8 detected by prenatal diagnosis. Prenat Diagn 23:319–322. https://doi.org/10.1002/pd.590

Stimpson KM, Song IY, Jauch A et al (2010) Telomere disruption results in non-random formation of de novo dicentric chromosomes involving acrocentric human chromosomes. PLoS Genet 6:e1001061–e1001119. https://doi.org/10.1371/journal.pgen.1001061

Thorvaldsdottir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14:178–192. https://doi.org/10.1093/bib/bbs017

Weckselblatt B, Rudd MK (2015) Human structural variation: mechanisms of chromosome rearrangements. Trends Genet. https://doi.org/10.1016/j.tig.2015.05.010

Yu S, Graf WD (2010) Telomere capture as a frequent mechanism for stabilization of the terminal chromosomal deletion associated with inverted duplication. Cytogenet Genome Res 129:265–274. https://doi.org/10.1159/000315887

Yu S, Fiedler S, Stegner A, Graf WD (2010) Genomic profile of copy number variants on the short arm of human chromosome 8. Eur J Hum Genet 18:1114–1120. https://doi.org/10.1038/ejhg.2010.66

Zhang F, Carvalho CMB, Lupski JR (2009a) Complex human chromosomal and genomic rearrangements. Trends Genet 25:298–307. https://doi.org/10.1016/j.tig.2009.05.005

Zhang F, Khajavi M, Connolly AM et al (2009b) The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. Nat Genet 41:849–853. https://doi.org/10.1038/ng.399

Zuffardi O, Bonaglia M, Ciccone R, Giorda R (2009) Inverted duplications deletions: underdiagnosed rearrangements?? Clin Genet 75:505–513. https://doi.org/10.1111/j.1399-0004.2009.01187.x

Affiliations

Takema Kato[1] · Hidehito Inagaki[1] · Syunsuke Miyai[1] · Fumihiko Suzuki[1] · Yuki Naru[1] · Yasuko Shinkai[1] · Asuka Kato[1] · Kazuo Kanyama[1] · Seiji Mizuno[2] · Yukako Muramatsu[3] · Toshiyuki Yamamoto[4] · Mitsuhisa Shinya[5,6] · Yukiko Tazaki[5,6] · Sayuri Hiwatashi[5,6] · Toshiro Ikeda[5,6] · Mamoru Ozaki[7] · Hiroki Kurahashi[1,2]

Takema Kato
takema@fujita-hu.ac.jp

Hidehito Inagaki
hinagaki@fujita-hu.ac.jp

Syunsuke Miyai
smiyai@fujita-hu.ac.jp

Fumihiko Suzuki
fsuzuki@fujita-hu.ac.jp

Yuki Naru
y.naru@ovus.co.jp

Yasuko Shinkai
y.shinkai@ovus.co.jp

Asuka Kato
asukak@fujita-hu.ac.jp

Kazuo Kanyama
kanyamk@labcorp.com

Seiji Mizuno
seiji_mizuno@aichi-colony.jp

Yukako Muramatsu
murayuka@med.nagoya-u.ac.jp

Toshiyuki Yamamoto
yamamoto.toshiyuki@twmu.ac.jp

Mitsuhisa Shinya
m-shinya@m3.kufm.kagoshima-u.ac.jp

Yukiko Tazaki
tazaki-y@m.kufm.kagoshima-u.ac.jp

Sayuri Hiwatashi
sayuri1030@yahoo.co.jp

Toshiro Ikeda
ikeda@m2.kufm.kagoshima-u.ac.jp

Mamoru Ozaki
m-ozaki@kanazawa-med.ac.jp

[1] Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, 1-98 Dengakugakubo, Kutsukake-cho, Toyoake, Aichi 470-1192, Japan

[2] Department of Clinical Genetics, Central Hospital, Aichi Developmental Disability Center, Kasugai, Kasugai, Japan

[3] Department of Pediatrics, Nagoya University Graduate School of Medicine, Nagoya, Japan

[4] Institute of Medical Genetics, Tokyo Women's Medical University, Shinjuku, Japan

[5] Genetic Counseling Room, Kagoshima University Hospital, Kagoshima, Japan

[6] Department of Obstetrics and Gynecology, Faculty of Medicine, Kagoshima, Japan

[7] Division of Genomic Medicine, Department of Advanced Medicine, Medical Research Institute, Kanazawa Medical University, Uchinada, Japan

Short Report

Analysis of the Origin of Double Mosaic Aneuploidy in Two Cases

Takema Kato[a]   Miki Kawai[a]   Shunsuke Miyai[a]   Fumihiko Suzuki[a]

Makiko Tsutsumi[a]   Seiji Mizuno[b, c]   Toshiro Ikeda[d, e]   Hiroki Kurahashi[a]

[a]Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Japan; [b]Department of Pediatrics, Aichi Human Service Center, Central Hospital, Kasugai, Japan; [c]Department of Clinical Genetics, Aichi Developmental Disability Center, Central Hospital, Kasugai, Japan; [d]Genetic Counseling Room, Kagoshima University Hospital, Kagoshima, Japan; [e]Department of Obstetrics and Gynecology, Faculty of Medicine, Kagoshima University, Kagoshima, Japan

Abstract

We present 2 cases of double mosaic aneuploidy harboring 2 or more different aneuploid cell lines, but no line with a normal chromosome constitution. One of these cases presented mosaicism of sex chromosome aneuploid cell lines (47,XXX/45,X) along with another line containing an autosomal trisomy (47,XX,+8), while the other case showed mosaicism of 2 different autosomal trisomy cell lines (47,XY,+5 and 47,XY,+8). To elucidate the mechanisms underlying these mosaicisms, we conducted molecular cytogenetic analyses. Genotyping data from the SNP microarray indicated that 2 sequential meiotic or early postzygotic segregation errors likely had occurred followed by natural selection. These cases suggest that frequent segregation errors and selection events in the meiotic and early postzygotic stages lead to this condition.

© 2020 S. Karger AG, Basel

In the course of blood karyotyping for the etiological analysis of multiple congenital anomalies or intellectual disabilities, or karyotyping of product of conception (POC) samples to investigate pregnancy loss, occasionally aneuploidy of different chromosomes is observed in 2 or more different cell lines as a mosaic. This is known as double mosaic aneuploidy. As this condition generally arises via 2 independent segregation errors in different cells, cells with a normal chromosome constitution are usually also present. However, in rare cases, double mosaic aneuploidy is seen in the absence of normal cell lines [Schofield et al., 1992; Harada et al., 1998; Schubert et al., 2002; Ryu et al., 2010; Kumar et al., 2014]. In these cases, the cytogenetic mechanisms leading to this condition are hard to understand.

In our current study, we describe 2 cases of double (triple) mosaic aneuploidy in which each cell harbored a single trisomic or monosomic chromosome, and no cells with normal chromosome content were observed. We performed molecular analyses and assessed the etiology of these cases.

T.K. and M.K. contributed equally to this work.

Material and Methods

For our analyses, we obtained blood samples and buccal mucosa cells from case 1 and POC samples from case 2.

Karyotyping

Chromosomal karyotyping of POC and blood samples was conducted using conventional G-banding methods.

Fluorescence in situ Hybridization

Interphase FISH analysis of peripheral blood and buccal mucosa cells was performed using a mixture of 2 probes: Vysis CEP 8 (D8Z2) Spectrum Green (Abbott Laboratories, IL) and Chromosome X Alpha Satellite (DXZ1) Texas Red (Cytocell Ltd, UK).

SNP Array

For SNP microarray analysis, genomic DNA was extracted from POC and blood specimens using a Gentra Puregene Tissue Kit (Qiagen, Germany). We then used the Affymetrix Cytoscan 750k array (Affymetrix, Santa Clara, CA) and the Illumina HumanCytoSNP-12 array (Illumina, San Diego, CA) for genotyping. Sample preparation was performed following the manufacturer's protocol. Copy number alterations and genotypes on the CytoScan 750k array were analyzed using Affymetrix® Chromosome Analysis Suite Software v4.0 (ChAS). Analysis of the CytoSNP-12 array data was carried out using GenomeStudio software (Illumina).

NGS Copy Number Analysis

Comprehensive copy number analysis was conducted using a next-generation sequencer (NGS), according to the manufacturer's protocol (Illumina). Genomic DNA was diluted and subjected to whole-genome amplification using SurePlex. Then, samples were prepared for NGS analysis. The library was sequenced using the VeriSeq PGS Kit - MiSeq system, and these sequencing data were analyzed using BlueFuse Multi Software (Illumina).

Results

Case 1

The 3-month-old girl was referred to our facility due to growth retardation and relative macrocephaly. She was the first-born child of a 35-year-old healthy mother and a 30-year-old healthy father. No dysmorphic features other than macrocephaly were observed, although agenesis of the corpus callosum was found by cerebral MRI, and a ventricular septal defect was also noted. The karyotype of the peripheral blood of the proband was mos 47,XX,+8[16]/47,XXX[4]. This double mosaic aneuploidy was found to be unique as it consisted of 2 cell lines that were aneuploid for a single different chromosome; no cells with normal chromosome content were observed. Since the proband had no clinical findings of sex chromosome aneuploidy or trisomy 8 mosaicism, we performed additional cytogenetic analyses. Repeated evaluations of the peripheral blood karyotype revealed mos 47,XX,+8[37]/47,XXX[8]/45,X[5]. FISH analysis of her buccal mucosa cells also indicated mosaic aneuploidy but with a different frequency, i.e., mos 47,XX,+8[5]/47,XXX[84]/45,X[11] (Fig. 1A, B). Her mild clinical findings might be explained by the low rate of trisomy 8 mosaicism in non-blood tissues. Unfortunately, we could not obtain the parental samples.

To further investigate the origin of the triple mosaic aneuploidy, we employed SNP microarray analysis of genomic DNA from blood. B allele frequency (BAF) analysis indicated that both chromosomes 8 and X had a 4-allele combination (Fig. 1C). The estimated mosaic rate was 50% for chromosome 8, whilst that for chromosome X was reduced, probably due to an compensation by the 45,X and 47,XXX lines. Since the other chromosomes did not show the 4-allele combination, a chimera of a zygote with trisomy 8 and another zygote with sex chromosome aneuploidy is unlikely. A 6-allele combination was not found in chromosome 8, suggesting an identical haplotype of its 2 copies. These data indicated that the segregation error for chromosome 8 might have been a meiosis II error between 2 sister chromatids without crossover in meiosis I, or a post-zygotic mitosis error. Because trisomy X and monosomy X were observed in case 1, a chromosome X nondisjunction should have occurred at the postzygotic stage.

From the aforementioned findings, we propose 2 possible pathways for this case of triple mosaic aneuploidy. Our first hypothesis is that a nondisjunction of chromosomes 8 and X occurred independently in each cell at the second postzygotic cell division. Monosomy 8 cells were then negatively selected (Fig. 1D-I). The other hypothesis assumes that the zygote is trisomic for chromosome 8 due to a meiosis II error. The extra chromosome 8 might be lost from one cell in the first postzygotic cell division. Then, in the following second cell division, the disomic cell might undergo mitotic nondisjunction of chromosome X (Fig. 1D-II).

Case 2

A 48-year-old woman presented at our clinic due to 3 consecutive miscarriages in which the pregnancy loss occurred at gestational week 7. Chromosome karyotyping of the POC sample dissected from chorionic villi revealed mos 47,XY,+5[48]/47,XY,+8[2] (Fig. 2A). To confirm this rare karyotype, we employed NGS chromosome analysis using genomic DNA from the surface of an uncultured POC sample. Full trisomy of chromosome 8 was observed, but the copy number was neutral for chromo-
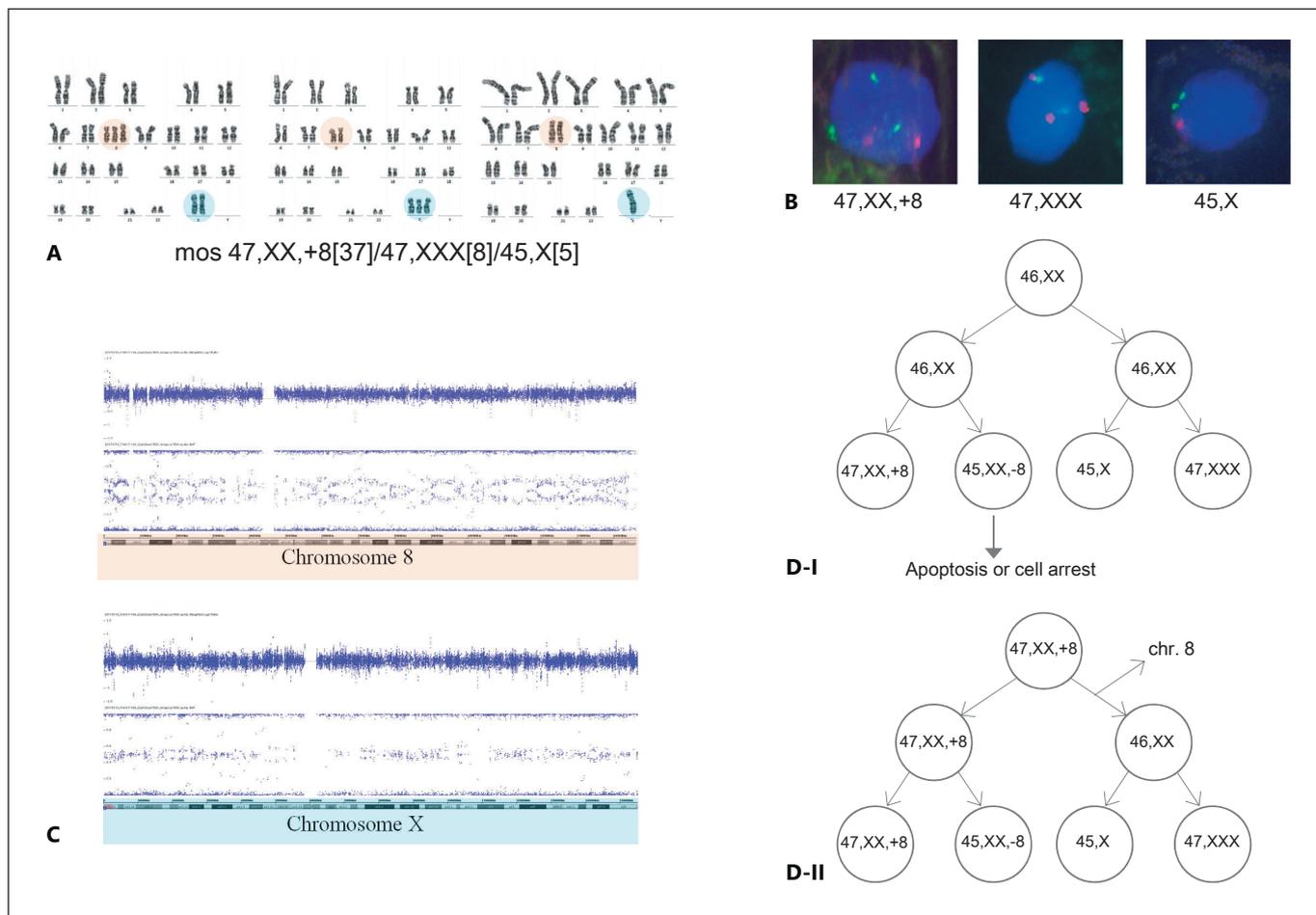
Kato/Kawai/Miyai/Suzuki/Tsutsumi/
Mizuno/Ikeda/Kurahashi

Fig. 1. Mosaic triple aneuploidy was identified in an infant with growth retardation and relative macrocephaly. **A** G-banding of peripheral blood lymphocytes revealed a mos 47,XX,+8/47,XXX/45,X karyotype. **B** Interphase FISH with alpha satellite probes specific for chromosome 8 (green) and chromosome X (red) on buccal mucosa cells. **C** SNP microarray analysis was used to determine a weighted log2 ratio (top) and B allele frequency (bottom) for chromosomes 8 and X. **D** Two possible mechanisms leading to the mosaic triple aneuploidy in this case (see text). The first 2 postzygotic divisions are shown.

some 5 (Fig. 2B). We speculated that this discordance might be attributable to differences in the cell source, i.e., trophoblasts in the uncultured sample and mesenchymal cell lineages in the cultured sample. The double mosaic aneuploidy detected in this POC was unique since it consisted of 2 cell lines that were aneuploid for a single, different chromosome, and no cells with normal chromosome content were observed. Unfortunately, we could not obtain the parental samples.

To determine the origin of the double mosaic aneuploidy in this case, we carried out SNP microarray analysis of genomic DNA isolated from an uncultured POC sample. The BAF results revealed a 4- or 6-allele combination pattern in both chromosomes 5 and 8 but in no other chromosome, indicating that a chimera of different zygotes each with a trisomy 5 or trisomy 8 was unlikely. The mosaic ratio estimated using the BAF was 20–40% for trisomy 5 and 60–70% for trisomy 8. The BAF of chromosome 5 showed a mixture of 4-allele and 6-allele combinations and a 4-allele combination surrounding the centromere, indicating a malsegregation event during meiosis II between sister chromatids with crossover (Fig. 2C). In contrast, chromosome 8 showed only a 4-allele combination throughout. This indicated that the missegregation of chromosome 8 would be due to a postzygotic mitotic error or meiosis II error between sister chromatids without crossover (Fig. 2C).
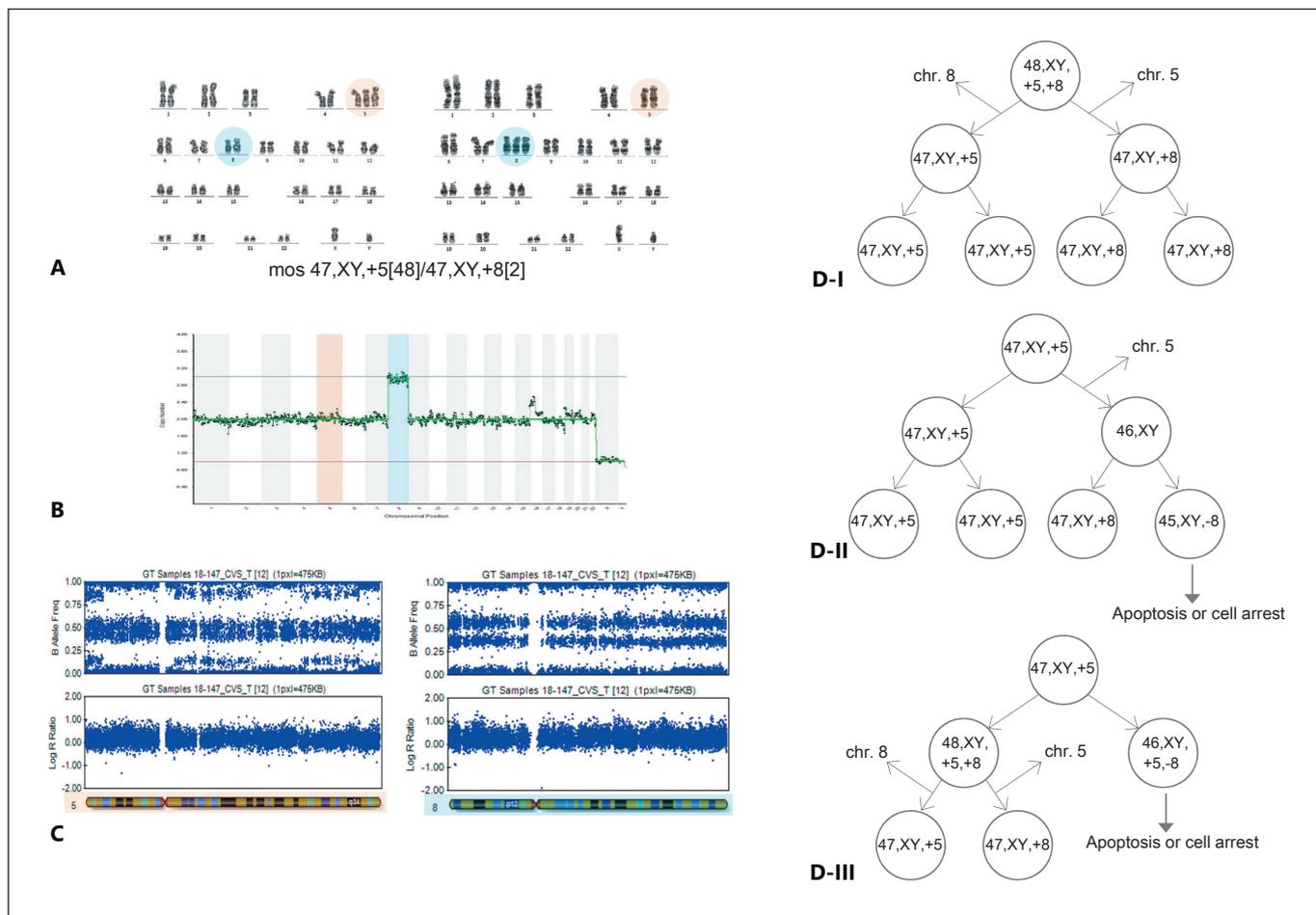
Fig. 2. Mosaic double trisomy was identified in the POC sample of a 48-year-old woman. **A** G-banding revealed a mos 47,XY,+5/47,XY,+8 karyotype in the POC sample. **B** NGS analysis of DNA from an uncultured sample indicated a neutral copy number for chromosome 5 and a full chromosome 8 trisomy. **C** SNP microarray analysis of the B allele frequency (top) and logR ratio (bottom) for chromosomes 5 and 8. **D** Three possible mechanisms for the mosaic double aneuploidy in this case (see text). The first 2 postzygotic divisions are shown.

Based on these observations, we surmised that there were 3 possible pathways that could have led to the double mosaic aneuploidy in case 2. The first was based on an original zygote with 48,XY,+5,+8 possibly derived from meiotic errors. During the first postzygotic mitotic cell division, loss of chromosome 5 or 8 in either of the 2 cells, respectively, resulted in 47,XY,+5 and 47,XY,+8 cells (Fig. 2D-I). A second possibility was that a 47,XY,+5 zygote, derived from a meiotic error, led by loss of chromosome 5 in the first postzygotic cell division to 47,XY,+5 and 46,XY cells, respectively. In the second cell division, the rescued 46,XY cell underwent mitotic nondisjunction of chromosome 8, followed by negative selection of the monosomy 8 cell (Fig. 2D-II). The third possible pathway was that an original

zygote with trisomy 5 underwent mitotic nondisjunction of chromosome 8 at the first cell division. This would have generated a 48,XY,+5,+8 and a 46,XY,+5,−8 cell underlying negative selection. In the subsequent second mitotic cell division, the 48,XY,+5,+8 cell would then have lost chromosome 5 or 8 in either of the 2 cells, resulting in 47,XY,+5 and 47,XY,+8 cells (Fig. 2D-III).

Discussion and Conclusion

Double mosaic aneuploidy without a euploid cell line is a very rare condition, and only a small number of cases have been reported to date [Schofield et al., 1992; Harada

et al., 1998; Schubert et al., 2002; Ryu et al., 2010]. This is likely due to a low fetal viability in the absence of euploid cells. When 2 different numerical chromosome abnormalities occur simultaneously or sequentially in the same cell line, a subset of cells should carry these 2 types of aneuploidy parallelly. When 2 different numerical chromosome abnormalities arise independently in different cell lines, each cell line should carry each single aneuploidy separately. In the second situation, original euploid cells should be generally present. However, we have here presented 2 cases of double mosaic aneuploidy in the absence of normal cells, in which the underlying cytogenetic mechanisms are hard to understand.

Our molecular cytogenetic analyses of these 2 cases indicated that the double mosaic aneuploidy arose via 2 successive segregation errors either during maternal meiosis or in early postzygotic stages, which was then followed by natural selection. The errors may have been nondisjunction or predivision during meiosis I or II, or missegregation during mitosis at the early postzygotic stage, such as nondisjunction or anaphase lagging [Conlin et al., 2010]. Such phenomena usually produce monosomy or trisomy, but in some cases can lead to monosomy or trisomy rescue. It has been amply documented that a high level of chromosome instability develops during meiosis in oogenesis and in cleavage-stage embryos [Vanneste et al., 2009; Nagaoka et al., 2012; Fragouli et al., 2013; Potapova and Gorbsky, 2017]. This was further validated by the recent advances in single-cell genomic analysis using NGS [Fiorentino et al., 2014]. Additionally, it has been demonstrated that the large amount of cytoplasm present in oocytes and cleavage stage embryos induces error-prone chromosome segregation via dilution of spindle assembly checkpoint proteins [Kyogoku and Kitajima, 2017]. This instability might contribute to the 2 successive segregation errors at this stage in a single zygote, thereby leading to the development of double mosaic aneuploidy.

Most of the reported double mosaic aneuploidies are combinations of autosomal and gonosomal aneuploidy [Schofield et al., 1992; Harada et al., 1998; Schubert et al., 2002; Ryu et al., 2010; Kumar et al., 2014; present case 1]. Trisomy 13, trisomy 18, trisomy 21, and sex chromosome aneuploidies are often involved due to phenotypic selection. In contrast, the combination of 2 different autosomal aneuploidies is very rare [Huijsdens-van Amsterdam et al., 2012; Jurcă et al., 2018; present case 2]. Notably, many reports of this rare condition, including our present study, have described chromosome 8 mosaic trisomy [Schofield et al., 1992; DeBrasi et al., 1995]. Prior

trisomy 8 nondisjunction studies have found that most mosaic trisomy 8 cases are likely due to mitotic duplication, with only a few cases arising from a maternal meiotic nondisjunction. Those results are in contrast to the findings for common autosomal trisomies where the majority of cases are due to errors in maternal meiosis [Karadima et al., 1998]. However, the reasons why chromosome 8 is frequently involved in double mosaic aneuploidy are unclear. Cell lineage-specific mitotic duplication or cell lineage-specific selection may be possible mechanisms. Future detailed case reports will be required to clarify the mechanisms underlying the involvement of specific chromosomes in double mosaic aneuploidy.

Author Contributions

T.K. and M.K. drafted the manuscript. T.K., M.K., S.M., F.S., M.T., S.M., and T.I. participated in designing the study and carried out the molecular biology work. H.K. coordinated and conceived the study and was involved in the critical revision of the manuscript for important intellectual content.

References

Conlin LK, Thiel BD, Bonnemann CG, Medne L, Ernst LM, et al: Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis. Hum Mol Genet 19:1263–1275 (2010).

DeBrasi D, Genardi M, D'Agostino A, Calvieri F, Tozzi C, et al: Double autosomal/gonosomal mosaic aneuploidy: study of nondisjunction in two cases with trisomy of chromosome 8. Hum Genet 95:519–525 (1995).

Fiorentino F, Biricik A, Bono S, Spizzichino L, Cotroneo E, et al: Development and validation of a next-generation sequencing-based protocol for 24-chromosome aneuploidy screening of embryos. Fertil Steril 101:1375–1382 (2014).

Fragouli E, Alfarawati S, Spath K, Jaroudi S, Sarasa J, et al: The origin and impact of embryonic aneuploidy. Hum Genet 132:1001–1013 (2013).

Harada N, Abe K, Nishimura T, Sasaki K, Ishikawa M, et al: Origin and mechanism of formation of 45,X/47,XX,+21 mosaicism in a fetus. Am J Med Genet 75:432–437 (1998).

Huijsdens-van Amsterdam K, Barge-Schaapveld DQ, Mathijssen IB, Alders M, Pajkrt E, Knegt AC: Prenatal diagnosis of a trisomy 7/trisomy 13 mosaicism. Mol Cytogenet 5:8 (2012).

Jurcă MC, Bembea M, Iuhas OA, Kozma K, Petcheşi CD, et al: Double autosomal trisomy with mosaicism 47,XY(+8)/47,XY(+21). Morphological and genetic changes of a rare case. Rom J Morphol Embryol 59:985–988 (2018).

Karadima G, Bugge M, Nicolaidis P, Vassilopoulos D, Avramopoulos D, et al: Origin of nondisjunction in trisomy 8 and trisomy 8 mosaicism. Eur J Hum Genet 6:432–438 (1998).

Kumar M, Lal V, Chapadgaonkar S, Bhattacharya SK: A rare case of mosaic 45,X/47,XX,+13 in 28-year-old women with secondary amenorrhoea: a case report and literature review. Meta Gene 3:8–13 (2014).

Kyogoku H, Kitajima TS: Large cytoplasm is linked to the error-prone nature of oocytes. Dev Cell 41:287–298 (2017).

Nagaoka SI, Hassold TJ, Hunt PA: Human aneuploidy: mechanisms and new insights into an age-old problem. Nat Rev Genet 13:493–504 (2012).

Potapova T, Gorbsky G: The consequences of chromosome segregation errors in mitosis and meiosis. Biology (Basel) 6:12 (2017).

Ryu SW, Lee G, Baik CS, Shim SH, Kim JT, et al: Down-Turner syndrome (45,X/47,XY,+21): case report and review (in Korean). Korean J Lab Med 30:195–200 (2010).

Schofield B, Babu A, Punales-Morejon D, Popescu S, Leiter E, et al: Double mosaic aneuploidy: 45,X/47,XY,+8 in a male infant. Am J Med Genet 44:7–10 (1992).

Schubert R, Eggermann T, Hofstaetter C, von Netzer B, Knöpfle G, Schwanitz G: Clinical, cytogenetic, and molecular findings in 45,X/47,XX,+18 mosaicism: clinical report and review of the literature. Am J Med Genet A 110:278–282 (2002).

Vanneste E, Voet T, Le Caignec C, Ampe M, Konings P, et al: Chromosome instability is common in human cleavage-stage embryos. Nat Med 15:577–583 (2009).

A case of a parthenogenetic 46,XX/46,XY chimera presenting ambiguous genitalia

Rie Kawamura[1] · Takema Kato[1] · Shunsuke Miyai[1] · Fumihiko Suzuki[1] · Yuki Naru[1] · Maki Kato[1] · Keiko Tanaka[2] ·
Miwako Nagasaka[2,3] · Makiko Tsutsumi[1] · Hidehito Inagaki[1] · Tomoaki Ioroi[4] · Makiko Yoshida[5] · Tomoya Nao[6] ·
Laura K. Conlin[7] · Kazumoto Iijima[2,3] · Hiroki Kurahashi[1] · Mariko Taniguchi-Ikeda[1,8]

Abstract

Sex-chromosome discordant chimerism (XX/XY chimerism) is a rare chromosomal disorder in humans. We report a boy with ambiguous genitalia and hypospadias, showing 46,XY[26]/46,XX[4] in peripheral blood cells. To clarify the mechanism of how this chimerism took place, we carried out whole-genome genotyping using a SNP array and microsatellite analysis. The B-allele frequency of the SNP array showed a mixture of three and five allele combinations, which excluded mosaicism but not chimerism, and suggested the fusion of two embryos or a shared parental haplotype between the two parental cells. All microsatellite markers showed a single maternal allele. From these results, we concluded that this XX/XY chimera is composed of two different paternal alleles and a single duplicated maternal genome. This XX/XY chimera likely arose from a diploid maternal cell that was formed via endoduplication of the maternal genome just before fertilization, being fertilized with both X and Y sperm.

These authors contributed equally: Rie Kawamura, Takema Kato

Supplementary information The online version of this article (https://doi.org/10.1038/s10038-020-0748-4) contains supplementary material, which is available to authorized users.

✉ Mariko Taniguchi-Ikeda
mtani@fujita-hu.ac.jp

1   Department of Molecular Genetics, Fujita Health University, Toyoake, Japan

2   Division of Genetic Counseling, Kobe University Hospital, Kobe, Japan

3   Department of Pediatrics, Kobe University Graduate School of Medicine, Kobe, Japan

4   Department of Pediatrics, Comprehensive Perinatal Maternal and Child Medical Center, Himeji Red Cross Hospital, Himeji, Japan

5   Department of Pathology, Hyogo Prefectural Kobe Children's Hospital, Kobe, Japan

6   Department of Urology, Hyogo Prefectural Kobe Children's Hospital, Kobe, Japan

7   Division of Genomic Diagnostics, Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, PA, USA

8   Department of Clinical Genetics, Fujita Health University Hospital, Toyoake, Japan

A chimera is defined as the fusion product of two different zygotes in a single embryo, whereas a mosaic results from a mitotic error in a single zygote. Sex-chromosome discordant chimerism in humans (XX/XY chimerism) is a rare chromosomal abnormality. Although the first case was described in 1962 [1], its incidence is still unknown. The XX/XY chimera manifests variable genital phenotypes, ranging from normal male or female genitalia to different degrees of ambiguous genitalia. Chimeras are thought to result from a defect in the processes near the time of fertilization. The XX/XY chimera is classified into some subtypes; tetragametic chimeras [2–5], parthenogenetic chimeras [5–9], androgenetic chimeras [5, 10, 11] and sesquizygotic twinning chimera [12]. Whereas tetragametic chimeras are known as the most common subtype of XX/XY chimeras, which are derived from the simple fusion of two different zygotes, parthenogenetic chimeras and androgenetic chimeras [8, 11] undergo endoreplication of one of the gametic genomes. Therefore, genotyping of XX/XY chimeras is important not only to clarify its developmental mechanism but also for diagnosis and treatment, as such patients occasionally present with fertility problems. In this report, we present a case of a parthenogenic chimera with a karyotype of XX/XY, together with a literature review.
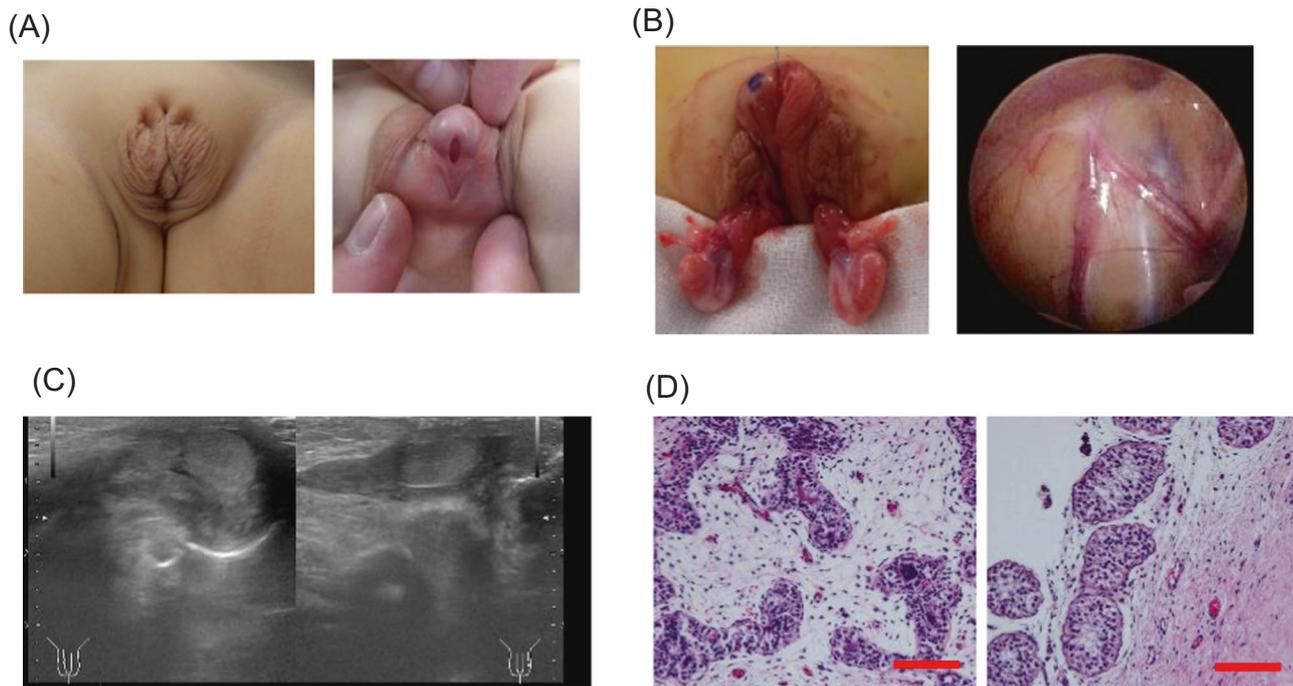
(A)

(B)

(C)

(D)

Fig. 1 Clinical findings of the patient. **a** On examination, the patient showed cryptorchidism (left), hypospadias (right), and anterior scrotum (right). **b** On operation, epididymes were identified macroscopically, but the tunica albuginea of the testis was absent bilaterally (left). Laparoscopy demonstrated that the vas deferens and gonadal veins flowed normally into the bilateral inguinal rings, and ovaries and a uterus were not detected (right). The precise position of the external urethral opening was normal. **c** The ultrasound echoic level of both gonads was homogeneous. No other tissue-like structures with margins, suggestive of ovarian tissue, were detected. **d** Testis biopsy of the patient (left). Compared to control tissue (right, age-matched normal testis), interstitial tissue of the patient was edematous and more prominent than the seminiferous tubules, which were tortuous and diverging. The seminiferous tubules appeared dysplastic, but detailed analysis demonstrated that they were not dysplastic. Bar, 100 μm

The patient was born from healthy parents at an affiliated hospital as a boy with ambiguous genitalia, hypospadias cryptorchidism, and an anterior scrotum (Fig. 1a). The pregnancy was not a result of in vitro fertilization treatment. His father was 43 years old and mother was 31 years old. Abdominal magnetic resonance imaging displayed no uterus or ovaries. His growth milestone at 1 year was normal. The patient's toy preference was wheeled vehicles and superheroes. Macro- and microscopically, the patient had ambiguous external genitalia, male-type internal genitalia (epididymis), and bilateral testes (Fig. 1b–d). The family chose a male sex to assign. Hypospadias was repaired. The parents were recommended to undergo genetic counseling regarding detailed genetic analysis as well as recurrence risk. G-banding from peripheral blood cells of the patient showed 46,XY[26]/46,XX[4]. To analyze the XX/XY ratio further, we performed fluorescence in situ hybridization (FISH) analysis on interphase nuclei from the patient's buccal mucosal cells with specific probes for chromosomes X and Y by AneuVysion Assay Kit (Abbott, Tokyo, Japan). FISH showed nuc ish XX[83]/XY[17]. The XY/XX ratio in buccal cells (17%) was lower than that in peripheral blood cells (87%). After receiving approval from the Ethics Review Board for Human Genome Studies at Fujita

Health University and written informed consent from the parents to participate in our study, genetic diagnosis was performed.

To confirm chimerism, we carried out SNP microarray analysis using a CytoScan 750 K Array (Affymetrix, Santa Clara, CA). We used genomic DNA, which was isolated from nail as a template for microarray analysis. Scan data were visualized using ChAS 3.2 software (Affymetrix). The copy number state of chromosome X was 1.5, which means that the ratio of XX to XY was about 50% (Fig. 2a). Next, to confirm chimerism, we analyzed the B-allele frequency of autosomes and the X chromosome [13, 14]. We found both a region showing five allele combinations, where possible SNP genotypes were BB–BB, BB–BA, BB–AA, BA–AA, and AA–AA, which means that there were three or four chromosome sets, and a region showing three allele combinations, where the possible SNP genotypes were BB–BB, AB–AB, and AA–AA, which means that there were two identical chromosome sets in the autosomes (Fig. 2b). The region showing four allele combinations, where the possible SNP genotypes were BB–B, BB–A, AA–B, and AA–A, were on chromosome X (Fig. 2b). These allele combinations were detected throughout the whole genome (Supplementary Fig. 1). This indicates the
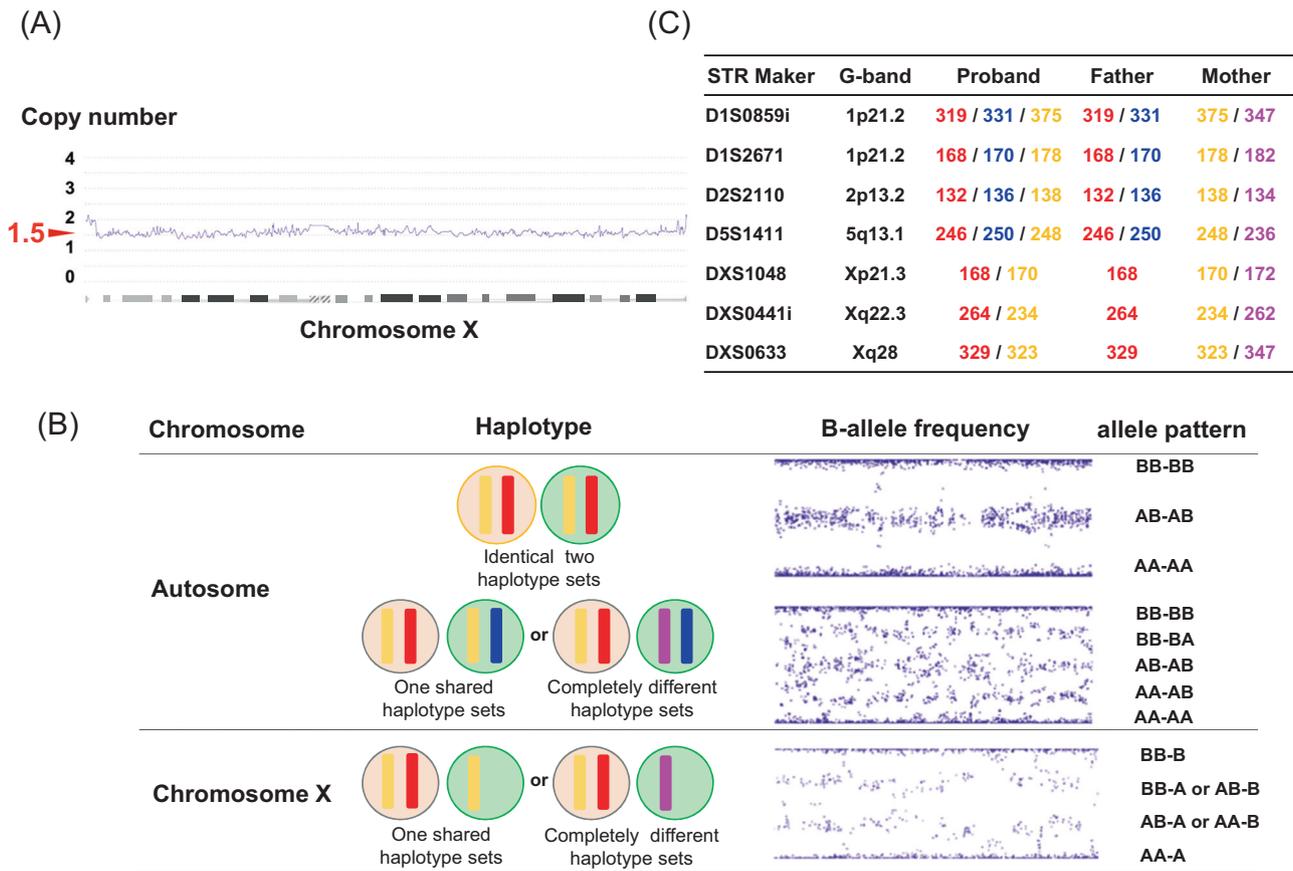
(A)

Copy number

Chromosome X

(C)

STR Maker	G-band	Proband	Father	Mother
D1S0859i	1p21.2	319 / 331 / 375	319 / 331	375 / 347
D1S2671	1p21.2	168 / 170 / 178	168 / 170	178 / 182
D2S2110	2p13.2	132 / 136 / 138	132 / 136	138 / 134
D5S1411	5q13.1	246 / 250 / 248	246 / 250	248 / 236
DXS1048	Xp21.3	168 / 170	168	170 / 172
DXS0441i	Xq22.3	264 / 234	264	234 / 262
DXS0633	Xq28	329 / 323	329	323 / 347

(B)

Chromosome	Haplotype	B-allele frequency	allele pattern
Autosome	Identical two haplotype sets		BB-BB / AB-AB / AA-AA
Autosome	One shared haplotype sets **or** Completely different haplotype sets		BB-BB / BB-BA / AB-AB / AA-AB / AA-AA
Chromosome X	One shared haplotype sets **or** Completely different haplotype sets		BB-B / BB-A or AB-B / AB-A or AA-B / AA-A

Fig. 2 SNP array and STR analysis. **a** Smooth signal track of the patient's X chromosome. Copy number (*Y* axis) showed 1.5, suggesting that the patient has more than two sets of X chromosomes. **b** A B-allele frequency (BAF) plot and possible allele patterns of the patient's nail DNA, showing a XX to XY ratio of 50% in SNP array analysis. The blue and red colored chromosomes in the circles are both of paternal origin. The yellow and pink colored chromosomes are of maternal origin. **c** STR maker analysis. Red and blue numbers indicate paternal origin, and yellow and pink numbers indicate maternal origin

presence of two different genomes, demonstrating that the patient does not show mosaicism but is an XX/XY chimera. Haplotype sharing region encompasses 1.5 Gbps, which is approximately half of the entire genome (Supplementary Fig. 1), suggesting that the chimera is possibly tetragenic, parthenogenic or androgenic. However, SNP array is not able to classify chimerism.

To investigate the parental origin of the genome in the patient, we performed microsatellite marker analysis, as known as short tandem repeat (STR) analysis. This time, genomic DNA was extracted from the patient's nails and parental peripheral blood cells. We chose STR markers located in the region showing the five allele combinations in autosomes or four allele combinations on chromosome X on SNP array analysis. We first determined the patient's genotype as well as the genotype of the parents. Next, the proband's genotype was determined by using completely different markers in autosomes. All markers showed the contribution of two paternal alleles and one maternal allele in the chimera (Fig. 2c). Therefore, we determined the dispermic origin of the chimera. We concluded that the

patient is a parthenogenetic chimera, which is composed of two paternal alleles and one duplicated maternal allele (Fig. 3a).

Although there are various mechanisms that lead to the development of chimeras, our results suggested that this patient was a parthenogenetic chimera [9]. Oocyte possibly started first mitosis before the disappearance of the pronuclear.

From the point of view of genetic counseling, the recurrence risk of this condition in this family is low because of the lack of a family history. On the other hand, careful follow-up of testicular function is important in considering future family planning and gender identity for the patient. However, as the chimeric ratio in this patient is different among various tissues in the body [15], androgen or estrogen exposure to the entire body cannot be assessed and it is difficult to predict the risk of problems, such as gonadoblastoma (Fig. 3b). In addition, genetic counseling should be recommended so that the patient's gender identity can be determined autonomously when the pathological condition is explained to the patient.
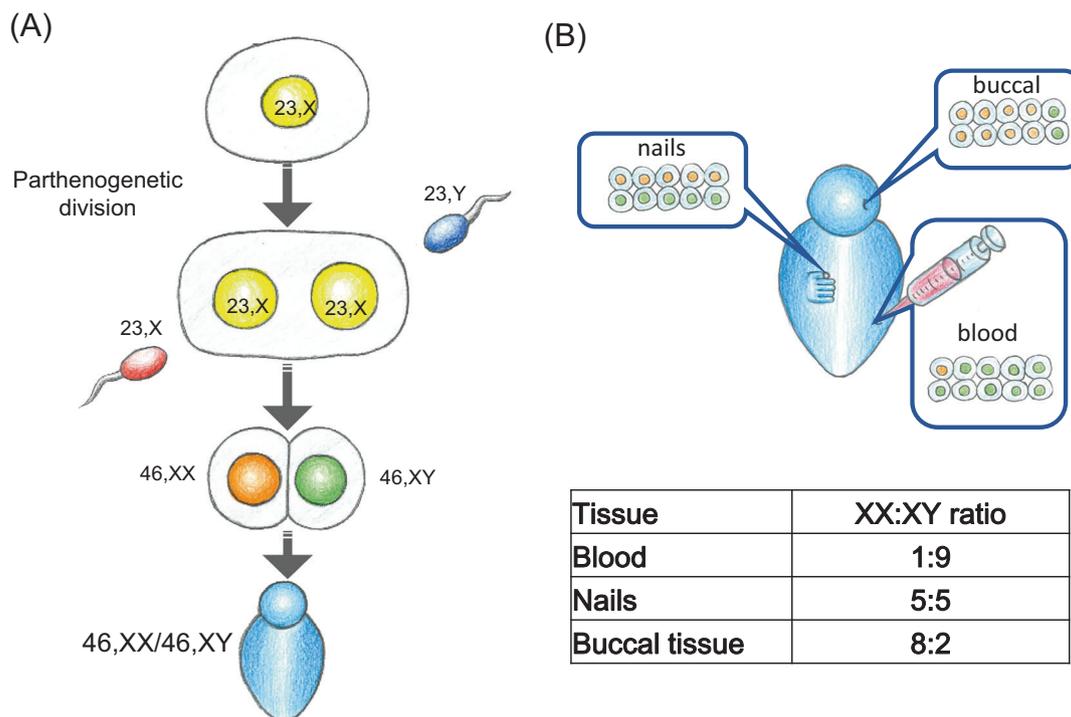
(A)

(B)

Tissue	XX:XY ratio
Blood	1:9
Nails	5:5
Buccal tissue	8:2

Fig. 3 Schematic drawing of how a parthenogenetic chimera is produced. **a** Schematic drawing of how a parthenogenetic chimera is produced. Endoduplication of the maternal genome occurs just before fertilization with two spermatozoa. The colors of the cells are consistent with Fig. 2b, c. **b** Top, diagram of XX and XY chimeric ratios of three tissues from the patient. Orange cells indicate 46,XX, and green cells indicate 46,XY. Bottom, whereas G-banding from the peripheral blood cells of the patient showed 46,XY[26]/46,XX[4] (XX:XY ratio, 1:9), FISH showed nuc ish XX[83]/XY[17] (XX:XY ratio, 8:2) in buccal cells, and the SNP array suggested that both XX and XY are present at about 50% (XX:XY ratio, 5:5)

In the present case, genotyping determined that the patient had a rare condition (XX/XY chimerism). Furthermore, it was suggested that the patient was not the common tetragametic chimera, but a parthenogenetic chimera. Genotyping of XX/XY chimeras is important not only for clarifying the pathogenesis of chimeras but also for understanding of the disease. Therefore, we expect this method to become the standard for patients who are chimeric for the XX/XY karyotype.

Compliance with ethical standards

References

1. Gartler SM, Waxman SH, Giblett E. An XX/XY human hermaphrodite resulting from double fertilization. Proc Natl Acad Sci USA. 1962;48:332–5.
2. Laursen RJ, Alsbjerg B, Vogel I, Gravholt CH, Elbaek H, Lildballe DL, et al. Case of successful IVF treatment of an oligospermic male with 46,XX/46,XY chimerism. J Assist Reprod Genet. 2018;35:1325–8.
3. Sheets KM, Baird ML, Heinig J, Davis D, Sabatini M, Starr DB. A case of chimerism-induced paternity confusion: what ART practitioners can do to prevent future calamity for families. J Assist Reprod Genet. 2018;35:345–52.
4. Malan V, Gesny R, Morichon-Delvallez N, Aubry MC, Benachi A, Sanlaville D, et al. Prenatal diagnosis and normal outcome of a 46, XX/46,XY chimera: a case report. Hum Reprod. 2007;22:1037–41.
5. Malan V, Vekemans M, Turleau C. Chimera and other fertilization errors. Clin Genet. 2006;70:363–73.
6. Shin SY, Yoo HW, Lee BH, Kim KS, Seo EJ. Identification of the mechanism underlying a human chimera by SNP array analysis. Am J Med Genet A. 2012;158:2119–23.
7. Lee KF, Hsu CS, Kuo PL, Chen JL, Jiang YH, Liu IY. The identification of a spontaneous 47,XX,+21/46,XY chimeric fetus with male genitalia. BMC Med Genet. 2012;13:85.
8. Yamazawa K, Nakabayashi K, Kagami M, Sato T, Saitoh S, Horikawa R, et al. Parthenogenetic chimaerism/mosaicism with a Silver-Russell syndrome-like phenotype. J Med Genet. 2010;47:782–5.
9. Giltay JC, Brunt T, Beemer FA, Wit JM, van Amstel HK, Pearson PL, et al. Polymorphic detection of a parthenogenetic

maternal and double paternal contribution to a 46,XX/46,XY hermaphrodite. Am J Hum Genet. 1998;62:937–40.

10. Kaiser-Rogers KA, McFadden DE, Livasy CA, Dansereau J, Jiang R, Knops JF, et al. Androgenetic/biparental mosaicism causes placental mesenchymal dysplasia. J Med Genet. 2006;43:187–92.

11. Giurgea I, Sanlaville D, Fournet JC, Sempoux C, Bellanné-Chantelot C, Touati G, et al. Congenital hyperinsulinism and mosaic abnormalities of the ploidy. J Med Genet. 2006;43:248–54.

12. Gabbett MT, Laporte J, Sekar R, Nandini A, McGrath P, Sapkota Y, et al. Molecular support for heterogonesis resulting in sesqui-zygotic twinning. N Engl J Med. 2019;28:842–9.

13. Davidsson J, Johansson B. Methylation and expression analyses of Pallister-Killian syndrome reveal partial dosage compensation of tetrasomy 12p and hypomethylation of gene-poor regions on 12p. Epigenetics. 2016;3:194–204.

14. Conlin LK, Thiel BD, Bonnemann CG, Medne L, Ernst LM, Zackai EH, et al. Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis. Hum Mol Genet. 2010;19: 1263–75.

15. Biesecker LG, Spinner NB. A genomic view of mosaicism and human disease. Nat Rev Genet. 2013;14:307–20.

BMC Cancer

An aggressive systemic mastocytosis preceded by ovarian dysgerminoma

Makiko Tsutsumi[1†], Hiroki Miura[2†], Hidehito Inagaki[1], Yasuko Shinkai[1], Asuka Kato[1,3], Takema Kato[1], Susumu Hamada-Tsutsumi[4], Makito Tanaka[2], Kazuko Kudo[2], Tetsushi Yoshikawa[2] and Hiroki Kurahashi[1*]

Abstract

Background: Aggressive systemic mastocytosis (ASM) is a rare malignant disease characterized by disordered mast cell accumulation in various organs. We here describe a female ASM patient with a previous history of ovarian dysgerminoma.

Methods: Molecular cytogenomic analyses were performed to elucidate an etiological link between the ASM and dysgerminoma of the patient.

Results: This patient was affected by ovarian dysgerminoma which was treated by chemotherapy and surgical resection. Having subsequently been in complete remission for 2 years, she developed symptoms of ASM. A somatic D816A mutation in the *KIT* gene was detected in her bone marrow, which facilitated the diagnosis of ASM. Unexpectedly, this *KIT* D816A variant was also detected in the prior ovarian dysgerminoma sample. Whole-exome sequencing allowed us to identify a somatic nonsense mutation of the *TP53* gene in the bone marrow, but not in the dysgerminoma. Microarray analysis of the patient's bone marrow revealed a copy-number-neutral loss of heterozygosity at the *TP53* locus, suggestive of the homozygous nonsense mutation in the *TP53* gene. In addition, the loss of heterozygosity at the *TP53* locus was also detected in the dysgerminoma.

Conclusions: These results indicated that either the mast cells causing the ASM in this case had originated from the preceding ovarian dysgerminoma as a clonal evolution of a residual tumor cell, which acquired the *TP53* mutation, or that both tumors developed from a common cancer stem cell carrying the *KIT* D816A variation.

Keywords: Aggressive systemic mastocytosis, *KIT*, Dysgerminoma, Germ cell tumor, *TP53*, Loss of heterozygosity

Background

ASM is one of the advanced forms of systemic mastocytosis (SM) with a poor prognosis. In this disorder, clonal mast cells become abnormally accumulated in the skin, lymph nodes, liver, gastrointestinal tract and bone marrow (BM) where they are activated and release mediators such as histamine, tryptase and cytokines that then cause organ damage [1–3]. The prevalence of ASM is 0.09 per 100,000, and the median age at diagnosis is over 60 years. ASM is quite rare in pediatric population [4–6]. The D816V mutation in *KIT* is frequently found in the tumor cells of SM patients and is an important part of the established diagnostic criteria for ASM. In addition to *KIT* variations, somatic mutations in other genes also occur in ASM that facilitate tumor growth [2, 7].

Ovarian dysgerminoma is one of the common malignant germ cell tumors believed to develop from primordial germ cells (PGCs) due to its morphology. These malignant tumors more frequently occur in adolescents and young adults, and surgical resection accompanied

* Correspondence: kura@fujita-hu.ac.jp
†Makiko Tsutsumi and Hiroki Miura contributed equally to this work.
[1]Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, 1-98 Dengakugakubo, Kutsukake-cho, Toyoake, Aichi 470-1192, Japan
Full list of author information is available at the end of the article

Tsutsumi *et al. BMC Cancer*   (2020) 20:1162

Page 2 of 7

with chemotherapy generally result in a good prognosis [8, 9].

In our present case report, we describe an adolescent case of ASM in a female with a previous history of ovarian dysgerminoma. Genetic analysis indicated a common origin for these malignancies and provided insights into the processes underlying the progression to ASM.

Methods

Samples for genetic analyses

Genomic DNA was isolated from peripheral blood (PB) and BM samples of the study patient using the QuickGene DNA whole blood DNA kit L (Kurabo, Osaka, Japan). Genomic DNA of the buccal mucosa was extracted using the DNeasy Blood and Tissue kit (Qiagen, Tokyo, Japan). A formalin-fixed paraffin-embedded specimen of the surgically dissected ovarian dysgerminoma was deparaffinized in xylene followed by proteinase treatment, phenol/chloroform extraction and ethanol precipitation of DNA. Conventional G-banding of the patient's bone marrow was performed using a standard method.

PCR amplification and sequencing

DNA fragments were amplified by PCR using Ex-Taq or LA-Taq polymerase (Takara, Kusatsu, Japan) followed by direct sequencing with the primers listed in Table S1 of the Additional file 1. Where indicated, PCR products were cloned into the pT7 Blue T-vector (Novagen, Madison, WI, USA) and then sequenced.

Real-time quantitative PCR

Quantitative PCR of the *KIT* gene was performed on the StepOnePlus Real-Time PCR system (Thermo Fisher Scientific, Waltham, MA, USA) using the PowerUp SYBR Green Master Mix (Thermo Fisher Scientific) with the primers listed in Table S1 of the Additional file 1. The *DROSHA* gene was used as an internal control.

Whole-exome sequencing

Whole-exome sequencing of the PB and BM specimens was performed as described previously [10]. The sequencing data were analyzed with Variant Studio 2.3 (Illumina, San Diego, CA, USA), Integrative Genomics Viewer ver.2.4.19 (Broad Institute, Cambridge, MA, USA) and Mutect2 software (Broad Institute). The list of known cancer genes in the Cancer Gene Census [11] was used to identify mutations in the study patient.

Cytogenetic microarray

Whole genomic microarray analysis of the BM sample was performed using the CytoScan 750 K array (Affymetrix, Santa Clara, CA) and analyzed using R package Rawcopy [12].
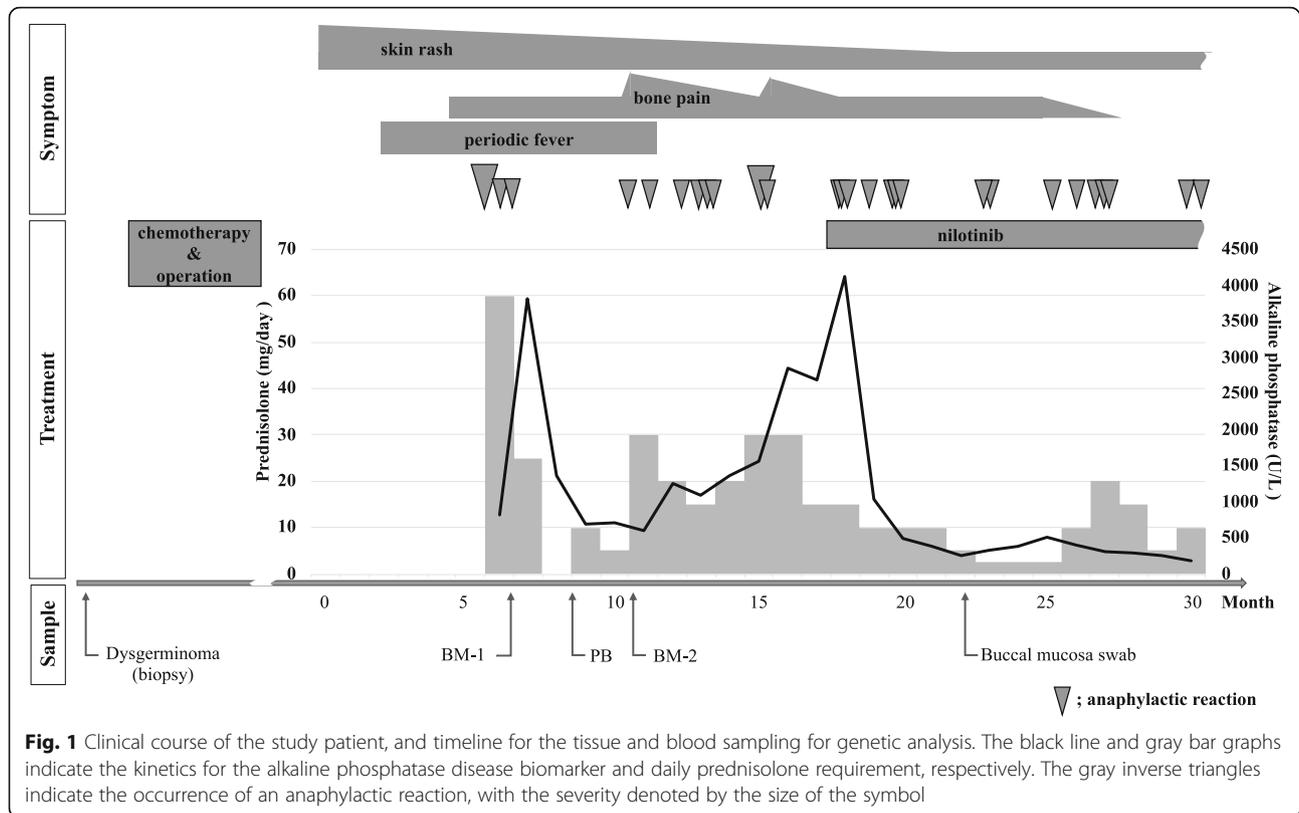
Results

Case presentation

The study case was affected with ovarian dysgerminoma when she was 13 years old. She received 4 cycles of chemotherapy, consisting of bleomycin, etoposide and cisplatin, followed by complete surgical resection of the tumor. After this resection, the patient's α-fetoprotein (AFP) level fell from 1053 ng/ml (normal range, 0–10.5 ng/ml) to normal levels and she received 2 cycles of postoperative chemotherapy consisting of carboplatin and etoposide.

She had been in complete remission for 2 years but developed recurring episodes of skin rash, bone pain, periodic fever and anaphylactic reactions when she was 16 years old (Fig. 1). A computed tomography scan suggested skeletal involvement with osteosclerosis mainly affecting the spine and osteolysis in a limited area of bones, but no tumor mass was observed suggesting that ovarian dysgerminoma recurrence was unlikely. BM examination of the osteolytic lesions revealed multifocal, dense infiltrates of mast cells that showed positive immunohistochemical staining for mast cell tryptase, CD25, CD33, and c-KIT, but no dysgerminoma cell was observed. A mutation at codon 816 of the KIT gene was further revealed in these cells, as detailed later. In addition, the serum tryptase level was markedly elevated at 276 μg/L (normal range, 1–15 μg/L), but AFP level was normal. The patient was subsequently diagnosed with ASM in accordance with the 2016 WHO classification of mastocytosis. The initial therapeutic intervention, including prednisolone, and histamine H1- and H2- receptor antagonists, improved her general condition, but the frequency of anaphylaxis did not decrease significantly and she became steroid dependence. Furthermore, her serum alkaline phosphatase level, which is indicative of disease activity, temporarily decreased and then rose again from 595 to 2857 U/L (normal range, 115–359 U/L). Mast cell accumulation in the BM was reevaluated before applying a second-line therapy. At that time, the proportion of the mast cells identified by c-KIT staining had decreased to 5% of all nucleated cells. The karyotype of the BM was normal (46,XX). After introducing nilotinib, a potent tyrosine kinase inhibitor for patients harboring a *KIT* codon 816 mutation, her clinical symptoms that had been occurring on an almost daily basis improved moderately and the prednisolone dose could therefore be reduced. Her serum alkaline phosphatase level was also normalized. However, she still suffered with anaphylactic reactions almost once per month, which suggested the existence of uncontrolled residual and reactive mast cells.

Detection of a KIT mutation

To make a definitive diagnosis of ASM in our study patient, the codon 816 region of the *KIT* gene was

Fig. 1 Clinical course of the study patient, and timeline for the tissue and blood sampling for genetic analysis. The black line and gray bar graphs indicate the kinetics for the alkaline phosphatase disease biomarker and daily prednisolone requirement, respectively. The gray inverse triangles indicate the occurrence of an anaphylactic reaction, with the severity denoted by the size of the symbol
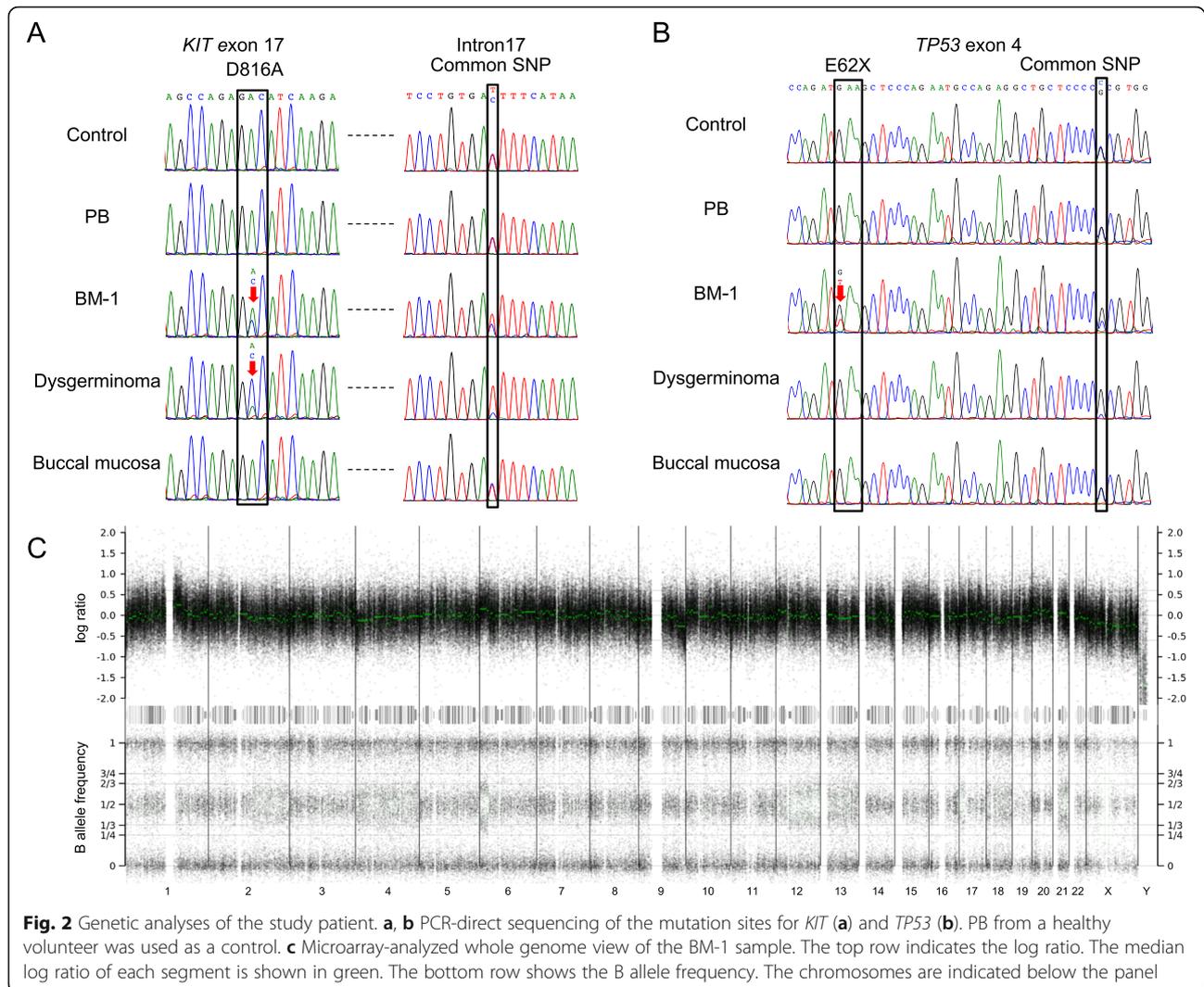
amplified by PCR and directly sequenced (Fig. 2a; Additional file 2: Fig. S1A). A somatic mutation, D816A (NM_000222.2:c.2447A > C, COSM24675) was identified in the patient's BM that had been sampled at the onset of the ASM-associated symptoms (BM-1). This mutation was not detectable in the PB, nor in the BM obtained after the initial treatment (BM-2). The percentage of mast cells in the BM-1 and BM-2 smears determined by microscopic examination was 50 and 5%, respectively. The variant allele frequency (VAF) in the BM-1 sample was 0.34, which may have been due to contamination by normal cells.

D816 mutations in the *KIT* gene are common to germ cell tumors [13] and we thus tested a biopsy specimen that was obtained from the dysgerminoma lesion in our study case prior to chemotherapy. The same D816A *KIT* mutation was detected. Notably also, the mutant allele ratio in the dysgerminoma was considerably higher than that of the normal allele (VAF = 0.75). *KIT* gene amplification was not evident by quantitative PCR (Additional file 2: Fig. S1B), suggesting that the normal allele had been deleted in the dysgerminoma cells. Indeed, the ratio between the two alleles of a heterozygous common single nucleotide polymorphism (SNP), C > T (rs1008658), located 115 bp downstream of the mutation site, was also skewed in samples carrying the D816A variant. PCR products including both the *KIT* D816 and rs1008658 sites were

cloned and sequenced individually (Additional file 2: Fig. S1C). The *KIT* D816A mutation was found to be linked to the T allele of the SNP. Since the VAF in the BM-1 sample was 0.64, the normal allele was thought be also lost in the ASM cells.

Genome-wide analysis of genetic alterations

We next explored the additional mutations associated with ASM development in our patient as ASM cells often carry mutations in cancer-related genes in addition to *KIT* [2]. Whole-exome sequencing of the genomic DNA from BM-1 and PB samples was performed and revealed a nonsense mutation in *TP53* and a missense mutation in *TET2* (Additional file 1: Tables S2 and S3). The nonsense mutation in *TP53*, E62X, (NM_000546: c.184G > T) was detected in BM-1, but not in the dysgerminoma nor the other samples (Fig. 2b). Normal cell contamination prevented us from determining the status of the normal allele (VAF = 0.31). When we examined for the presence of a common heterozygous SNP (rs1042522), located 31 bp downstream of the mutation site, an allelic imbalance was found in the BM-1 cells (Fig. 2b). Next-generation sequencing reads demonstrated that the mutant allele was present in the G allele of the SNP (Additional file 3: Fig. S2). From this phasing data, loss of heterozygosity (LOH) at this locus was evident in the BM-1 sample, suggesting a biallelic

Tsutsumi *et al. BMC Cancer* (2020) 20:1162

Page 4 of 7

Fig. 2 Genetic analyses of the study patient. **a**, **b** PCR-direct sequencing of the mutation sites for *KIT* (**a**) and *TP53* (**b**). PB from a healthy volunteer was used as a control. **c** Microarray-analyzed whole genome view of the BM-1 sample. The top row indicates the log ratio. The median log ratio of each segment is shown in green. The bottom row shows the B allele frequency. The chromosomes are indicated below the panel

inactivation of the *TP53* gene in the ASM cells. To next determine whether the normal allele of the *TP53* gene had been deleted, its copy number was analyzed by whole-genomic microarray. LOH spanning the 17p13.1 region that incorporates the *TP53* locus was demonstrated in the BM-1 cell genomes, although the copy number of this region was normal (Fig. 2c). This copy-number-neutral LOH (CN-LOH) suggested that second hit was not the deletion, but that both alleles of the *TP53* gene in the ASM cells carried an E62X mutation, possibly due to mitotic recombination. Surprisingly, an LOH of the rs1042522 SNP was also found in the dysgerminoma cells without the E62X mutation (Fig. 2b).

The V1846F missense variant of *TET2* (NM_001127208.3:c.5536G > T) was found from our analysis of a buccal mucosal sample from the current study patient to be a germline variant (Additional file 4: Fig. S3). In silico analysis further predicted this to be a deleterious variation (PolyPhen-2 = 0.852, SIFT = 0.034) that did not appear in

the databases (Additional file 1: Table S3). Notably, *TET2* is one of the tumor suppressor genes associated with ASM [14–16]. Both *TET2* and *KIT* are located on the chromosome 4 and we found by microarray analysis that the copy number of chromosome 4 was decreased in BM-1 (Fig. 2c). To then investigate whether the inactivation of the *TET2* gene is involved in development of ASM, we examined the phase of the *TET2* variant and the rs1008658 SNP. According to the determined allele frequencies, the *TET2* mutation was linked to the chromosome harboring the C allele of rs1008658 and was lost in the tumor cells. We thus concluded that the *TET2* variant was not associated with the development of ASM, since this rare mutation had been lost in the ASM cells. Further to this, the presence of this rare *TET2* variant in the healthy father of our patient was consistent with it having a benign nature (Additional file 4: Fig. S3).

In addition to chromosomes 4 and 17, copy-number aberrations or LOHs were detected in chromosomes 1,
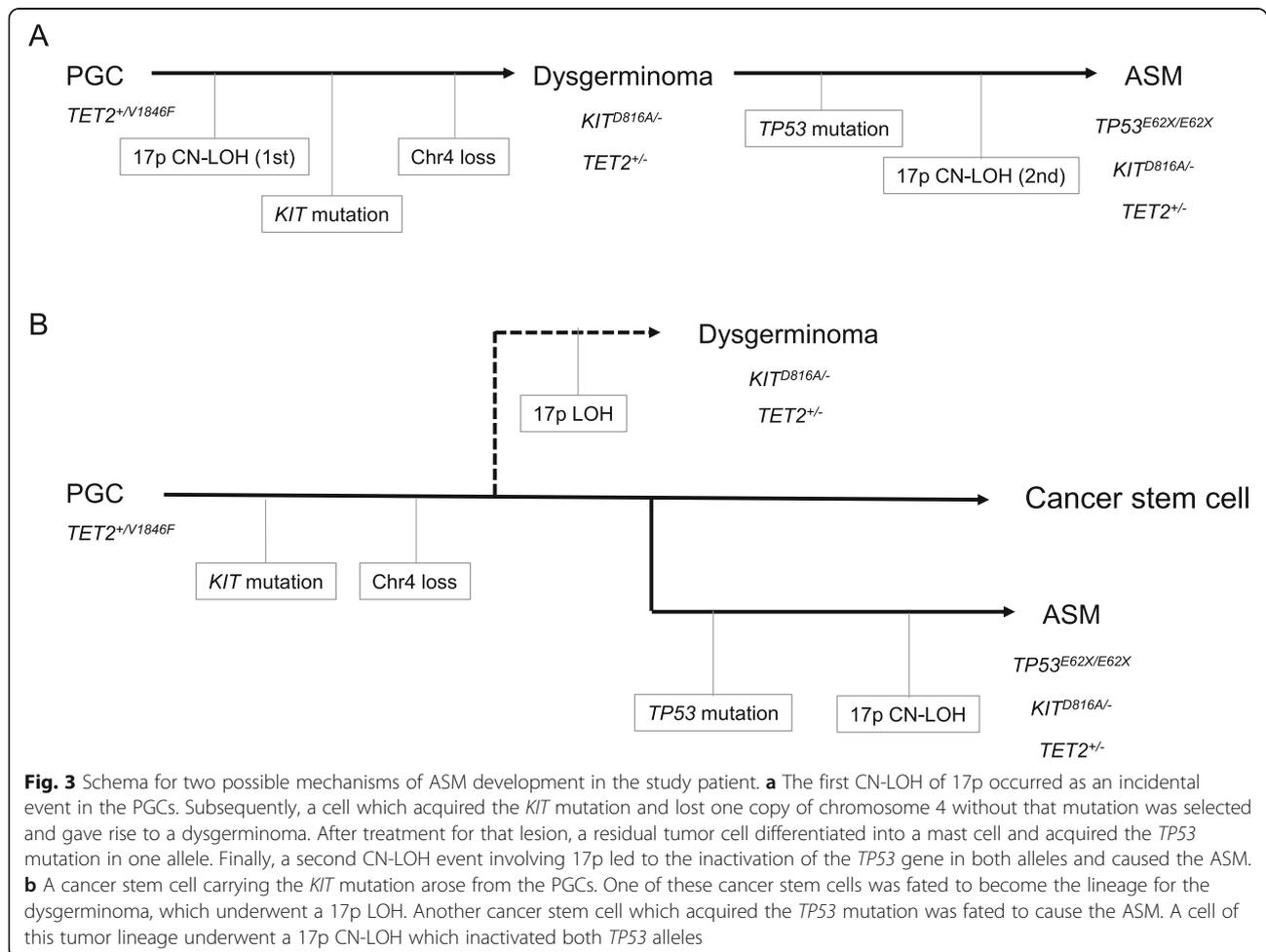
2, 6, 9, 12, 13, 14, 16, 18, 20, 21 and X in the BM-1 cells from our patient (Fig. 2c). No other mutations with the potential to be cancer driver candidates were detected by whole-exome sequencing of the affected regions of these chromosomes (Additional file 1: Table S3).

Discussion

We have here presented an adolescent case of ASM which is a rare neoplasm at such a young age. Our female patient developed the ASM after prior treatment for a dysgerminoma. Significantly, the pathogenic *KIT* gene variant, D816A, was identified in both the dysgerminoma and BM samples in this case. *KIT* mutations are frequently found in the tumor cells of SM patients and are an important part of the established diagnostic criteria for ASM, but the most prevalent variant of these is the D816V mutation, which has been observed in > 60% of ASM patients [7]. The D816A mutation has been occasionally identified also in SM with an associated hematologic neoplasm [17–19] but not in ASM. *KIT* D816 mutations are also commonly observed in germ cell tumors, and have been found in one-third of ovarian

dysgerminomas [13]. Notably however, the D816A mutation has not been reported previously in these tumors. Thus, the presence of the *KIT* D816A variant in both the ASM and dysgerminoma cells in our present patient is unlikely to be a coincidence. There are considerable number of reports describing the association between mastocytosis and germ cell tumor [20–24]. The etiological linkage between ASM and a preceding germ cell tumor caused by a *KIT* D816 mutation is further supported by a similar prior report of an ASM patient carrying *KIT* D816V who had previously had an ovarian germ cell tumor harboring this same mutation [25]. Another similar case of an ovarian germ cell tumor carrying a *KIT* D816H mutation was also recently reported in which the chemotherapy was complicated by the development of SM with chronic myelomonocytic leukemia harboring this same mutation [26].

One simple hypothesis to explain our current findings is clonal evolution (Fig. 3a). In brief, the *KIT* D816A variation initially induces the development of ovarian dysgerminoma. Although the tumor was removed by surgical resection, minimally residual cells that

Fig. 3 Schema for two possible mechanisms of ASM development in the study patient. **a** The first CN-LOH of 17p occurred as an incidental event in the PGCs. Subsequently, a cell which acquired the *KIT* mutation and lost one copy of chromosome 4 without that mutation was selected and gave rise to a dysgerminoma. After treatment for that lesion, a residual tumor cell differentiated into a mast cell and acquired the *TP53* mutation in one allele. Finally, a second CN-LOH event involving 17p led to the inactivation of the *TP53* gene in both alleles and caused the ASM. **b** A cancer stem cell carrying the *KIT* mutation arose from the PGCs. One of these cancer stem cells was fated to become the lineage for the dysgerminoma, which underwent a 17p LOH. Another cancer stem cell which acquired the *TP53* mutation was fated to cause the ASM. A cell of this tumor lineage underwent a 17p CN-LOH which inactivated both *TP53* alleles

differentiated into hematologic cells possibly underwent subsequent biallelic mutation of the *TP53* gene, resulting in malignant transformation. Our microarray analysis indicated a complex karyotype and suggested the mast cells underwent repeated genetic rearrangements, although an abnormal karyotype is less common in ASM [5, 27, 28]. *TP53* is a less frequently affected gene in advanced SM [3], but is the most commonly affected gene in therapy-related myeloid neoplasms [29, 30]. The chemotherapy for the previous dysgerminoma in our present patient may have contributed to the mutation induction in a residual tumor cell and then expansion of these mutated mast cells, resulting in the younger onset of ASM in this case.

It is noteworthy that the LOH at the *TP53* locus was present both in the dysgerminoma and ASM, but that the ASM had the CN-LOH. The ASM cells were found to carry a biallelic homozygous E62X mutation of the *TP53* gene, suggesting that the CN-LOH was established by mitotic recombination after acquisition of the *TP53* mutation. However, the dysgerminoma was found to carry a 17p LOH without *TP53* mutation, suggesting that the 17p LOH in the dysgerminoma and ASM were independent events. A possible alternative hypothesis therefore is that the dysgerminoma and ASM originated from a common cancer stem cell harboring the *KIT* D816A mutation and developed independently (Fig. 3b). Briefly, the deletion of 17p possibly occurred in the dysgerminoma lineage. After treatment for this lesion, another cancer stem cell acquired the *TP53* mutation and underwent chromosome rearrangements, which evolved into the origin of another tumor lineage and caused the ASM.

Previously reported experiments using murine cells expressing pathogenic *KIT* mutants have demonstrated that the loss of one copy of *Tet2* accelerates mast cell growth [15, 31]. Although the effect of the *TET2* V1846F variation in tumorigenesis is not clear, cells with only one normal *TET2* allele may have a growth advantage under KIT D816A expression, even after deletion of the *TET2* V1846F allele in our current patient, suggesting that the *TET2* V1846F variant was unlikely to have affected the tumor progression.

Our current genetic analyses were useful in not only reaching a conclusive diagnosis but also for determining the treatment strategies in our study patient. Although the patient's condition had not progressed during the current treatment and the genetic alternations in her bone marrow were not among the negative prognostic indicators reported previously [32], the comprehensive cancer-related genetic alterations that were detected in this case in addition to the *KIT* mutation suggested that a tyrosine kinase inhibitor would not be sufficient to achieve remission. Notably in this regard, the multikinase inhibitor midostaurin is not currently available in Japan. An allogenic bone marrow transplantation from an HLA 1-locus mismatched sibling donor is being planned for our study patient.

Conclusions

Our current cytogenomic analyses demonstrated that mast cells causing the ASM shared a common origin with the dysgerminoma, suggesting that a careful follow-up should be required for the dysgerminoma patients after treatment.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-020-07653-z.

Additional file 1: Table S1. PCR primers used in this study. **Table S2.** BM-1-specific candidate variants of cancer-related genes analyzed by Mutect2. **Table S3.** Candidate variants of cancer-related genes identified in the PB of the study patient.

Additional file 2: Figure S1. Genetic analyses of the study patient. (A) PCR-direct sequencing of the *KIT* mutation site in the BM-2 sample. (B) Real-time quantitative PCR of the *KIT* gene. The *SNX25* gene is located on the 4q35.1. Ratio to the healthy control (mean ± SD) obtained from the two independent experiments are shown. (C) Cloning and sequence analysis of the PCR products from the BM-1.

Additional file 3: Figure S2. Integrative Genomics Viewer image of the next-generation sequencing reads of the *TP53* mutation site. The arrow indicates the transcriptional direction of the *TP53* gene. The 'A' at the E62X site and 'C' in the common SNP are the complementary bases of 'T' and 'G' in the context of gene coding, respectively.

Additional file 4: Figure S3. PCR-direct sequencing of the *TET2* mutation site in the PB, BM-1, dysgerminoma, and buccal mucosa samples from the patient, and PB samples from her parents and a healthy control.

Abbreviations
AFP: α-fetoprotein; ASM: Aggressive systemic mastocytosis; BM: Bone marrow; CN-LOH: Copy-number-neutral loss of heterozygosity; LOH: Loss of heterozygosity; PB: Peripheral blood; PGC: Primordial germ cells; SM: Systemic mastocytosis; SNP: Single nucleotide polymorphism; VAF: Variant allele frequency

Tsutsumi *et al. BMC Cancer*   (2020) 20:1162

Page 7 of 7

Author details
[1]Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, 1-98 Dengakugakubo, Kutsukake-cho, Toyoake, Aichi 470-1192, Japan. [2]Department of Pediatrics, Fujita Health University School of Medicine, Toyoake, Japan. [3]ITOCHU Collaborative Research-Molecular Targeted Cancer Treatment for Next Generation, Graduate School of Medicine, Nagoya University, Nagoya, Japan. [4]Department of Virology and Liver Unit, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan.

References
1. Theoharides TC, Valent P, Akin C. Mast cells, Mastocytosis, and related disorders. N Engl J Med. 2015;373:163–72.
2. Valent P, Akin C, Hartmann K, Nilsson G, Reiter A, Hermine O, et al. Advances in the classification and treatment of Mastocytosis: current status and outlook toward the future. Cancer Res. 2017;77:1261–70.
3. Leguit R, Hebeda K, Kremer M, van der Walt J, Gianelli U, Tzankov A, et al. The Spectrum of aggressive Mastocytosis: a workshop report and literature review. Pathobiology. 2020;87:2–19.
4. Cohen SS, Skovbo S, Vestergaard H, Kristensen T, Møller M, Bindslev-Jensen C, et al. Epidemiology of systemic mastocytosis in Denmark. Br J Haematol. 2014;166:521–8.
5. Pieri L, Bonadonna P, Elena C, Papayannidis C, Grifoni FI, Rondoni M, et al. Clinical presentation and management practice of systemic mastocytosis. A survey on 460 Italian patients. Am J Hematol. 2016;91:692–9.
6. Méni C, Bruneau J, Georgin-Lavialle S, Le Saché de Peufeilhoux L, Damaj G, Hadj-Rabia S, et al. Paediatric mastocytosis: a systematic review of 1747 cases. Br J Dermatol. 2015;172:642–51.
7. Arock M, Sotlar K, Akin C, Broesby-Olsen S, Hoermann G, Escribano L, et al. KIT mutation analysis in mast cell neoplasms: recommendations of the European competence network on Mastocytosis. Leukemia. 2015;29:1223–32.
8. Kraggerud SM, Hoei-Hansen CE, Alagaratnam S, Skotheim RI, Abeler VM, Rajpert-De Meyts E, et al. Molecular characteristics of malignant ovarian germ cell tumors and comparison with testicular counterparts: implications for pathogenesis. Endocr Rev. 2013;34:339–76.
9. Shaaban AM, Rezvani M, Elsayes KM, Baskin H Jr, Mourad A, Foster BR, et al. Ovarian malignant germ cell tumors: cellular classification and clinical and imaging features. Radiographics. 2014;34:777–801.
10. Boda H, Uchida H, Takaiso N, Ouchi Y, Fujita N, Kuno A, et al. A PDE3A mutation in familial hypertension and brachydactyly syndrome. J Hum Genet. 2016;61:701–3.
11. Cancer Gene Census. https://cancer.sanger.ac.uk/census. Accessed 1 Oct 2019.
12. Mayrhofer M, Viklund B, Isaksson A. Rawcopy: improved copy number analysis with Affymetrix arrays. Sci Rep. 2016;6:36158.
13. Cheng L, Roth LM, Zhang S, Wang M, Morton MJ, Zheng W, et al. KIT gene mutation and amplification in dysgerminoma of the ovary. Cancer. 2011; 117:2096–103.
14. Tefferi A, Levine RL, Lim KH, Abdel-Wahab O, Lasho TL, Patel J, et al. Frequent TET2 mutations in systemic mastocytosis: clinical, KITD816V and FIP1L1-PDGFRA correlates. Leukemia. 2009;23:900–4.
15. Soucie E, Hanssens K, Mercher T, Georgin-Lavialle S, Damaj G, Livideanu C, et al. In aggressive forms of mastocytosis, TET2 loss cooperates with c-KITD816V to transform mast cells. Blood. 2012;120:4846–9.
16. Schwaab J, Schnittger S, Sotlar K, Walz C, Fabarius A, Pfirrmann M, et al. Comprehensive mutational profiling in advanced systemic mastocytosis. Blood. 2013;122:2460–6.
17. Cornet E, Dumézy F, Roumier C, Lepelley P, Jouy N, Philippe N, et al. Involvement of a common progenitor cell in core binding factor acute myeloid leukaemia associated with mastocytosis. Leuk Res. 2012;36:1330–3.
18. Yabe M, Masukawa A, Kato S, Yabe H, Nakamura N, Matsushita H. Systemic mastocytosis associated with t(8;21) acute myeloid leukemia in a child: detection of the D816A mutation of KIT. Pediatr Blood Cancer. 2012;59: 1313–6.
19. Frederiksen JK, Shao L, Bixby DL, Ross CW. Shared clonal cytogenetic abnormalities in aberrant mast cells and leukemic myeloid blasts detected by single nucleotide polymorphism microarray-based whole-genome scanning. Genes Chromosomes Cancer. 2016;55:389–96.
20. Chariot P, Monnet I, LeLong F, Chleq C, Droz JP, de Cremoux H. Systemic mast cell disease associated with primary mediastinal germ cell tumor. Am J Med. 1991;90:381–5.
21. Chariot P, Monnet I, Gaulard P, Abd-Alsamad I, Ruffié P, De Cremoux H. Systemic mastocytosis following mediastinal germ cell tumor: an association confirmed. Hum Pathol. 1993;24:111–2.
22. Delacrétaz F, Stalder M, Meugé-Moraw C, Schmidt PM, Joris F, Kurt AM, et al. Systemic mastocytosis following a malignant ovarian germ cell tumour. Histopathology. 1997;30:582–4.
23. Teitell M, Rowland JM. Systemic mast cell disease associated with primary ovarian mixed malignant germ cell tumor. Hum Pathol. 1998;29:1546–7.
24. Miyagawa S, Hirota S, Park YD, Yamasaki M, Daikoku N, Morikawa H, et al. Cutaneous mastocytosis associated with a mixed germ cell tumour of the ovary: report of a case and review of the literature. Br J Dermatol. 2001;145: 309–12.
25. Lee JW, Yang WS, Chung SY, Kang JH, Cho B, Kim HK, et al. Aggressive systemic mastocytosis after germ cell tumor of the ovary: C-KIT mutation documentation in both disease states. J Pediatr Hematol Oncol. 2007;29: 412–5.
26. Mitchell SG, Bunting ST, Saxe D, Olson T, Keller FG. A variant c-KIT mutation, D816H, fundamental to the sequential development of an ovarian mixed germ cell tumor and systemic mastocytosis with chronic myelomonocytic leukemia. Pediatr Blood Cancer. 2017;64:e26282.
27. Naumann N, Jawhar M, Schwaab J, Kluger S, Lübke J, Metzgeroth G, et al. Incidence and prognostic impact of cytogenetic aberrations in patients with systemic mastocytosis. Genes Chromosomes Cancer. 2018;57:252–9.
28. Shah S, Pardanani A, Elala YC, Lasho TL, Patnaik MM, Reichard KK, et al. Cytogenetic abnormalities in systemic mastocytosis: WHO subcategory-specific incidence and prognostic impact among 348 informative cases. Am J Hematol. 2018;93:1461–6.
29. Ok CY, Patel KP, Garcia-Manero G, Routbort MJ, Peng J, Tang G, et al. TP53 mutation characteristics in therapy-related myelodysplastic syndromes and acute myeloid leukemia is similar to de novo diseases. J Hematol Oncol. 2015;8:45.
30. Chung J, Sallman DA, Padron E. TP53 and therapy-related myeloid neoplasms. Best Pract Res Clin Haematol. 2019;32:98–103.
31. De Vita S, Schneider RK, Garcia M, Wood J, Gavillet M, Ebert BL, et al. Loss of function of TET2 cooperates with constitutively active KIT in murine and human models of mastocytosis. PLoS One. 2014;9:e96209.
32. Rossignol J, Polivka L, Maouche-Chrétien L, Frenzel L, Dubreuil P, Hermine O. Recent advances in the understanding and therapeutic management of mastocytosis. F1000Res. 2019;8. https://doi.org/10.12688/f1000research.19463.1.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

ORIGINAL ARTICLE

Molecular analysis of low-level mosaicism of the *IKBKG* mutation using the X Chromosome Inactivation pattern in Incontinentia Pigmenti

Miki Kawai[1,2] | **Takema Kato**[1] | **Makiko Tsutsumi**[1] |
Yasuko Shinkai[1] | **Hidehito Inagaki**[1] | **Hiroki Kurahashi**[1,2] (ORCID)

[1]Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Japan

[2]Department of Clinical Genetics, Fujita Health University Hospital, Toyoake, Japan

Correspondence
Hiroki Kurahashi, Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, 1-98 Dengakugakubo, Kutsukake-cho, Toyoake, Aichi 470-1192, Japan.
Email: kura@fujita-hu.ac.jp

Funding information
This work was supported by grants-in-aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan (15H04710); from the Ministry of Health, Welfare and Labor (H27-Nanchitou-Ippan[nan]-024); and from the Japan Agency for Medical Research and Development

Abstract

Background: Incontinentia pigmenti (IP) is a rare X-linked disorder affecting the skin and other ectodermal tissues that is caused by mutation of the *IKBKG/NEMO* gene. Previous studies have reported that the overall mutation detection rate in IP is ~75%. We hypothesized that a low-level mosaicism existed in the remaining cases.

Methods: Genomic variations in the *IKBKG* gene were examined in 30 IP probands and their family members. Standard mutational analyses were performed to detect common deletions, nucleotide alterations, and copy number variations. To assess skewing of the X chromosome inactivation (XCI) pattern, a HUMARA assay was performed. We compared the results of this analysis with phenotype severity.

Results: Pathogenic variants were identified in 20 probands (66.7%), the rate of detection was suboptimal. The remaining 10 probands tended to manifest a mild phenotype with no skewed X chromosome inactivation that is generally observed in IP patients. Quantitative nested PCR and digital droplet PCR were performed for the 10 patients and mosaicism of the common *IKBKG* deletion were identified in five patients.

Conclusion: Overall, we detected 25 *IKBKG* mutations (83.3%). Determination of the XCI value in advance of mutational analyses for IP could improve the mutation detection rate. Our improved detection rate for these mutations, particularly those with a low-level mosaicism, may present opportunities for appropriate genetic counseling.

KEYWORDS
IKBKG, incontinentia pigmenti, mosaicism, X chromosome inactivation

1 | INTRODUCTION

Incontinentia pigmenti (IP; MIM #308300) is a rare X-linked genodermatosis with an estimated prevalence at birth of 1.2/100,000 (Orphanet, 2018). This disorder affects the skin and other ectodermal tissues including the eyes, teeth, hair, nail, and central nervous system (CNS). Skin lesions are the first diagnostic manifestations of IP

that appear in the neonatal period and comprise a vesicu-bullous stage (stage I) followed by a verrucous stage (stage II) persisting for months to years, a hyperpigmented stage (stage III), and finally a hypopigmented stage (stage IV) usually continuing throughout life. These skin defects follow Blaschko lines and present in almost all IP patients, thus constituting the principal diagnostic IP criteria for this condition (Landy & Donnai, 1993; Minić et al., 2014). Systemic involvement includes ocular and neurologic impairment. Although IP is X-linked dominant and is usually lethal in males, some affected males have been reported, representing approximately 10% (Fusco et al., 2014) of the patient population. Affected males present with somatic mosaicism and in some reported cases a concomitant diagnosis of Klinefelter syndrome (Conte et al., 2014).

IP is caused by a mutation of the *IKBKG/NEMO* gene (Inhibitor of Kappa polypeptide gene enhancer in B-cells, Kinase Gamma/Nuclear Factor κB, Essential Modulator, GenBank NM_003639.3, OMIM#300248). This gene encodes NEMO/IKKγ which is required for the activation of the nuclear factor-kappa B (NF-κB) transcription factor. NEMO/IKKγ acts as a regulatory subunit of the inhibitor of the κB (IκB) kinase (IKK) complex. The absence of NEMO/IKKγ protein renders the cells sensitive to apoptosis, leading to lethality in males (Aradhya, et al., 2001) and selective skewed X-inactivation in females (Fusco et al., 2004). In most patients, IP onset is due to loss-of-function (LoF) mutations, although in some case *IKBKG* hypomorphic mutations have been reported (Fusco et al., 2008) that reduce but do not eliminate NF-κB activation, thus explaining why some affected male patients survive (Döffinger et al., 2001). Female patients carrying these hypomorphic mutations show mild signs of IP (Aradhya, et al., 2001).

IKBKG is a 23-kb gene composed of nine coding exons, four alternative noncoding first exons (from exons 1A to 1D), and two promoters. The *IKBKG* gene also shares a part of a 35.7 kb segmental duplication arranged in the opposite direction, one covering the genuine gene and the other a part of a pseudogene copy. The non-functional *IKBKGP* spans the region between exon 3 to 10. The most frequent IP mutation is a recurrent deletion produced by non-allelic homologous recombination (NAHR) due to a misalignment between approximately two 650 bp short interspersed nuclear element (SINE) of a medium reiterated 67B (MER67B), one of which is located in intron 3 and the other distally to exon 10. This deletion removes exons 4-10, spanning 11.7 kb. In addition, the NAHR mechanism can also generate benign copy number variations (CNVs) through an exon 4-10 deletion in *IKBKGP* or an exon 4-10 duplication in the *IKBKG* gene. However, both benign CNVs are risk alleles for de novo generation of a pathological *IKBKG* exon 4-10 deletion (Fusco et al., 2009). The

IKBKG region contains many repeat sequences and is considered to be prone to rearrangement.

Previous studies have reported that the overall mutation detection rate in IP is 77.6%. *IKBKG* gene targeted analysis of the exon 4-10 deletion can detect 65% of affected patients, single-nucleotide variant (SNV) analysis can detect 8.6%, and large duplication and deletion analysis can detect 4% of clinically affected individuals (Fusco et al., 2008; Fusco et al., 2007; Fusco et al., 2012). As *IKBKG* is the only gene responsible for IP, these rates are low in comparison with other genetic diseases, even when taking into consideration that the *IKBKG* locus has a complex genomic structure. In the course of the mutational analyses of our own IP patients, we observed that patients with no detectable mutation showed a mild phenotype. We hypothesized that a low-level mosaicism existed in these cases which would have made it difficult to detect the mutations in blood samples.

In this study, we conducted *IKBKG* gene mutational analysis in 30 IP patients. We compared the results of this analysis with phenotype severity and also with the X chromosome inactivation (XCI) pattern that would not appear skewed in patients with mosaicism due to the presence of normal cells presenting with a random XCI pattern. We describe observed relationships between phenotype severity, XCI pattern and somatic mosaicism. We also discuss the clinical implications of our results for future genetic testing and counseling.

2 | MATERIALS AND METHODS

2.1 | DNA samples

All research was performed in accordance with the principles of the Declaration of Helsinki, and the Ethical Guidelines for Human Genome/Gene Analysis Research by the Ministry of Education, Culture, Science, and Technology, the Ministry of Health, Labor, and Welfare, and the Ministry of Economy, Trade, and Industry of Japan.

Peripheral blood samples were obtained from 30 Japanese IP patients (28 females) in accordance with the research protocol approved by the local ethics committee of Fujita Health University. The IP patients enrolled in our study were selected on the basis of the previous Landy and Donnai (Landy & Donnai, 1993) diagnostic criteria, and also because they met the recently updated criteria (Minić et al., 2014). The severity of disease in each patient was evaluated using manifestation score of skin leision, nervous system defect, ocular system defect, dental system defect, hair defect, and nail defect (Table S1). We collected DNA samples from both parents in 10 cases and a single parent in six cases. Written informed consent to participate in the study was obtained from the patients or their parents. Genomic DNA from peripheral

blood was extracted using a conventional salt precipitation technique.

2.2 | Analysis of common Incontinentia Pigmenti deletions

Common IP deletions have previously been characterized (Bardaro et al., 2003). Long-range PCR was performed in our current analysis to detect the specific *IKBKG* exon 4-10 deletion using the primers listed in Table S3. The PCR reactions were performed in 25 µl volumes containing 1 µl of sample, 2.5 U of LA Taq Polymerase (Takara Bio Inc.), LA Taq buffer II, 200 µM dNTPs, 2 mM $MgCl_2$ and 10 pmol of each primer. The amplification protocol comprised 30 cycles of 10 seconds at 98°C and 3 min at 65°C. To detect *IKBKGP*, PCR was performed as described above with the previously reported Rev-2 and JF3R primers (Bardaro et al., 2003).

Nested PCR was performed to detect low-level mosaic deletions. The primers used for these second amplifications are also listed in Table S3. The resulting PCR products were diluted 1:100 with Tris-EDTA Buffer (TE). These nested reactions contained 1 µl of template DNA in an identical reaction mixture for 20 cycles of 10 seconds at 98°C and 1 min at 60°C. PCR products were visualized on a 2% agarose gel, quantified using an image analyzer, and compared with the products amplified from serial dilutions of genomic DNA from patients harboring the common *IKBKG* deletion.

To detect mosaic deletions, we performed droplet digital PCR using the custom probes listed in Table S4. Droplet generation, PCR cycling, and droplet reading were performed in accordance with the manufacturer's protocol (Bio-Rad). The first PCR was performed over 13 cycles with 10 ng of DNA sample as the template. The resulting PCR products were diluted 1:10 with Tris-EDTA Buffer. Probes and primers were mixed with a 2x ddPCR supermix and with 1 µl of template DNA. In total, 22 µl volume reactions were loaded into an 8-channel droplet generator cartridge (Bio-Rad) and droplets were then generated with 70 µl of droplet generation oil (Bio-Rad) using a manual QX200 Droplet Generator. Following droplet generation, samples were transferred to a 96-well PCR plate, heat sealed and then amplified on a thermal cycler using the following cycling conditions: 95℃ for 10 min followed by 40 cycles at 95℃ for 30 seconds and 60℃ for 10 min, one cycle at 98℃ for 10 min, and maintenance at 12℃. Post-PCR products were analyzed on the QX200 droplet reader (Bio-Rad) using QuantaSoft software.

2.3 | Analysis of structural variants

Copy number variations were analyzed at the IP locus, including the *IKBKG* gene and *IKBKGP*, by applying the SALSA

MLPA probemix P073-A1 (MRC-Holland). Test fragments have been designed previously to evaluate the *IKBKG* gene and the sizes and locations of most of these test probes as well as reference probes have been defined. MLPA and capillary electrophoresis-based amplification product separation (ABI3130, Life Technologies) was performed in accordance with the manufacturer's instructions. Relative copy numbers were obtained after normalization of the peaks against blood-derived controls. Sequences were analyzed using Gene mapper software.

2.4 | Analysis of nucleotide alterations

To detect SNVs within *IKBKG* coding sequences, the coding exons were amplified using nested PCR without amplification of *IKBKGP*. Electropherograms were then obtained by Sanger sequencing using Big Dye Terminator Cycle Sequencing Reactions on an ABI 3100 device (PE Applied Biosystems) and were compared with those for genomic sequences from control samples. To evaluate the low-level mosaicism of the SNVs, we performed deep sequencing of the *IKBKGP* region using next generation sequencer (NGS). We performed long-range PCR for three regions including the entire *IKBKG* gene and pseudogene. The primers used are listed in Table S5. Three PCR reactions were performed under the following cycling conditions: an initial denaturation of 94℃ for 2 min followed by 5 cycles of 98℃ for 10 s and 74℃ for 5 min, 5 cycles of 98℃ for 10 s and 72℃ for 5 min, 5 cycles of 98℃ for 10 s and 70℃ for 5 min, and 20 cycles of 98℃ for 10 s and 68℃ for 5 min. Successful amplification of the three PCR products was confirmed by agarose gel electrophoresis. Pooled DNA libraries were prepared using a Nextera XT DNA Sample Preparation Kit in accordance with the manufacturer's protocol (Illumina). Paired ends were sequenced for 150 bp using a Miseq Reagent Kit v2 (Illumina). The sequence reads were aligned with the human genome (hg38) reference sequence after manual editing to use between chrX:154541001-154567000 with BWA aligner. Approximately 10,000 reads from a single patient were obtained and analyzed for genotyping. Downstream processing was carried out with the Genome analysis toolkit (GATK), SAMtools (http://samtools.sourceforge.net/) and Picard Tools (http://broadinstitute.github.io/picard/). Substitution and Indel calls were made with a GATK HaplotypeCaller and GenotypeGVCF (Boisson et al., 2019). Approximately 10,000 reads from a single patient were obtained and analyzed for genotyping.

Variants that have not been registered in any database, are submitted as a new variant in Leiden Open variation Database (http://www.lovd.nl/3.0/home).

2.5 | X chromosome inactivation studies

To examine the X chromosome inactivation (XCI) pattern, we performed a HUMARA assay according to a previously described protocol (Beever et al., 2003) to assess skewing of the X chromosome inactivation. We first digested the genomic DNA with the methylation-sensitive restriction enzyme *Hpa*II. PCR primers, one of which was labeled with FAM, were designed across the polymorphic CAG repeat and two *Hpa*II sites in the androgen receptor (*AR*) gene on the X chromosome. In addition to the *AR* locus, the *SLITRK4* and *PCSK1N* loci were used in case the *AR* locus was not informative (Bertelsen & Tümer, 2011). PCR amplification would be achieved only from the inactivated allele harboring methylated *Hpa*II sites. PCR products were analyzed by capillary electrophoresis (ABI3730 Genetic Analyzer) and quantified via the area under the curve using GeneMapper software. A

skewed XCI was determined when the inactivated allele was biased at more than 90% (Beever et al., 2003).

2.6 | Statistical analysis

Intergroup comparisons were made using the Student's *t*-test or one way analysis of variance method and *P* values of less than 0.05 were considered statistically significant.

3 | RESULTS

In our current IP cohort, we detected the *IKBKG* exon 4-10 common deletion in 13 patients (03f, 05f, 06f, 08f, 13f, 14f, 16f, 18f, 21f, 22f, 28f, 29f, and 30f; Figure 1a) which was a prevalence (detection rate) of 43.3% among the total series
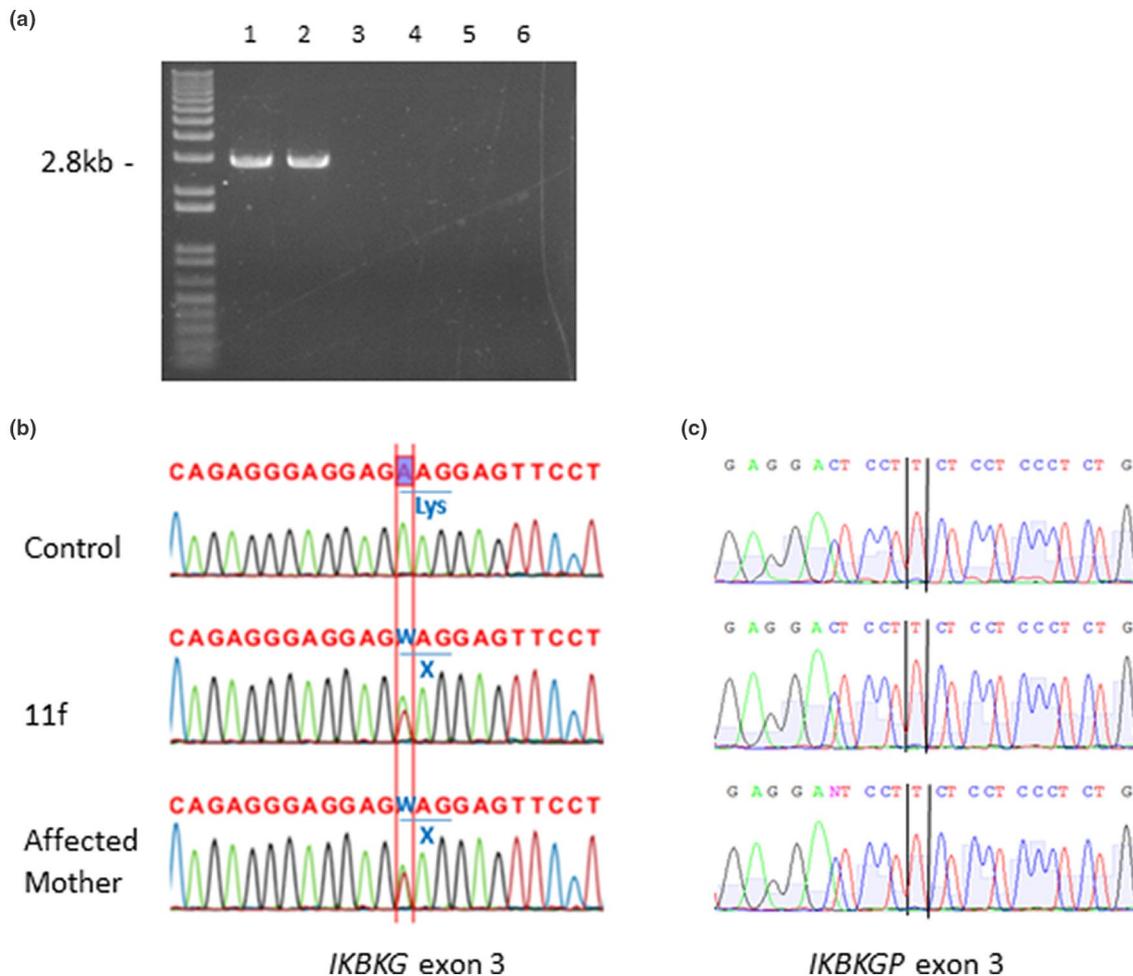
FIGURE 1 Mutational analysis of the *IKBKG* gene in the IP study patients. (a) PCR analysis of the IP patient blood samples. Lane 1, *IKBKG* deletion exon 4-10 control. Lane 2, patient 18f. The *IKBKG* deletion was detected. Lanes 3 and 4 are the father and mother of 18f. Lanes 5 and 6 are a healthy control and distilled water (DW). (b) Sanger sequencing of patient 11f and her affected mother. Coding exons for the *IKBKG* gene were amplified by PCR, being not amplified together with IKBKGP.NM_001099857.2:c.268A>T, NP_001093327.1: p.Lys90X is confirmed. (c) Pseudogene-specific sequencing. Sequencing of 11f and her mother did not show the pathogenic variant at chrX:154648193 on *IKBKGP*.

TABLE 1 Summary of the mutation.

Patient	Age at genetic testing	Female/male	Mutation[a]	Exon/intron	Inheritance	Phenotype[b] Score	XCI pattern
01f	0	F	ND	—	Sporadic	2	91.8%
02f	3	F	ND	—	Sporadic	2	79.0%
03f	0	F	ex 4-10 deletion	ex 4-10	Familial	4	91.3%
04m	0	M	ND	—	Sporadic	2	(male)
05f	0	F	ex 4-10 deletion	ex 4-10	Sporadic	3	NI
06f	0	F	ex 4-10 deletion	ex 4-10	Sporadic	3	99.0%
07f	0	F	c.343A>T (K115X)	ex 3-10	Familial	3	82.4%
08f	3	F	ex 4-10 deletion	ex 4-10	Sporadic	3	100.0%
09f	0	F	c.896delC (P299RfsX3)	ex 7-10	Familial	4	99.1%
10f	0	F	94 kb deletion	ex 3-10	Familial	4	97.3%
11f	0	F	c.268A>T (K90X)	ex 3-10	Familial	3	98.4%
12f	2	F	Mosaic ex 4-10 deletion	ex 4-10	Sporadic	2	67.1%
13f	0	F	ex 4-10 deletion	ex 4-10	Sporadic	3	100.0%
14f	48	F	ex 4-10 deletion	ex 4-10	Familial	5	100.0%
15f	0	F	ND	—	Sporadic	2	72.2%
16f	0	F	ex 4-10 deletion	ex 4-10	Familial	3	100.0%
17f	0	F	c.184C>T (R62X)	ex 2	ND	3	97.4%
18f	0	F	ex 4-10 deletion	ex 4-10	Sporadic	5	NI
19f	0	F	Mosaic ex 4-10 deletion	ex 4-10	Sporadic	1	50.0%
20f	0	F	Mosaic ex 4-10 deletion	ex 4-10	Sporadic	1	58.4%
21f	2	F	ex 4-10 deletion	ex 4-10	Sporadic	3	100.0%
22f	34	F	ex 4-10 deletion	ex 4-10	ND	3	100.0%
23f	44	F	c.913-2A>G (p.spl)	int 7	Familial	5	100.0%
24m	0	M	Mosaic ex 4-10 deletion	ex 4-10	Sporadic	2	(male)
25f	0	F	Mosaic ex 4-10 deletion	ex 4-10	Sporadic	2	65.3%
26f	2	F	c.976_978delAAG (K326del)	ex 8	ND	3	75.7%
27f	33	F	ND	—	Sporadic	2	92.8%
28f	30's	F	ex 4-10 deletion	ex 4-10	ND	5	98.5%
29f	1	F	ex 4-10 deletion	ex 4-10	Sporadic	3	51.3%
30f	1	F	ex 4-10 deletion	ex 4-10	Sporadic	3	98.3%

Note: Homozygote for the polymorphic CAG repeat in the *AR* gene.

Abbreviations: ND, not determined. NI, not informative.

[a]The mutation numbering is based on the IKBKG cDNA sequence according to the GenBank Accession number NM_003639.3. Codon numbering starts from the translation initiation codon 1 according to the GenBank Accession number NP_003630.1.

[b]Phenotype score was allocated from Table S2.

of 30 patients (Table 1). Among these 13 patients with the common deletion, three were familial cases (03f, 14f, 16f) and we confirmed in one case that the same deletion was present in the mother (03f). We next performed Sanger sequencing in the 17 IP patients who did not carry the recurrent deletion. Pathogenic SNVs were detected in 6 cases [07f, c.343A>T (p.K115X); 09f, c.896delC (p.P299RfsX3); 11f, c.268A>T (p.K90X; Figure 1b,c); 17f, c.184C>T (p.R62X); 23f, c.913-2A>G; 26f, c.976_978delAAG (p.K326del)]. These identified mutations included three

nonsense mutations (07f, 11f, 17f), one frameshift mutation (09f), one in-frame amino acid deletion (26f), and one splicing mutation (23f). To identify structural rearrangements caused by large deletions or duplications at the IP locus, including *IKBKG* and neighboring genes, we performed MLPA analysis to investigate copy number variation in the region. We thereby identified a large deletion including the *IKBKGP* pseudogene (10f). Although we identified pathogenic variants in 20 of the 30 study patients (66.7%), this rate of detection was suboptimal.

In the remaining 10 patients with no detectable mutations via standard analyses, we observed that their skin symptoms were less severe than the patient with detectable *IKBKG* mutations. In the evaluations of phenotype severity using our scoring system, the detected mutation cases had a score of 3.55 (n = 20), and those without a detectable mutation had a score of 1.8 (n = 10), which was significantly low ($p < 0.01$; Table 1, Table S1). To exclude the possibility that these patients had another autosomal disease that mimics IP, their XCI status was analyzed using a HUMARA assay since female X-linked IP patients

generally manifest a selective skewed X-inactivation. We performed the HUMARA assay on all of the female patients in our current IP cohort (n = 28) as well as some of their female family members (n = 16; Figure 2a, Table S2). We could not however determine the allelic status in some of the cases since the polymorphic CAG repeat in the *AR* gene was homozygous in five of the women, including the mother of 03f, whose values of XCI pattern was determined by the other method (described later). We obtained the values of XCI pattern of 40 females (31 patients and 9 unaffected family members).
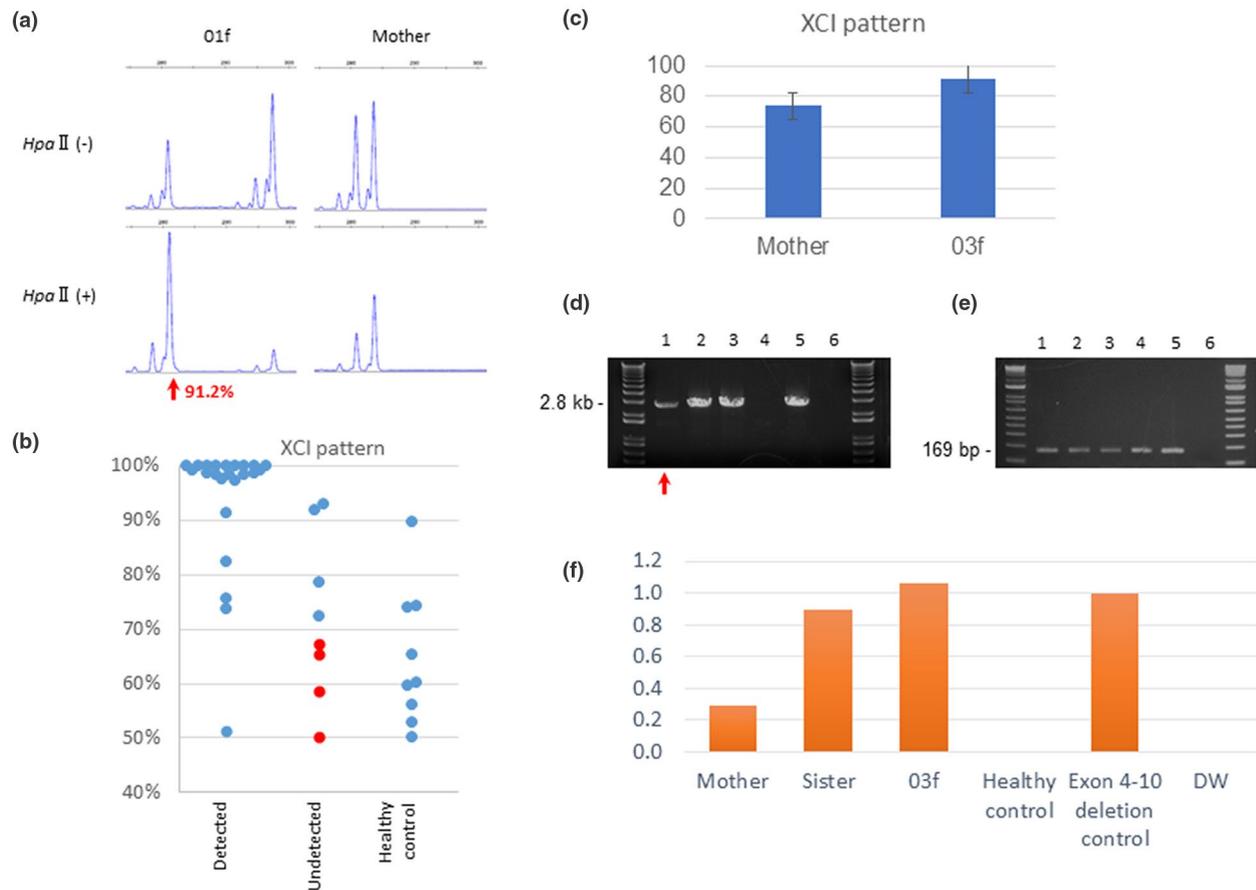

FIGURE 2 XCI patterns among IP patients in the study cohort. (a) Electropherogram of the HUMARA assay data showing the fragment analysis of PCR products amplified from undigested and digested DNA of patient 01f and her mother. The two major peaks represent two alleles with different numbers of short tandem repeats at the HUMARA locus (red arrows). After digestion, the DNA of the mother displayed a matching pattern of preferential loss of the short alleles. (b) Scatter plot of the informative XCI pattern of female IP probands and their mothers. "Detected" represents 18 probands, 4 carrier mothers, and 1 carrier sisters in whom an *IKBKG* mutation was detected. The median XCI value was 98.5% in these 23 cases. "Undetected" represents 8 probands with a median XCI value of 69.7%. The median XCI value of the healthy controls was 60.2%. Mosaic mutations are denoted by red dots. 11 probands, 1 carrier mother and 1 carrier sister carried the exon 4-10 deletion; one patient harbored a 94 kb deleted rearrangement; six probands and their three carrier mothers harbored SNVs; eight female probands had unknown mutations. The median age of the patients was 0 years and of the healthy controls was 40 years. (c) Median XCI values at the *AR*, *ZDHHC15*, *SLITRK4*, and *PCSK1N* loci. A lower XCI level in the mother may indicate somatic and germline mosaicism. (d) Common deletion-specific PCR analysis. 2.8 kb PCR products indicate the common exon 4-10 deletion of the *IKBKG* gene. Lanes 1-3 represent the mother of 03f, the sister of 03f and 03f, respectively. Lane 4 is a healthy control. Lane 5 is an *IKBKG* exon 4-10 deletion control and lane 6 is DW. (e) Internal standard for semi-quantitative PCR. The same DNA samples were used as templates for control PCR to amplify intron 45 of the *DMD* gene. (f) Semi-quantitative PCR. Using image analyzer, measurements of deletion specific PCR products were divided by that of internal control PCR products (lane 5). Mother of 03f is likely to have a mosaic deletion since the amount of the PCR products was less than that of 03f and her sister with heterozygous deletion.

Among the informative 31 patients, the skewed XCI values ranged from 50% to 100%, indicating that some patients showed a selectively skewed XCI pattern while others showed a random XCI pattern. We divided the patients into two categories: patients with an already detected or undetected *IKBKG* mutation. The mean value for the XCI patterns of the patients with a detected mutation was 98.5% (n = 23). In contrast, the XCI pattern values of the patients without a detected mutation (n = 8) ranged from 50% to 92.8% (mean 69.7%), which was significantly low ($p < 0.01$; Figure 2b, Table S2). The mean value for the unaffected family members was 60.2% (n = 9; Figure 2b). It is possible that our remaining 10 patients without a detected mutation and with a mild IP phenotype might have had a disorder other than X-linked IP. However, one of our patients with an in-frame deletion of a single amino acid (p.K326del) in the *IKBKG* gene (26f) had an XCI value of 75.7%, indicating that a hypomorphic variant does not lead to a selectively skewed XCI. Although the K326 residue in the *IKBKG* gene product is located at a linear polyubiquitinated site (Ikeda et al., 2011) and is an evolutionarily conserved position, protein products with an in-frame deletion may possibly have some residual activity that prevents cellular lethality. Hence, it is likely that the patients with a mild phenotype and with an almost random XCI pattern harbor weak variants such as somatic mosaicism.

Further evidence of the relationship between phenotype severity and the XCI pattern was observed in patient 03f and her mildly affected mother, both of whom carried the common *IKBKG* deletion (Table 2). Since the polymorphic CAG repeat in the *AR* gene was homozygous in the mother, we instead assessed three alternative methylation–sensitive restriction enzyme sites in the *ZDHHC15*, *SLITRK4* and *PCSKIN* genes, respectively. The XCI patterns for patient 03f were 89.2% in *SLITRK4*; 93.1% in *PCSKIN*, a median of 91.3%,

whereas those for the mother were 67.7% in the *ZDHHC15* locus; 73.3% in *SLITRK4*; 82.5% in *PCSKIN*, a median of 73.7% (Figure 2c, Table 3). These data indicated that patient 03f manifests a selective skewed pattern, but that her mildly affected mother does not. Retrospectively, the amount of deletion-specific PCR products in the mother was less than that of the proband (Figure 2d). This finding indicated that the mother, who had a mild phenotype, may have somatic mosaicism of the common *IKBKG* deletion.

This means that the patients who are positive in PCR for common *IKBKG* deletion might possibly include those with somatic mosaicism. We reanalyzed the PCR-positive patients (03f, 05f, 06f, 08f, 15f, 17f, 28f, 29f, 30f, mother of 03f) by less sensitive MLPA. MLPA detected the common *IKBKG* deletion in all patients except for 29 f. Although the possibility of the effect of duplication polymorphism has not been ruled out, this results suggest that 29f might also harbor somatic mosaicism since the XCI pattern of 29f (51.3%) support this speculation.

A nested PCR assay in the remaining 10 patients revealed low-level mosaicisms of the common deletion in five cases (12f, 19f, 20f, 24m, 25f; Figure 3a-d). Using an image analyzer, the mosaic ratios in 12f, 19f, 20f, 24m and 25f were found to be 1/70, 1/20, 1/43, 1/20, and 1/62, respectively. With the ddPCR assay, the mosaic ratios in those patients were 1/75, 1/57, 1/39, 1/76, and 1/56, respectively (Figure 3e). Except for one male patient, we reviewed the XCI pattern in four of the female IP patients with a mosaic deletion and found XCI patterns of 67.1%, 50%, 58.4%, and 65.3%, respectively, with a mean value of 61.9%. The mean XCI value of the remaining four cases with undetected mutations was 79.0%, compared to 60.2% in the healthy controls. These XCI values thus showed distinct differences. Hence, as the XCI pattern of the mother, who had clinical symptoms, of 03f was 73.3%, this lower XCI may indicate a somatic and germline

TABLE 2 Phenotype score of the 03f family.

Patient	Age at genetic testing	Female/ male	Skin leision[a]	Nervous system defect[b]	Ocular system defect[c]	Dental system defect[d]	Hair defect[e]	Nails defect[f]	Phenotype score[g]
03f (proband)	0	F	3	1	ND	ND	ND	ND	4
Sister of 03f	3	F	3	0	0	0	1	0	4
Mother of 03f	38	F	2	0	0	0	0	0	2

Note: IP phenotype score analysis. A phenotype score of clinical severity was derived when possible in IP patients whose clinical data were available.

Abbreviation: ND, not determined.

Phenotype score represents the addition of the single score values for each system/organ.

[a]We assigned a score 1 for little skin abnormality in limbs; a score 2 for skin abnormality in some parts of limbs or trunk; a score of 3 for skin abnormality in all limbs and trunk. All IP reported patients suffered skin abnormality.

[b]A score of 1 was added for each nervous system (NS) defect (seizures, or spastic paresis, or motor retardation, or mental retardation or microcephaly).

[c]A score of 1 was added for each ocular system defect (strabismus, or cataracts, or optic atrophy, or retinal vascular pigmentary abnormalities, ormicrophthalmos, or pseudogliomas).

[d]A score of 1 was added for each Dental System defect (partial anodontia, or delayed dentition, or cone/peg shaped teeth, or impactions).

[e]A score of 1 was added for each Hair defect (vertex alopecia, or wooly hair nevus, or eyelash and eyebrow hypogenesis).

[f]A score of 1 was added for each Nails defect (onychogryposis, or pitting, or ridging).

[g]Phenotype score represents the addition of the single score values for each system/organ.

	AR	*ZDHHC15*	*SLITRK4*	*PCSKIN*	**Median (%)**
03f	91.3	NI	89.2	93.1	91.3
Sister of 03f	98.3	—	—	—	98.3
Mother of 03f	NI	67.7	73.7	82.5	73.7

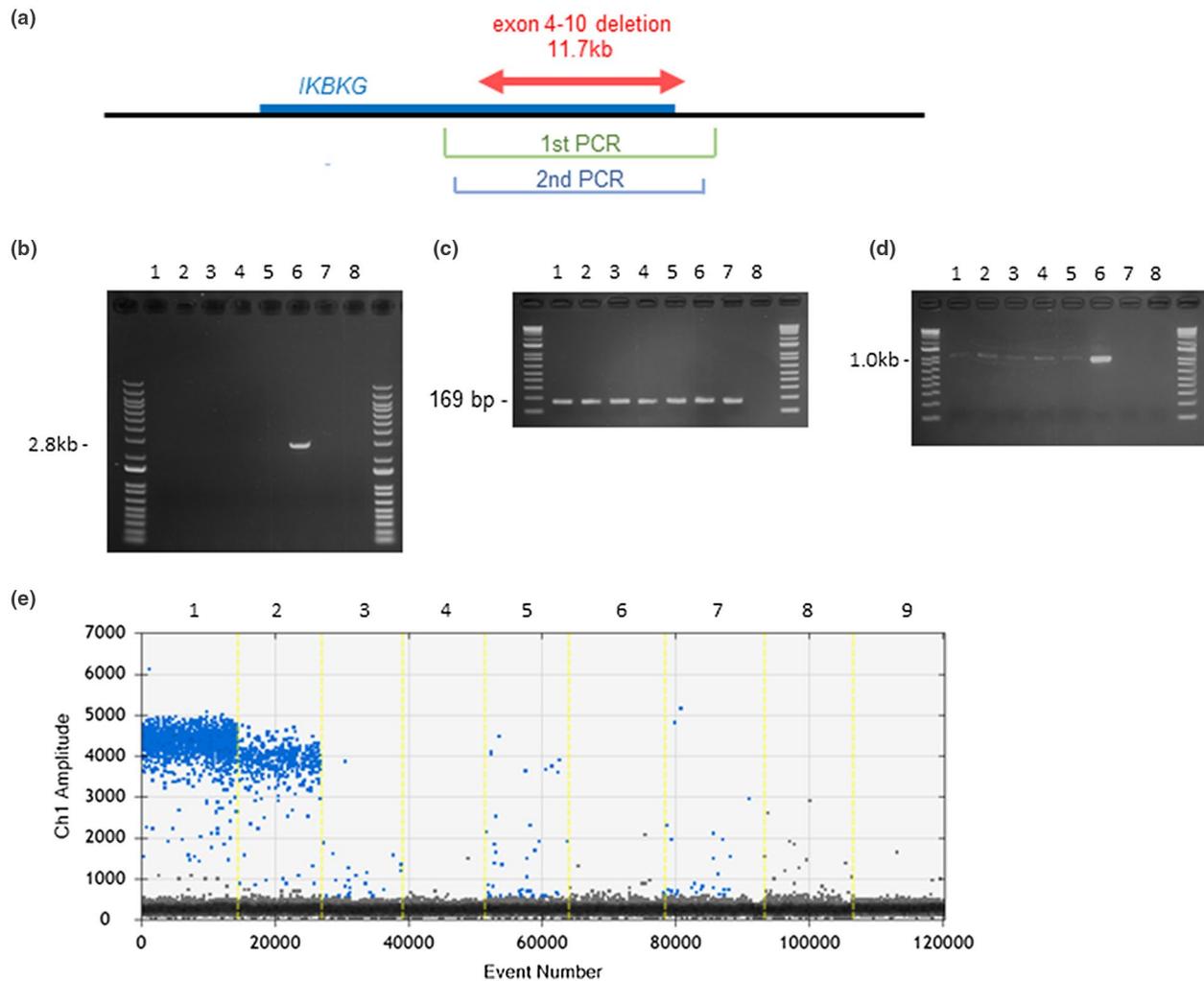Abbreviation: NI, not informative.

FIGURE 3 Mosaic analysis of the *IKBKG* gene. (a) Schematic showing the relative location of the *IKBKG* gene and its exons 4-10. The position of the common exon 4-10 deletion in IP is indicated by red arrows. The green line denotes the region of the first PCR and the blue line the second PCR. (b) First PCR. A 2.8 kb amplified product indicates an *IKBKG* exon 4-10 deletion. Lanes 1-5 represent 12f, 19f, 20f, 20m, and 25f, respectively. Lane 6, *IKBKG* exon 4-10 deletion control; lane 7, healthy control; lane 8; DW. (c) Internal standard for semi-quantitative PCR. The same DNA samples were used as templates for control PCR to amplify intron 45 of the *DMD* gene. (d) Second PCR. A 1.0 kb amplified product indicates a mosaic deletion, which was evident in lanes 1-5. Lanes 6-8 are as described in (b). (e) ddPCR. Plots show a quantification of fluorescence signal of droplet. Green dots indicate fluorescence-positive droplets, while black dots indicate negative droplets. Lane 1 represent *IKBKG* exon 4-10 deletion control. Lane 2, 3, 4, 5, 6, 7 represent the mother of 03f, 12f, 19f, 20f, 24m, 25f, respectively. Lane 8,9 represent healthy control and DW, respectively.

mosaicism. Our result of PCR assay and ddPCR assay revealed a mosaic ratio of 1/4 and 1/5 for the mother of 03f, respectively (Figure 3e). To confirm the somatic mosaicism, we attempted to obtain somatic tissues other than blood in these patients. Cheek swab sample was obtained only from 24 m, which did not show the common deletion by the nested PCR (data not shown).

We finally performed NGS analysis to detect mosaic SNVs in the remaining subjects (01f, 02f, 04m, and 15f).

Approximately, 90% of the cumulative target region was covered with a sequence depth of more than 200X. We evaluate SNVs and indels at a greater than 1% frequency in the coding exons of the _IKBKG_ gene and _IKBKGP_ pseudogene. No SNVs that indicated low-level mosaicism were identified in these patients. These data are summarized in Table 1. When including the low-level mosaic mutations, we detected a total of 25 _IKBKG_ mutations in 30 samples (83.3%).

4 | DISCUSSION

In our current study of 30 unrelated IP families from Japan, we identified _IKBKG_ mutations in 25 families, including 5 low-level mosaic mutations. In a previous large-scale study of IP, 65% of the cases had the common exon 4-10 deletion in the _IKBKG_ gene (Fusco et al., 2008), 8.6% harbored SNVs (Fusco et al., 2007), and 4% had a larger deletion at the _IKBKG_ locus (Conte et al., 2014). In our current standard protocol, we initially identified the common _IKBKG_ deletion at a much reduced frequency of 43.3% (13/30), which made our initial mutation detection efficiency low (20/30, 66.7%) in comparison to previous studies. This outcome might have partly been due to ethnic differences. Another possible reason may be that there was a bias in the patients referred to our institute, that is, they were mainly non-deleted cases that came via deletion screening in a prior facility. We finally identified a total of 25 mutations in 30 IP patients after introducing a detection system for low-level mosaicism (83.3%).

We identified low-level mosaicism of the common _IKBKG_ deletion in 16.7% of our current IP cases (5/30). Thus, as much as 27.8% of the common deletions in our current study series were mosaic (5/18). This suggested that our diagnosis system, semi-quantitative nested PCR and ddPCR, could efficiently detect low-level mosaicism. We conclude from this that clinical geneticists should become aware of the high incidence of low-level mosaicism in this disorder. Our present data have indicated that IP patients with mild symptoms only affecting the skin and those with low XCI values tended to include cases of low-level mosaic mutations in the _IKBKG_ gene. It is well documented that XCI values are elevated with age (Hatakeyama et al., 2004). Our current patients median age was 0 years old in both groups (no significant difference), although that of the unaffected family members was 38.5 years. Thus, these are likely predictors of a low-level mosaicism when a mutation survey fails to detect mutations in the _IKBKG_ gene. Hence, a HUMARA assay may be a powerful method of differentiating the mutational status of the _IKBKG_ gene in IP patients. We recommend the use of this assay when the standard mutational analyses fail to detect mutations in the _IKBKG_ gene in IP patients.

It is well acknowledged that random XCI in humans occurs in early embryogenesis after the blastocyst stage (Lee &

Bartolomei, 2013). If a zygote with a germinal mutation in the _IKBKG_ gene undergoes random XCI, about half of the cells carry the mutation on the activated X chromosome and the other half carry the mutation on the inactive X. During early embryogenesis, the cells with the mutation on the activated X chromosome would undergo apoptosis and be negatively selected. Thus, a skewed XCI pattern would be established. A small number of surviving cells harboring the _IKBKG_ mutation on the activated X chromosome would then induce local mild symptoms. In contrast, if the mutation in the _IKBKG_ gene arises in the very early embryonic stages prior to XCI determination, most of the cells would be normal and lack the mutation. Even if the cells with the mutation on the activated X chromosome would be negatively selected, XCI data would show a random XCI pattern reflecting a majority of normal cells. Finally, if the mutation in the _IKBKG_ gene arises after XCI determination, XCI data would show a random XCI pattern reflecting the majority of normal cells regardless of whether the mutation occurred on the activated or inactivated X chromosome. According to their mosaic ratio, our current study probands were found to have undergone the somatic mutation event at the embryonic stage from the end of the morula stage, which consists of 16-36 cells, to the blastocyst stage. However, whether these mutations occurred before or after the XCI patterns are determined, patients with low-level mosaicism would not manifest skewed XCI patterns.

We did not detect any mutations in the _IKBKG_ gene in four patients (01f, 02f, 04m, 15f and 27f). One of these cases (27f) showed a skewed XCI pattern (92.8%), suggesting that she had a germline mutation that we could not detect with our methods. Two of our current study patients (02f and 15f) did not show a skewed XCI pattern (79.0% and 72.2%, respectively). The other case (04m) was a male patient with a normal karyotype. These three IP patients may therefore have low-level mosaic mutations not of the common deletion, but at the nucleotide level. The detection sensitivity of the deep sequencing we used prevented us from identifying mosaic nucleotide mutations. With a read depth of 200-400, up to a 1% mutation frequency in the coding exons of the _IKBKG_ gene could be theoretically detected in spite of the presence of random artifacts such as misincorporations during DNA synthesis. However, a level of mosaicism below 1% could not be detected even when this deep sequencing was performed. Further investigations using ultra deep sequencing might elucidate the mutations of these remaining patients.

IP females have a risk of having an affected baby, with the risk of inheritance generally thought to be 50%, that is, male embryonic lethality and the birth of affected females. However, our current analyses indicate that the apparently affected IP females include those harboring germline mutations and those with a low level of mosaicism. Given that IP females can have low-level mosaic somatic mutations, the risk of inheritance becomes very small. In sporadic IP

cases, a mother with an apparently normal phenotype may have a hidden low level of mosaicism. Among the IP patients who carry the *IKBKG* exon 4-10 deletion, 65% are sporadic cases (Fusco et al., 2009). In addition, 3.8% of IP offspring with a de novo mutation have parents who shows 1% mosaicism in their blood cells (Rahbari et al., 2015). Clinical symptoms of IP appearing in patients with low-level mosaic mutations might be too mild to be diagnosed properly. Our current findings suggest that most mothers of sporadic cases may have no germline mutation, but some might show a low level of mosaicism when a highly sensitive mutation search is performed. Furthermore, even when the mother of a child with IP is correctly diagnosed with a low level of mosaicism, they would not necessarily have a small risk of having affected children. The mother of our current study patient 03f showed a 1/4 mosaic mutation, but was a parent of three girls of which two were affected with IP. A noteworthy limitation of our current analysis is that we only have the mosaicism information from blood samples and not from the germinal cells. In any event, a correct diagnosis of germline or somatic mosaicism in relation to IP will facilitate more informed genetic counseling for family planning purposes.

In conclusion, we have here identified five patients with low level of mosaicism of the common *IKBKG* exon 4-10 deletion that causes IP. These patients manifest a mild IP phenotype only and no skewed XCI pattern, suggesting that XCI pattern values can predict the possibility of mosaicism for this disorder. Determination of the XCI value in advance of mutational analyses for IP could improve the mutation detection rate. Detecting mosaic mutations will also be beneficial for genetic counseling of affected individuals.

CONFLICT OF INTEREST
The authors declare no competing interests in relation to this study.

AUTHOR CONTRIBUTIONS
MK carried out the genomic analysis and drafting of the manuscript. TK, MT, and HI was responsible for genomic analysis. YS carried out NGS analysis. HK conceived the study and participated in its design and drafted the manuscript. All authors read and approved the final version of the manuscript and agree with the order of presentation of the authors.

DATA AVAILABILITY STATEMENT
The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

ORCID
Hiroki Kurahashi https://orcid.org/0000-0002-5690-5218

REFERENCES
Aradhya, S., Bardaro, T., Galgóczy, P., Yamagata, T., Esposito, T., Patlan, H., Ciccodicola, A., Munnich, A., Kenwrick, S., Platzer, M., & D'Urso, M. (2001). Multiple pathogenic and benign genomic rearrangements occur at a 35 kb duplication involving the NEMO and LAGE2 genes. *Human Molecular Genetics*, *10*(22), 2557-2567. https://doi.org/10.1093/hmg/10.22.2557

Aradhya, S., Courtois, G., Rajkovic, A., Lewis, R. A., Levy, M., Israël, A., & Nelson, D. L. (2001). Atypical forms of incontinentia pigmenti in male individuals result from mutations of a cytosine tract in Exon 10 of NEMO (IKK- g). *American Journal of Human Genetics*, *68*(3), 765-771.

Bardaro, T., Falco, G., Sparago, A., Mercadante, V., Gean Molins, E., Tarantino, E., Valeria Ursini, M., & D'Urso, M. (2003). Two cases of misinterpretation of molecular results in incontinentia pigmenti, and a PCR-based method to discriminate NEMO/IKKγ gene deletion. *Human Mutation*, *21*(1), 8-11. https://doi.org/10.1002/humu.10150

Beever, C. L., Stephenson, M. D., Peñaherrera, M. S., Jiang, R. H., Kalousek, D. K., Hayden, M., Field, L., Brown, C. J., & Robinson, W. P. (2003). Skewed X-chromosome inactivation is associated with trisomy in women ascertained on the basis of recurrent spontaneous abortion or chromosomally abnormal pregnancies. *The American Journal of Human Genetics*, *72*(2), 399-407. https://doi.org/10.1086/346119

Bertelsen, B., Tümer, Z., & Ravn, K. (2011). Three new loci for determining X chromosome inactivation patterns. *The Journal of Molecular Diagnostics*, *13*(5), 537–540. https://doi.org/10.1016/j.jmoldx.2011.05.003

Boisson, B., Honda, Y., Ajiro, M., Bustamante, J., Bendavid, M., Gennery, A. R., Kawasaki, Y., Ichishima, J., Osawa, M., Nihira, H., Shiba, T., Tanaka, T., Chrabieh, M., Bigio, B., Hur, H., Itan, Y., Liang, Y., Okada, S., Izawa, K., … Yasumi, T. (2019). Rescue of recurrent deep intronic mutation underlying cell type – dependent quantitative NEMO deficiency Find the latest version : Rescue of recurrent deep intronic mutation underlying cell type – dependent quantitative NEMO deficiency. *The Journal of Clinical Investigation*, *129*(2), 583-597. https://doi.org/10.1172/JCI124011

Conte, M. I., Pescatore, A., Paciolla, M., Esposito, E., Miano, M. G., Lioi, M. B., McAleer, M. A., Giardino, G., Pignata, C., Irvine, A. D., & Scheuerle, A. E. (2014). Insight into IKBKG/NEMO locus: Report of new mutations and complex genomic rearrangements

leading to incontinentia pigmenti disease. *Human Mutation*, *35*(2), 165-177. https://doi.org/10.1002/humu.22483

Döffinger, R., Smahi, A., Bessia, C., Geissmann, F., Feinberg, J., Durandy, A., Bodemer, C., Kenwrick, S., Dupuis-Girod, S., Blanche, S., Wood, P., Rabia, S. H., Headon, D. J., Overbeek, P. A., Le Deist, F., Holland, S. M., Belani, K., Kumararatne, D. S., Fischer, A., … Casanova, J.-L. (2001). X-linked anhidrotic ectodermal dysplasia with immunodeficiency is caused by impaired NF- κ B signaling. *Nature Genetics*, *27*, 277-285.

Fusco, F.,Bardaro, T., Fimiani, G., Mercadante, V., Miano, M. G., Falco, G., Israël, A., Courtois, G., D'Urso, M., & Ursini, M. V. (2004). Molecular analysis of the genetic defect in a large cohort of IP patients and identification of novel NEMO mutations interfering with NF-κB activation. *Human Molecular Genetics*, *13*(16), 1763-1773. https://doi.org/10.1093/hmg/ddh192

Fusco, F., Fimiani, G., Tadini, G., Michele, D'. U., & Ursini, M. V. (2007). Clinical diagnosis of incontinentia pigmenti in a cohort of male patients. *Journal of the American Academy of Dermatology*, *56*(2), 264-267. https://doi.org/10.1016/j.jaad.2006.09.019

Fusco, F., Mariateresa, P., Alessandra, P., Brigida, L. M., Carmen, A., Francesca, F., Mattia, G., Marcella, Z., Michele, D. U., Giuseppina, M. M., & Valeria, U. M. (2009). Microdeletion/duplication at the Xq28 IP Locus causes a de novo IKBKG/NEMO/IKKgamma exon4-10 deletion in families with incontinentia pigmenti. *Human Mutation*, *30*(9), 1284-1291. https://doi.org/10.1002/humu.21069

Fusco, F., Paciolla, M., Conte, M., Pescatore, A., Esposito, E., Mirabelli, P., Lioi, M., & Ursini, M. (2014). Incontinentia pigmenti: Report on data from 2000 to 2013. *Orphanet Journal of Rare Diseases*, *9*(1), 1-5. https://doi.org/10.1186/1750-1172-9-93

Fusco, F., Paciolla, M., Napolitano, F., Pescatore, A., D'Addario, I., Bal, E., Lioi, M. B., Smahi, A., Miano, M. G., & Ursini, M. V. (2012). Genomic architecture at the incontinentia pigmenti locus favours de novo pathological alleles through different mechanisms. *Human Molecular Genetics*, *21*(6), 1260-1271. https://doi.org/10.1093/hmg/ddr556

Fusco, F., Pescatore, A., Bal, E., Ghoul, A., Paciolla, M., Lioi, M. B., D'Urso, M., Rabia, S. H., Bodemer, C., Bonnefont, J. P., Munnich, A., Miano, M. G., Smahi, A., & Ursini, M. V. (2008). Alterations of the IKBKG locus and diseases: An update and a report of 13 novel mutations. *Human Mutation*, *29*(5), 595-604. https://doi.org/10.1002/humu.20739

Hatakeyama, C., Anderson, C. L., Beever, C. L., Peñaherrera, M. S., Brown, C. J., & Robinson, W. P. (2004). The dynamics of

X-inactivation skewing as women age. *Clinical Genetics*, *66*(4), 327-332. https://doi.org/10.1111/j.1399-0004.2004.00310.x

Ikeda F., Deribe Y. L., Skånland S. S., Stieglitz B., Grabbe C., Franz-Wachtel M., van Wijk S. J. L., Goswami P., Nagy V., Terzic J., Tokunaga F., Androulidaki A., Nakagawa T., Pasparakis M., Iwai K., Sundberg J. P., Schaefer L., Rittinger K., Macek B., Dikic I. (2011). SHARPIN forms a linear ubiquitin ligase complex regulating NF-κB activity and apoptosis. *Nature*, *471*, (7340), 637–641. http://dx.doi.org/10.1038/nature09814.

Landy, S. J., & Donnai, D. (1993). Incontinentia pigmenti (Bloch-Sulzberger syndrome). *Journal of Medical Genetics*, *30*(1), 53-59.

Lee, J. T., & Bartolomei, M. S. (2013). Review X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell*, *152*(6), 1308-1323. https://doi.org/10.1016/j.cell.2013.02.016

Minić, S., Trpinac, D., & Obradović, M. (2014). Incontinentia pigmenti diagnostic criteria update. *Clinical Genetics*, *85*(6), 536-542. https://doi.org/10.1111/cge.12223

Orphanet. (2018). Prevalence and incidence of rare diseases: Bibliographic data. *Number*, 1(1). www.orpha.net%0Awww.orpha.net

Rahbari, R., Wuster, A., Lindsay, S. J., Hardwick, R. J., Alexandrov, L. B., Al Turki, S., Dominiczak, A., Morris, A., Porteous, D., Smith, B., Stratton, M. R., & Hurles, M. E. (2015). Articles Timing, rates and spectra of human germline mutation. *Nature Genetics*, *48*(2), 126-133. https://doi.org/10.1038/ng.3469

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.
Supplementary Material

How to cite this article: Kawai M, Kato T, Tsutsumi M, Shinkai Y, Inagaki H, Kurahashi H. Molecular analysis of low-level mosaicism of the *IKBKG* mutation using the X Chromosome Inactivation pattern in Incontinentia Pigmenti. *Mol Genet Genomic Med*. 2020;00:e1531. https://doi.org/10.1002/mgg3.1531

ORIGINAL ARTICLE

Cancer Science WILEY

Impact of *DPYD*, *DPYS*, and *UPB1* gene variations on severe drug-related toxicity in patients with cancer

Katsuyuki Yokoi[1,2] (iD) | Yoko Nakajima[1] | Hiroshi Matsuoka[3] | Yasuko Shinkai[2] | Takuma Ishihara[4] | Yasuhiro Maeda[5] | Takema Kato[2] | Hidetoshi Katsuno[3] | Koji Masumori[3] | Kenji Kawada[6] | Tetsushi Yoshikawa[1] | Tetsuya Ito[1] | Hiroki Kurahashi[2]

[1]Department of Pediatrics, Fujita Health University School of Medicine, Toyoake, Japan

[2]Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Japan

[3]Department of Gastrointestinal Surgery, Fujita Health University School of Medicine, Toyoake, Japan

[4]Innovative and Clinical Research Promotion Center, Gifu University Hospital Gifu University, Gifu, Japan

[5]Center for Joint Research Facilities Support, Fujita Health University, Toyoake, Japan

[6]Department of Medical Oncology, Fujita Health University School of Medicine, Toyoake, Japan

Correspondence
Yoko Nakajima, Department of Pediatrics, Fujita Health University School of Medicine, 1-98 Dengakugakubo, Kutsukake-cho, Toyoake, Aichi 470-1192, Japan.
Email: yonaka@fujita-hu.ac.jp

Abstract

Cancer treatment with a fluoropyrimidine (FP) is often accompanied by severe toxicity that may be dependent on the activity of catalytic enzymes encoded by the *DPYD*, *DPYS*, and *UPB1* genes. Genotype-guided dose individualization of FP therapy has been proposed in western countries, but our knowledge of the relevant genetic variants in East Asian populations is presently limited. To investigate the association between these genetic variations and FP-related high toxicity in a Japanese population, we obtained blood samples from 301 patients who received this chemotherapy and sequenced the coding exons and flanking intron regions of their *DPYD*, *DPYS*, and *UPB1* genes. In total, 24 single nucleotide variants (15 in *DPYD*, 7 in *DPYS* and 2 in *UPB1*) were identified including 3 novel variants in *DPYD* and 1 novel variant in *DPYS*. We did not find a significant association between FP-related high toxicity and each of these individual variants, although a certain trend toward significance was observed for p.Arg181Trp and p.Gln334Arg in *DPYS* ($P = .0813$ and .087). When we focused on 7 *DPYD* rare variants (p.Ser199Asn, p.Ile245Phe, p.Thr305Lys, p.Glu386Ter, p.Ser556Arg, p.Ala571Asp, p.Trp621Cys) which have an allele frequency of less than 0.01% in the Japanese population and are predicted to be loss-of-function mutations by in silico analysis, the group of patients who were heterozygous carriers of at least one these rare variants showed a strong association with FP-related high toxicity ($P = .003$). Although the availability of screening of these rare loss-of-function variants is still unknown, our data provide useful information that may help to alleviate FP-related toxicity in Japanese patients with cancer.

KEYWORDS

5-fluorouracil, *DPYD*, *DPYS*, fluoropyrimidine, *UPB1*

- -

1 | INTRODUCTION

Fluoropyrimidines including 5 5FU and its prodrugs are widely used in the treatment of various malignancies including head and neck, gastrointestinal, or breast cancers.[1-3] FPs have a narrow therapeutic index and up to 30% of treated patients develop early-onset severe toxicity such as diarrhea, nausea, mucositis, stomatitis, myelosuppression, neurotoxicity, and hand-foot syndrome.[4-6] FP toxicity is largely dependent on its catabolism. Most FP molecules are inactivated by DPD and FP-related toxicity is often caused by an inherited reduced activity of this enzyme.[7-9] Patients with a DPD deficiency have an increased risk of developing severe treatment-related toxicity from a standard dose of FP.[10] A partial DPD deficiency is present in 3%-5% of the North American and European general population.

The DPD gene, *DPYD*, is located on chromosome 1p21 and is comprised of 23 exons.[11] The 4 *DPYD* variants considered most clinically relevant with statistically significant associations with severe toxicity are c.1905 + 1G>A (*DPYD**2A, rs3918290, IVS14 + 1G>A), c.2846A>T (rs67376798, D949V), c.1679T>G (rs55886062, *DPYD*13, I560S) and c.1236G>A (rs56038477, E412E, in haplotype B3).[12,13] Hence, *DPYD* genotype-guided dose individualization of FP therapy has now been conducted in some western countries.[10] However, to our knowledge it has not been performed yet in Japan, possibly because none of these *DPYD* variants have been identified in the Asian population.[14] Recently, 21 *DPYD* allelic variants were identified in 1070 healthy Japanese individuals.[15] The functional alterations caused by these variants were analyzed in vitro and their enzyme activities were characterized.[14] However, there has been no report to date on the clinical relevance of *DPYD* variants as predictors of FP-associated toxicity in Japanese people.

It is thought that decreased activity of the enzymes DHP and βUP, which are located downstream of DPD in FP catabolism, may also play a role in FP-associated toxicity.[16] The DHP-encoding gene, *DPYS*, is located on chromosome 8q22,[17] and the βUP-encoding gene, *UPB1*, is located on chromosome 22q11.[18] A relationship between *DPYS* and *UPB1* gene variations and severe FP-related toxicity has been reported,[16] but no other data currently support this association, which thus remains to be fully elucidated. It is known that the Japanese prevalence of βUP deficiency is relatively high (1 per 6000 newborns).[19] We have also reported that some *DPYS* variants may be more common than expected in East Asian groups.[20] These findings have prompted us to screen for variants of the genes associated with FP-related toxicity in Japanese subjects.

We have here evaluated the association between *DPYD*, *DPYS*, and *UPB1* gene variations and severe FP-related toxicity in Japanese patients with cancer. This is the first report to assess the clinical relevance of *DPYD*, *DPYS*, and *UPB1* variants as predictors of severe FP-associated toxicity in East Asians.

2 | MATERIALS AND METHODS

2.1 | Patients and sample collection

Blood samples of 301 consenting patients who received or were receiving FP-based chemotherapy were collected between 2018 and 2020. All patients were of East Asian origin. These 301 patients were recruited at Fujita Health University. All treatments, patient characteristics, concurrent therapy and adverse effects (gastrointestinal [nausea, vomiting, diarrhea, oral mucositis], neutropenia, hand-foot syndrome, acute kidney injury) developed within the first 2 cycles of treatment in this cohort and were classified according to the CTCAE v4.0. We divided these subjects into 2 groups in accordance with the grade of toxicity for statistical purposes. The high-toxicity group included patients who experienced severe toxicity presenting with CTCAE grade 3-5 adverse events in any category. The low-toxicity group included patients who experienced low toxicity involving CTCAE grade 0-2 adverse events.

2.2 | DPYD, DPYS, UPB1 sequencing analysis

Genomic DNA was extracted from aliquots of the study patient blood specimens using a standard procedure. We designed a custom AmpliSeq panel for the sequencing of coding exons and flanking intron regions (± 10 bp) of *DPYD*, *DPYS*, and *UPB1*. Library preparation for amplicon sequencing was performed using AmpliSeq Library PLUS for Illumina. Libraries were sequenced on the MiSeq platform with 150 bp paired-end reads (Illumina, San Diego, CA). Sequencing data were analyzed with Illumina Basespace DNA Amplicon App. We used the UCSC genome browser (http://genome-asia.ucsc.edu/ human GRCh37/hg19) as the human genome assembly. Illumina Variant Studio was used for annotation and filtration of genomic variants with a Pass Filter read depth > 50×. Allele frequency was investigated with gnomAD browser beta (http://gnomad.broadinstitute.org/) and Jmorp https://jmorp.megabank.tohoku.ac.jp).

The in silico analysis of each variant was performed using Polymorphism Phenotyping ver. 2 (PolyPhen-2; http://genetics.bwh.harvard.edu/pph2) and SIFT (http://sift-dna.org) to predict the functional impact on the protein product. In the PolyPhen-2 program, the investigated variant is categorized as probably damaging (probability score > 0.85), possibly damaging (probability score between 0.16 and 0.85), or benign (probability score less than or equal to 0.15). SIFT is a tool for sorting intolerant from tolerant amino acids. The evaluated amino acid substitution is predicted as damaging if the score is <.05 and is predicted to be tolerated if the score is greater than or equal to .05.

2.3 | Statistical analyses

The study patient characteristics were presented using median and range for continuous variables, and frequencies and proportions for categorical variables. The Fisher exact test was used to identify the association of

DPYD, *DPYS*, and *UPB1* variations with FP-related high toxicity because some categories would have an expected count of 5 or less. The frequencies of high-toxicity and low-toxicity groups were compared between each genotype. Due to the exploratory nature of this study, a *P*-value less than .05 was considered statistically significant, and .05 < *P* < .1 was recognized as indicating a certain trend toward significance. Statistical analyses were conducted using R software version 3.6.2 (www.r-project.org).

3 | RESULTS

3.1 | Patient characteristics

The age, treatments, and cancer types of the 301 patients are listed in Table 1. The most commonly used regimen in this cohort was 5FU + L-OHP/CPT-11 + α(molecular target) (54.5%, n = 164), followed by 5FU monotherapy (28.9%, n = 87). A majority of patients (68.1%) had a colorectal tumor (n = 205). During the first 2 cycles of chemotherapy, 18.3% (n = 55) of the patients developed high-toxicity responses (CTCAE grade 3-4), and 81.7% (n = 246) showed low toxicity (CTCAE grade 0-2). No patient in our current series developed a grade 5 adverse event. The most frequent adverse event category observed in the high-toxicity group was neutropenia, followed by gastrointestinal-related issues (Table 2).

Toxicity rates varied and significantly depended on the FP regimen (Table 1). 80% (n = 4/5) and 53.8% (n = 7/13) of patients who received

DTX + CDDP+5FU and FOLFOXIRI + α, respectively, developed high toxicity, while, only 1 patient (1.8%) showed high toxicity in 5FU monotherapy.

3.2 | Variant analysis of the *DPYD*, *DPYS*, and *UPB1* genes

In total, 24 non-synonymous single nucleotide variants (15 *DPYD*, 7 *DPYS* and 2 *UPB1*) were identified in our present study population including 3 novel variants in *DPYD* and 1 novel variant in *DPYS*. (Table 3) Seven *DPYD* variants and 2 *DPYS* variants were rare variants with a minor allele frequency in the Japanese population of less than 0.01% (Jmorp). (Table 3) Out of the 24 variants identified, we excluded 1 *DPYD* nonsense variant and one *DPYS* noncoding variant and analyzed the remaining 22 by PolyPhen-2 and SIFT. The results indicated that 18 variants, including all of the rare variants, were predicted to be probably damaging by PolyPhen-2 and/or damaging by SIFT (Table 3). The number of heterozygous and homozygous individuals in the high-toxicity and low-toxicity groups for each of the 24 variants is shown in Table S1.

3.3 | Association of *DPYD*, *DPYS*, and *UPB1* variations with FP-related high toxicity

The *P*-values for each of the variants in relation to an association with FP-related high toxicity are shown in Table 4. In the single

TABLE 1 Baseline characteristics of the study population: n = 301

	Total sample	Low toxicity (grade 0-2) n (%)	High toxicity (grade 3-4) n (%)
Total	301	246	55
Age			
Median	67	66	68
Range	22-85	25-85	22-81
Sex			
Male	179	155 (86.6%)	24 (13.4%)
Female	122	91 (74.6%)	31 (25.4%)
Tumor			
Stomach	70	59 (84.3%)	11 (15.7%)
Colorectal	205	167 (81.5%)	38 (18.5%)
Other tumors	26	20 (76.9%)	6 (23.1%)
5FU + CDDP	17	15 (88.2%)	2 (11.8%)
5FU + L-OHP/CPT-11 + α(molecular target)	164	127 (77.4%)	37 (22.6%)
5FU mono	87	86 (98.9%)	1 (1.1%)
DTX + CDDP+5FU	5	1 (20%)	4 (80%)
FOLFOXIRI + α	13	6 (46.2%)	7 (53.8%)
5FU + PTX	1	1 (100%)	0 (0%)
5FU + GEM/DTX	14	10 (71.4%)	4 (28.6%)

Abbreviations: 5FU, 5 fluorouracil; CDDP, cisplatin; L-OHP, oxaliplatin; CPT-11, irinotecan hydrochloride hydrate; DTX, docetaxel hydrate; PTX, paclitaxel; GEM, gemcitabine hydrochloride.

TABLE 2　Numbers and proportions (%) of patients experiencing different categories of toxicity during the first 2 therapy cycles

Toxicity category	Total sample (n = 301) n (%)	Low toxicity (grade 1-2) (n = 246) n (%)	High toxicity (grade 3-4) (n = 55) n (%)
Gastrointestinal			
Nausea	61 (20.3%)	58 (23.6%)	3 (5.5%)
Vomiting	21 (7%)	18 (7.3%)	3 (5.5%)
Diarrhea	43 (14.3%)	36 (14.6%)	7 (12.7%)
Oral Mucositis	29 (9.6%)	28 (11.4%)	1 (1.8%)
Neutropenia	125 (41.5%)	82 (33.3%)	43 (78.2%)
Hand-foot syndrome	30 (10%)	30 (12.2%)	0 (0%)
Acute kidney injury	10 (3.3%)	10 (4.1%)	0 (0%)

variant analysis, in which we analyzed each variant individually, we did not find any significant association with high toxicity, although a certain trend toward significance was observed for p.Arg181Trp and p.Gln334Arg of *DPYS* (P = .081 and .087). *DPYS* p.Arg181Trp was a common variant and was heterozygous in 14 patients who were all classified as low-toxicity group cases. In contrast, heterozygosity for *DPYS* p.Gln334Arg was observed in only 3 patients, 2 of whom developed high toxicity (Table 4). Clinical and genetic information for these 2 heterozygous *DPYS* p.Gln334Arg patients with high toxicity are provided in Table 5. Both patients received 5FU + L-OHP and developed grade 3 neutropenia. In addition, 1 patient presented with grade 1 vomiting and the other presented with grade 2 nausea. Although both patients were simultaneous carriers of other variants including *DPYD* p.Arg29Cys, *DPYD* p.Ile543Val and *DPYS* c.-1T>C, these additional variants are common benign variants found in the total cohort with an allele frequency of 96%, 27% and 68%, respectively. Hence, our results suggested that p.Gln334Arg may contribute to the susceptibility to severe FP-related toxicity.

We next focused on 7 rare *DPYD* variants that have an allele frequency in the Japanese population of less than 0.01%. Six show loss-of-function by in silico analysis ie probably damaging by PolyPhen-2 and/or deleterious by SIFT (p.Ser199Asn, p.Ile245Phe, p.Thr305Lys, p.Ser556Arg, p.Ala571Asp, p.Trp621Cys), and one is a nonsense mutation (p.Glu386Ter). We divided our patients into a rare pathogenic *DPYD* variant group consisting of individuals heterozygous for these 7 rare variants (n = 7), and a group of all other individuals without these rare variants (n = 294). Using the Fisher exact test, we found that the rare pathogenic *DPYD* variant group showed a significant association with FP-related high toxicity (P = .003; Table 4). Detail information on the 7 patients in the rare *DPYD* variant group is presented in Table 6. Although these 7 patients also carried other variants (*DPYD* p.Arg29Cys, *DPYD* p.Met166Val, *DPYD* p.Ile543Val, *DPYD* p.Thr768Lys and *DPYS* c.-1T>C), these additional variants showed a frequency of more than 1% and appeared benign. In the rare *DPYD* variant group also, 1 patient carrying a heterozygous *DPYD* p.Ala571Asp variant received 5FU monotherapy which is known as a more tolerable chemotherapy protocol, but developed severe toxicity including grade 3 nausea, grade 3 diarrhea, and grade 1 neutropenia.

4 | DISCUSSION

More than 450 *DPYD* variants have been identified to date as a cause of 5FU-related toxicity in patients with cancer.[14] In the context of 5FU, 4 *DPYD* variants identified in the White population are known to have an impact on enzyme function and FP-related toxicity risk.[21] However, none of these *DPYD* variants has been identified to date in an Asian population.[22] Prospective *DPYD* genotyping has thus proved feasible and effective in White but not in Japanese cases. In our present study, we revealed that *DPYD* nonsynonymous variants with allele frequencies of less than 0.01% in the Japanese population, and with an in silico analysis prediction of loss of function, may be associated with severe FP-related toxicity. Our data lend support to the concept that *DPYD* variants exist also in East Asian populations that affect the enzymatic activity of the protein product and thereby the severity of FP-related toxicity. In this study, we found 7 rare *DPYD* variants that were not found in Japanese genome variation databases. For Japanese allele frequency, we used the Jmorp database, which is based on the data of approximately 4000 Japanese individuals mainly living in the northeastern area of Japan. Since our 301 patients were recruited at our hospital at the central area of Japan, the discordance might be due to regional difference in the allele frequency.

The aim of our present study was the establishment of *DPYD* genotype-guided dose individualization of FP therapy in Japanese patients with cancer that have been performed in some western countries. However, we could not find a specific common variant in our present Japanese cohort that was highly associated with FP-related high toxicity. Sequencing of all coding DNA in the *DPYD* gene has some advantages in relation to screening high-risk individuals for severe FP-related toxicity, although it would not seem reasonable to reduce an FP treatment dose based on insufficient in silico findings. It may therefore be difficult to introduce *DPYD* genotyping as useful prospective screening in Japan. Previously, analysis of DPD enzyme activity has been proposed to be the most reliable method for identifying at-risk patients.[23] For the interpretation of a novel or very rare *DPYD* variant, it is useful to measure the DPD activity in the individuals who carry the variants.

TABLE 3 Information and *P*-values for the variants examined in this study

Genotype	dbSNP	In silico function (PolyPhen-2)	In silico function (SIFT)	Allele frequency (%) (Japanese/east Asian/Total)	*P*-value
DPYD					
NM_000110.3:c.85C>T NP_000101.2:p.Arg29Cys	rs1801265	Benign (0)	Tolerated (0.18)	96.85/92.8/76.6	.507
NM_000110.3:c.496A>G NP_000101.2:p.Met166Val	rs2297595	Probably damaging (1)	Tolerated (0.07)	2.18/1.524/8.585	.146
NM_000110.3:c.596G>A NP_000101.2:p.Ser199Asn	rs776973423	Probably damaging (1)	Damaging (0.02)	No data/0/0.006371	.183
NM_000110.3:c.733A>T NP_000101.2:p.Ile245Phe	rs767836989	Possibly damaging (0.853)	Damaging (0)	No data/0/0.004376	.183
NM_000110.3:c.914C>A NP_000101.2:p.Thr305Lys	No number	Probably damaging (0.999)	Damaging (0.01)	No data/no data/no data	1
NM_000110.3:c.1003G>A NP_000101.2:p.Val335Met	rs72549306	Probably damaging (1)	Damaging (0)	0.12/0.01632/0.001989	1
NM_000110.3:c.1156G>T NP_000101.2:p.Glu386Ter	rs78060119			No data/0/0.0007974	.183
NM_000110.3:c.1627A>G NP_000101.2:p.Ile543Val	rs1801159	Benign (0)	Tolerated (0.44)	27.62/25.34/19.52	.974
NM_000110.3:c.1666A>C NP_000101.2:p.Ser556Arg	rs755407188	Probably damaging (1)	Damaging (0)	No data/0.02176/0.001596	1
NM_000110.3:c.1712C>A NP_000101.2:p.Ala571Asp	No number	Probably damaging (1)	Damaging (0)	No data/no data/no data	.183
NM_000110.3:c.1863G>T NP_000101.2:p.Trp621Cys	No number	Probably damaging (1)	Damaging (0)	No data/no data/no data	.183
NM_000110.3:c.2194G>A NP_000101.2:p.Val732Ile	rs1801160	Probably damaging (0.999)	Damaging (0)	19.7/1.887/4.531	.266
NM_000110.3:c.2303C>A NP_000101.2:p.Thr768Lys	rs56005131	Possibly damaging (0.579)	Damaging (0)	24.1/0.236/0.01948	.429
NM_000110.3:c.2476G>A NP_000101.2:p.Val826Met	No number	Probably damaging (0.975)	Damaging (0)	0.14/no data/no data	1
NM_000110.3:c.2678A>G NP_000101.2:p.Asn893Ser	rs188052243	Benign (0)	Tolerated (0.41)	0.22/0.04903/0.003989	1
DPYS					
NM_001385.2:c.-1T>C	rs2959023			69.14/70.45/59.17	.45
NM_001385.2:c.17G>A NP_001376.1:p.Arg6Gln	rs199618701	Benign (0.028)	Damaging (0.02)	0.13/0.3628/0.05538	1
NM_001385.2:c.541C>T NP_001376.1:p.Arg181Trp	rs36027551	Benign (0.024)	Tolerated (0.18)	3.02/5.928/0.9123	.0813
NM_001385.2:c.884A>G NP_001376.1:p.His295Arg	rs996605020	Probably damaging (0.985)	Tolerated (0.27)	No data/no data/no data	1
NM_001385.2:c.1001A>G NP_001376.1:p.Gln334Arg	rs121964923	Probably damaging (1)	Damaging (0)	0.41/0.06516/0.004597	.087
NM_001385.2:c.1253C>T NP_001376.1:p.Thr418Ile	No number	Probably damaging (1)	Damaging (0)	0.01/no data/no data	1
NM_001385.2:c.1469G>A NP_001376.1:p.Arg490His	rs189448963	Probably damaging (1)	Damaging (0)	0.06/0.01504/0.02369	1
UPB1					
NM_016327.2:c.91G>A NP_057411.1:p.Gly31Ser	rs200145797	Probably damaging (1)	Damaging (0)	0.12/0.4612/0.03339	1
NM_016327.2:c.977G>A NP_057411.1:p.Arg326Gln	rs118163237	Probably damaging (1)	Tolerated (0.29)	0.85/2.611/0.192	.671

A recent study has reported the functional characterization of 21 allelic variants of *DPYD* identified in 1070 Japanese individuals.[14] Five of the variants (p.Val335Met, p.Ile543Val, p.Val732Ile, p.Thr768Lys, p.Asn893Ser) identified in our present analysis were among those described in that earlier study. Among these 5 variants, the activity of the p.Val335Met and p.Thr768Lys mutant DPDs exhibited significantly lower intrinsic clearance ($CLint = Vmax/Km$) values compared to the wild-type enzyme (47.4% and 47.9% respectively). However, our present analysis did not find an association between any of these previously reported *DPYD* single variants and severe FP-related toxicity. This may be due to the small number of subjects we analyzed and a further investigation with an increased number of patients is thus warranted to further clarify this issue.

DHP is the second enzyme in the catabolic pathway of uracil and thymine. There are some reports of variants in *DPYS* that may explain the occurrence of severe toxicity from FP-based chemotherapy. For example, c.−1T>C is a common noncoding variant in this gene reported to have an impact on toxicity in patients receiving FP.[24] Our current results have also revealed a high allele frequency of 68% for c.−1T>C, but did not demonstrate a clear relationship between FP-related high toxicity and this variant. With regard to *DPYS* gene coding regions, a prior study has described a patient with severe adverse events from FP therapy harboring the *DPYS* compound heterozygous missense and nonsense variants p.Gly334Arg and p.Arg465Ter.[7] The p.Gln334Arg variant had been previously identified in Japanese patients with DHP deficiency and functional analysis revealed that the corresponding mutant enzyme had only 2.5% residual activity.[17] Until now, it was unknown whether a heterozygous p.Gly334Arg patient would be at a high risk for severe FP-related toxicity, but our current findings have suggested that this might be a possibility. Because the frequency of the p.Gly334Arg is higher in Japanese people than in other ethnic groups (0.41% vs 0.004597% Jmorp, genomeAD), genetic analysis of the *DPYS* gene is important, at least in Japanese patients. Conversely, our present data have indicated that no patients who are heterozygous for *DPYS* p.Arg181Trp developed a severe adverse event following FP treatment. The kinetic parameters of the corresponding mutant enzyme were assessed in a previous report and no markedly reduced activity relative to wild-type DHP was evident.[25] This variant may have protective effects against the development of FP-related toxicity in vivo, but the mechanism is unknown.

The contribution of the some *UPB1* gene alterations to the development of FP-related toxicity was also analyzed previously in White patients with cancer.[1] There have been few reports to date however on coding region variants in this gene. In our previous study, we revealed that the *UPB1* pathogenic variant c.977G>A p.Arg326Gln was prevalent in the Japanese population at a rate of 1.8% but was not found in more than 8000 European and more than 4000 African American alleles.[26] However, the association of this variant with FP-related toxicity is unknown. Our current results found no clear association between this *UPB1* variant and FP-related toxicity, suggesting that a standard regimen with this chemotherapeutic would be tolerated by heterozygous carriers of this pathogenic variant.

Rare *DPYD* variants that cause loss of function in silico and a *DPYS* pathogenic variant p.Gly334Arg may be associated with severe FP-related toxicity in Japanese patients with cancer. However, the common *UPB1* pathogenic variant p.Arg326Gln in the Japanese population does not show a clear association with toxicity in heterozygous individuals.

TABLE 4 Frequency of *DPYS* p.Arg181Trp, *DPYS* p.Gln334Arg and rare pathogenic *DPYD* variants found in the high-toxicity and low-toxicity groups

Genotype	Low toxicity (grade 0-2)	High toxicity (grade 3-4)	Total
DPYS c.541C>T (p.Arg181Trp)			
TT	0	0	0
CT	14	0	14
CC	232	55	287
Total	246	55	301
P-value = .0813			
DPYS c.1001A>G (p.Gln334Arg)			
GG	0	0	0
AG	1	2	3
AA	245	53	298
Total	246	55	301
P-value = .087			
Frequency of patients who had a rare and pathogenic variant of DPYD			
Hetero	2	5	7
Reference	244	50	294
Total	246	55	301
P-value = .0271			

TABLE 5 Clinical and genetic information for 2 heterozygous *DPYS* p.Gln334Arg patients with high toxicity

Patient no.	Age	Sex	Cancer	Regimen	Side effects	Other DPYD variants	Other DPYS variants	Other UPB1 variants
Patient 1	68	Female	Colorectal	5FU + L-OHP	Vomiting Grade 1 Neutropenia Grade 3	p.Ile543Val het p.Arg29Cys hom	c.-1T>C hom	No variant
Patient 2	81	Female	Colorectal	5FU + L-OHP	Nausea Grade 2 Neutropenia Grade 3	p.Arg29Cys het	c.-1T>C hom	No variant

TABLE 6 Detailed information on the 7 patients in the rare pathogenic *DPYD* variant group

Rare pathogenic variant	Age	Sex	Cancer	Regimen	Side effects	Other DPYD variants	Other DPYS variants	Other UPB1 variants
c.596G>A (p.Ser199Asn)	67	Female	Colorectal	5FU + L-OHP	Vomiting Grade 3 Diarrhea Grade 3 Neutropenia Grade 3	p.Ile543Val hom p.Arg29Cys hom	c.-1T>C het	No variant
c.733A>T (p.Ile245Phe)	63	Female	Stomach	5FU + L-OHP	Nausea Grade 2 Vomiting Grade 2 Diarrhea Grade 3	p.Ile543Val het p.Met166Val het p.Arg29Cys hom	c.-1T>C hom	No variant
c.914C>A (p.Thr305Lys)	66	Female	Colorectal	5FU + L-OHP	Diarrhea Grade 1 Oral Mucositis Grade 1 Hand-foot syndrome Grade 1 Neutropenia Grade 2	p.Arg29Cys hom	c.-1T>C het	No variant
c.1156G>T (p.Glu386Ter)	48	Male	Colorectal	FOLFOXIRI + α	Nausea Grade 1 Neutropenia Grade 3	p.Arg29Cys hom	c.-1T>C hom	No variant
c.1666A>C (p.Ser556Arg)	25	Male	Colorectal	5FU + L-OHP	Nausea Grade 1 Vomiting Grade 1 Oral Mucositis Grade 1 Neutropenia Grade 1	p.Arg29Cys hom	c.-1T>C hom	No variant
c.1712C>A (p.Ala571Asp)	70	Female	Colorectal	5FU mono	Nausea Grade 3 Diarrhea Grade 3 Neutropenia Grade 1	p.Arg29Cys het	c.-1T>C hom	No variant
c.1863G>T (p.Trp621Cys)	72	Female	Colorectal	5FU + L-OHP	Nausea Grade 1 Neutropenia Grade 3	p.Ile543Val het p.Arg29Cys hom p.Thr768Lys het	c.-1T>C hom	No variant

ETHICAL APPROVAL

We obtained approval for this study from the Ethical Review Board for Human Genome Studies at Fujita Health University. Written informed consent was obtained from all patients. All experiments were carried out in accordance with the relevant guidelines and regulations.

ORCID

Katsuyuki Yokoi https://orcid.org/0000-0002-4760-0420

REFERENCES

1. Fidlerova J, Kleiblova P, Kormunda S, Novotny J, Kleibl Z. Contribution of the β-ureidopropionase (UPB1) gene alterations to the development of fluoropyrimidine-related toxicity. *Pharmacol Rep.* 2012;64(5):1234-1242.
2. Jennings BA, Loke YK, Skinner J, et al. Evaluating predictive pharmacogenetic signatures of adverse events in colorectal cancer patients treated with fluoropyrimidines. *PLoS One.* 2013;8(10):e78053.
3. Longley DB, Harkin DP, Johnston PG. 5-fluorouracil: mechanisms of action and clinical strategies. *Nat Rev Cancer.* 2003;3(5):330-338.
4. Kuilenburg ABPV, Meijer J, Tanck MWT, et al. Phenotypic and clinical implications of variants in the dihydropyrimidine dehydrogenase gene. *Biochim Biophys Acta.* 2016;1862(4):754-762.
5. Lunenburg CATC, Henricks LM, Guchelaar HJ, et al. Prospective DPYD genotyping to reduce the risk of fluoropyrimidine-induced severe toxicity: Ready for prime time. *Eur J Cancer.* 2016;54:40-48.
6. Wigle TJ, Tsvetkova EV, Welch SA, Kim RB. DPYD and fluorouracil-based chemotherapy: mini review and case report. *Pharmaceutics.* 2019;11(5):199.
7. Hiratsuka M, Yamashita H, Akai F, et al. Genetic polymorphisms of dihydropyrimidinase in a Japanese patient with capecitabine-induced toxicity. *PLoS One.* 2015;10(4):e0124818.
8. Loganayagam A, Arenas Hernandez M, Corrigan A, et al. Pharmacogenetic variants in the DPYD, TYMS, CDA and MTHFR genes are clinically significant predictors of fluoropyrimidine toxicity. *Br J Cancer.* 2013;108(12):2505-2515.
9. van Kuilenburg AB, Meinsma R, Zonnenberg BA, et al. Dihydropyrimidinase deficiency and severe 5-fluorouracil toxicity. *Clin Cancer Res.* 2003;9(12):4363-4367.
10. Henricks LM, Lunenburg CATC, de Man FM, et al. DPYD genotype-guided dose individualisation of fluoropyrimidine therapy in patients with cancer: a prospective safety analysis. *Lancet Oncol.* 2018;19(11):1459-1467.
11. Lu ZH, Zhang R, Diasio RB. Purification and characterization of dihydropyrimidine dehydrogenase from human liver. *J Biol Chem.* 1992;267(24):17102-17109.
12. Meulendijks D, Henricks LM, Sonke GS, et al. Clinical relevance of DPYD variants c.1679T>G, c.1236G>A/HapB3, and c.1601G>A as predictors of severe fluoropyrimidine-associated toxicity: a systematic review and meta-analysis of individual patient data. *Lancet Oncol.* 2015;16(16):1639-1650.
13. van Kuilenburg AB, Meijer J, Maurer D, et al. Severe fluoropyrimidine toxicity due to novel and rare DPYD missense

mutations, deletion and genomic amplification affecting DPD activity and mRNA splicing. *Biochim Biophys Acta Mol Basis Dis.* 2017;1863(3):721-730.

14. Hishinuma E, Narita Y, Saito S, et al. Functional Characterization of 21 Allelic Variants of Dihydropyrimidine Dehydrogenase Identified in 1070 Japanese Individuals. *Drug Metab Dispos.* 2018;46(8):1083-1090.

15. Nagasaki M, Yasuda J, Katsuoka F, et al. Rare variant discovery by deep whole-genome sequencing of 1,070 Japanese individuals. *Nat Commun.* 2015;21(6):8018.

16. Kummer D, Froehlich TK, Joerger M, et al. Dihydropyrimidinase and β-ureidopropionase gene variation and severe fluoropyrimidine-related toxicity. *Pharmacogenomics.* 2015;16(12):1367-1377.

17. Hamajima N, Kouwaki M, Vreken P, et al. Dihydropyrimidinase deficiency: structural organization, chromosomal localization, and mutation analysis of the human dihydropyrimidinase gene. *Am J Hum Genet.* 1998;63(3):717-726.

18. Vreken P, van Kuilenburg AB, Hamajima N, et al. cDNA cloning, genomic structure and chromosomal localization of the human BUP-1 gene encoding beta-ureidopropionase. *Biochim Biophys Acta.* 1999;1447(2-3):251-257.

19. Kuhara T, Ohse M, Inoue Y, Shinka T. Five cases of beta-ureidopropionase deficiency detected by GC/MS analysis of urine metabolome. *J Mass Spectrom.* 2009;44(2):214-221.

20. Nakajima Y, Meijer J, Dobritzsch D, et al. Dihydropyrimidinase deficiency in four East Asian patients due to novel and rare DPYS mutations affecting protein structural integrity and catalytic activity. *Mol Genet Metab.* 2017;122(4):216-222.

21. Froehlich TK, Amstutz U, Aebi S, Joerger M, Largiadèr CR. Clinical importance of risk variants in the dihydropyrimidine dehydrogenase gene for the prediction of early-onset fluoropyrimidine toxicity. *Int J Cancer.* 2015;136(3):730-739.

22. Maekawa K, Saeki M, Saito Y, et al. Genetic variations and haplotype structures of the DPYD gene encoding dihydropyrimidine dehydrogenase in Japanese and their ethnic differences. *J Hum Genet.* 2007;52(10):804-819.

23. Thomas F, Hennebelle I, Delmas C, et al. Genotyping of a family with a novel deleterious DPYD mutation supports the pretherapeutic screening of DPD deficiency with dihydrouracil/uracil ratio. *Clin Pharmacol Ther.* 2016;99(2):235-242.

24. Fidlerova J, Kleiblova P, Bilek M, et al. Contribution of dihydropyrimidinase gene alterations to the development of serious toxicity in fluoropyrimidine-treated cancer patients. *Cancer Chemother Pharmacol.* 2010;65(4):661-669.

25. Hishinuma E, Akai F, Narita Y, et al. Functional characterization of 21 allelic variants of dihydropyrimidinase. *Biochem Pharmacol.* 2017;1(143):118-128.

26. Nakajima Y, Meijer J, Dobritzsch D, et al. biochemical and molecular analysis of 13 Japanese patients with β-ureidopropionase deficiency demonstrates high prevalence of the c.977G>A (p. R326Q) mutation. *J Inherit Metab Dis.* 2014;37(5):801-812.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

MAJOR ARTICLE

Inherited Chromosomally Integrated Human Herpesvirus 6 Is a Risk Factor for Spontaneous Abortion

Hiroki Miura,[1] Yoshiki Kawamura,[1] Tamae Ohye,[2] Fumihiko Hattori,[1] Kei Kozawa,[1] Masaru Ihira,[3] Hiroshi Yatsuya,[4] Haruki Nishizawa,[5] Hiroki Kurahashi,[6] and Tetsushi Yoshikawa[1]

[1]Department of Pediatrics, Fujita Health University School of Medicine, Toyoake, Japan, [2]Department of Clinical Laboratory Medicine, Graduate School of Health Sciences, Fujita Health University, Toyoake, Japan, [3]Faculty of Clinical Engineering, Fujita Health University School of Health Sciences, Toyoake, Japan, [4]Department of Public Health, Fujita Health University School of Medicine, Toyoake, Japan, [5]Department of Obstetrics and Gynecology, Fujita Health University School of Medicine, Toyoake, Japan, and [6]Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Japan

Background. Human herpesvirus 6 (HHV-6) can be genetically transmitted from parent to child as inherited chromosomally integrated HHV-6 (iciHHV-6). HHV-6 reactivation occurs in pregnant women with iciHHV-6. We found no sex differences in the frequency of index cases with iciHHV-6 but inheritance from the father was more common. We evaluated the association between iciHHV-6 status and spontaneous abortion.

Methods. iciHHV-6 was confirmed by high viral DNA copy numbers in whole blood and somatic cells. The origin of integrated viral genome, paternal or maternal, was examined using the same method. The pregnancy history of 23 mothers in families with iciHHV-6 and 285 mothers in families without iciHHV-6 was abstracted.

Results. Of 23 iciHHV-6 index cases, 8 mothers and 15 fathers had iciHHV-6. Spontaneous abortion rates in mothers with and mothers without/fathers with iciHHV-6 and mothers in families without iciHHV-6 were 27.6%, 10.3%, and 14.8%, respectively ($P = .012$). Mothers with iciHHV-6 (odds ratio [OR], 6.41; 95% confidence interval [CI], 1.10–37.4) and maternal age at the most recent pregnancy ≥40 years (OR, 3.91; 95% CI, 1.30–11.8) were associated with 2 or more spontaneous abortions.

Conclusions. Mothers with iciHHV-6 is a risk factor for spontaneous abortion.

Keywords. iciHHV-6; spontaneous abortion; pregnancy.

Human herpesvirus 6 (HHV-6) was the sixth human herpesvirus discovered. It belongs to the β-Herpesvirinae subfamily. This virus is categorized into 2 distinct species (HHV-6A and HHV-6B) with 90% overall nucleotide sequence identity [1, 2]. Primary HHV-6B infection generally occurs in infancy and causes exanthema subitum, a common febrile exanthematous disease [3, 4]. In contrast to HHV-6B, the clinical features of primary HHV-6A infection remain unclear. Although HHV-6B is a ubiquitous virus throughout the world, HHV-6A infection is considered to be endemic only in a limited geographical area [5].

Although horizontal transmission is considered to be the main route of HHV-6 infection, it can be genetically transmitted from parent to child as inherited chromosomally integrated HHV-6 (iciHHV-6) [6, 7]. The complete HHV-6 genome is integrated into every nucleated cell of an individual with iciHHV-6. Extremely high copy numbers of HHV-6 DNA can

be detected in clinical specimens containing nucleated cells, which can lead to a misdiagnosis of active viral infection [8]. The frequency of iciHHV-6 in the general population is approximately 0.8%–1.5% in the United Kingdom [9, 10] compared with 0.2%–0.6% in Japan [6, 7]. This low frequency has hampered the elucidation of the link between iciHHV-6 and disease.

Previous studies have suggested that iciHHV-6 is associated with clinical manifestations such as encephalitis [11, 12], cognitive dysfunction, fatigue [13, 14], and angina pectoris [15]. However, the precise relationship between iciHHV-6 and these conditions remains under investigation. In addition, viral reactivation from the integrated viral genome has been demonstrated in an immunocompromised patient with iciHHV-6 [16] and pregnant women with iciHHV-6 [17, 18]. Moreover, viral reactivation can be induced in iciHHV-6 cells by some chemical compound treatments [19]. These findings suggest that iciHHV-6 may cause physiological changes in the host through viral reactivation from the integrated viral genome. For example, hematopoietic stem cell transplant recipients with iciHHV-6 have been shown to have an increased frequency of acute graft-versus-host disease [20].

Theoretically, a parent with iciHHV-6 has a 50% chance of transmitting the integrated HHV-6 genome to the next generation. Our previous study demonstrated no sex difference among index cases with iciHHV-6. However, when we examined the

origin of the viral genome, we noticed that there are many cases of paternal origin [21]. Therefore, to determine if paternal origin is more common than maternal origin and to evaluate the association between iciHHV-6 and spontaneous abortion, we sought to examine the origin of the HHV-6 genome and pregnancy history of mothers in families with iciHHV-6.

METHODS

Subjects and Samples

Individuals with iciHHV-6 with a confirmed origin of the viral genome were enrolled in this study. Index cases were diagnosed with iciHHV-6 based on the following procedures. Screening for HHV-6 was performed using peripheral whole blood or serum obtained from patients with febrile illness enrolled in another clinical study at Fujita Health University (ethics committee accession number 14-096). In addition, harvested cord blood from healthy neonates born at Fujita Health University Hospital underwent HHV-6 screening. Furthermore, whole blood or serum samples obtained from patients with suspected iciHHV-6 at other hospitals were also analyzed.

Screening and Criteria for iciHHV-6

To screen for HHV-6 infection, DNA was extracted from 200 μL of whole blood or serum samples using the QIAamp DNA Blood Mini kit (QIAGEN). DNA samples were stored at −30°C until polymerase chain reaction (PCR) analysis. HHV-6 DNA loads were measured in these samples using real-time PCR [22]. If a high copy number of HHV-6 DNA (whole blood, $>5.5 \times 10^5$ copies/mL; serum, $>1.0 \times 10^5$ copies/mL) was detected, the subject was suspected of having iciHHV-6. HHV-6 species were determined by restriction fragment length polymorphism (RFLP) analysis of loop-mediated isothermal amplification products as previously described [23].

If a subject was suspected of having iciHHV-6, additional whole blood, hair follicle, or buccal mucosa swab samples were collected for further investigation. Whole blood, hair follicle, or buccal mucosa swab samples were used to measure HHV-6 DNA load with real-time PCR analysis. Whole blood samples were used for fluorescent in situ hybridization (FISH) assays to identify the integration site. A subject was defined as having iciHHV-6 if either of the following 2 criteria was fulfilled: (1) detection of more than 5.5×10^5 copies/mL of HHV-6 DNA in whole blood and HHV-6 DNA in somatic cells (hair follicle or buccal mucosa swab samples) or (2) confirmation of HHV-6 DNA integration in a chromosome by FISH.

To determine the origin of the endogenous HHV-6 genome, whole blood, hair follicle, or buccal mucosa swab samples were collected from the parents of the index cases. The samples were used to measure HHV-6 DNA load and for FISH analysis. The origin of iciHHV-6 was confirmed using the criteria described above.

FISH Analysis for Integration Site Determination

To identify the site of chromosomal integration, 5 mL of heparinized peripheral blood was collected from subjects with iciHHV-6 to perform FISH analysis as previously described [24]. Briefly, phytohemagglutinin-stimulated lymphocytes or Epstein-Barr virus-transformed lymphoblasts were arrested by treatment with colcemid. Metaphase preparations were obtained via hypotonic treatment with 0.075 M KCl followed by methanol and acetate fixation. PCR products from HHV-6 sequences were used as probes for FISH, which were labelled by nick translation with biotin-16-dUTP or digoxigenin-11-dUTP. After hybridization, the probes were detected using either Alexa Fluor 488-conjugated streptavidin or rhodamine-conjugated antidigoxigenin and visualized through counter-staining with 4′,6-diamino-2-phenylindole. Chromosome-specific probes were used as reference standards.

Pregnancy History of Mothers in Families with iciHHV-6

Pregnancy history was obtained from mothers in families with iciHHV-6 and 285 women in families without iciHHV-6 as a control group for comparison of the incidence of spontaneous abortion. Control subjects had given birth at Fujita Health University Hospital between December 2017 and March 2019. Pregnancy histories of cases and controls who could be reviewed from medical records were abstracted and those of the others were elicited by telephone interview. These reviews and interviews were performed by a person blinded to the subjects' iciHHV-6 status. All cord blood samples obtained from control subjects did not contain HHV-6 DNA.

Statistical Analysis

The sex of iciHHV-6 index cases and the maternal/paternal origin were compared using the binomial test for differences in proportion. The number of spontaneous abortions per mother and maternal age at the most recent pregnancy in the iciHHV-6 versus non–iciHHV-6 groups were evaluated using the Kruskal-Wallis test. The proportion of mothers having 2 or more spontaneous abortions in the 2 groups was evaluated using Fisher exact test. Univariable associations between potential factors described below and history of spontaneous abortion were evaluated using Fisher exact test. We analyzed the association between 2 or more spontaneous abortion and the following factors: iciHHV-6 family, iciHHV-6 species, iciHHV-6 integration site, mother with iciHHV-6, and maternal age at the most recent pregnancy. Variables that were statistically significantly associated with spontaneous abortion in univariable analyses were included in multivariable logistic regression analyses. Statistical significance was defined as a 2-sided $P < .05$. JMP version 13.1 (SAS Institute) was used for analyses.

Ethics Approval

This study was approved by the Ethics Review Board for Human Genome Studies at Fujita Health University (accession number HG19-006), 12 June 2019. If the subjects had suspected iciHHV-6 based on HHV-6 DNA loads in blood samples, written informed consent was obtained from subjects or their guardians to carry out examinations for definitive diagnosis of iciHHV-6 and collection of clinical information.

RESULTS

Characteristics of Index Cases With iciHHV-6

Twenty-three iciHHV-6 index cases were enrolled in this study. Sex, iciHHV-6 origin, iciHHV-6 integration site, iciHHV-6 species, and maternal pregnancy history are summarized in Table 1. No sex difference was observed among index cases ($P = .84$) but paternal origin accounted for 68% of iciHHV-6 cases ($P = .088$). The site of integration was determined in 21 of 23 subjects with iciHHV-6 using FISH analysis. One copy of the HHV-6 genome was integrated in 20 of 21 cases (22q, 11 cases; Xp, 2 cases; Yp, 2 cases; 1q, 2 cases; 16q, 1 case; 17p, 1 case; and 20p, 1 case). Two copies of the HHV-6 genome were integrated in patient 8; 1 copy was derived from each parent. The copy integrated on chromosome 20p was of paternal origin whereas the copy integrated on chromosome 22q was of maternal origin.

RFLP analysis indicated that 11 (48%) of the 23 iciHHV-6 cases were HHV-6A and 12 (52%) were HHV-6B.

Regarding the birth history of the 8 index cases whose mothers had iciHHV-6, 1 was preterm, 1 had low birth weight, and 1 was preterm and had low birth weight. The remaining 5 index cases had no abnormalities at the time of birth. In the 15 index cases whose mothers did not have iciHHV-6, 1 was preterm and had low birth weight and another was born through emergency cesarean delivery because of reduced baseline variability. The remaining 13 index cases had no abnormalities.

Maternal Pregnancy History by iciHHV-6 Status

In 23 families with iciHHV-6, 8 mothers with iciHHV-6 (patients 1 to 8) had 8 spontaneous abortions in 29 pregnancies whereas 15 mothers without/fathers with iciHHV-6 (patients 9 to 23) had 3 spontaneous abortions in 29 pregnancies (Table 1). In the control group, 285 mothers in families without iciHHV-6 had 87 spontaneous abortions in 589 pregnancies. The spontaneous abortion rate and numbers of mothers having 2 or more spontaneous abortions were compared among 3 groups (mothers with iciHHV-6, mothers without/fathers with iciHHV-6, and control mothers in families without iciHHV-6) (Table 2). Spontaneous abortion rates among mothers with iciHHV-6, mothers without/fathers with iciHHV-6 and control

Table 1. Characteristics of Index Cases With iciHHV-6 and Maternal Pregnancy History

Patient	Sex	Origin[a]	Integration Site	iciHHV6 Species	No. of Pregnancies	No. of Spontaneous Abortions	Maternal Age at Most Recent Pregnancy, y
1	M	Maternal	22q	HHV-6A	5	2	NA
2	F	Maternal	22q	HHV-6A	3	1	35
3	F	Maternal	22q	HHV-6A	4	1	25
4	F	Maternal	22q	HHV-6A	2	0	34
5	M	Maternal	22q	HHV-6A	2	0	33
6	M	Maternal	16p	HHV-6B	2	0	32
7	M	Maternal	20p	HHV-6B	6	2	40
8	F	Paternal and maternal	20p and 22q	HHV-6B	5	2	32
9	F	Paternal	22q	HHV-6A	3	0	34
10	F	Paternal	22q	HHV-6A	1	0	25
11	F	Paternal	22q	HHV-6A	2	0	25
12	M	Paternal	22q	HHV-6A	2	0	NA
13	M	Paternal	22q	HHV-6B	2	1	30
14	F	Paternal	22q	HHV-6B	1	0	NA
15	F	Paternal	Xp	HHV-6B	2	0	27
16	M	Paternal	Xp	HHV-6B	2	0	NA
17	M	Paternal	Yp	HHV-6B	4	1	34
18	M	Paternal	Yp	HHV-6B	2	0	39
19	F	Paternal	1q	HHV-6B	2	1	26
20	M	Paternal	1q	HHV-6B	2	0	NA
21	M	Paternal	17p	HHV-6A	1	0	30
22	M	Paternal	ND	HHV-6A	1	0	36
23	M	Paternal	ND	HHV-6B	2	0	32

Abbreviations: [a]Paternal or maternal origin of chromosomally integrated HHV-6.

Table 2. Prevalence of Spontaneous Abortion by iciHHV-6 Status

Status	Spontaneous Abortion Rate, %	No. of Mothers With 2 or More Spontaneous Abortions, n (%)	Maternal Age at Most Recent Pregnancy, y, Median (IQR)
Mother with iciHHV-6 (n = 8)	27.6	3 (37.5)	33.0 (28.0–36.0)
Mother without/father with iciHHV-6 (n = 15)	10.3	0 (.0)	30.0 (26.5–34.0)
Mother in family without iciHHV-6 (n = 285)	14.8	18 (6.3)	33.0 (32.0–34.5)
P value	**.012**	**.002**	.46

Bold values are statistically significant.

Abbreviations: iciHHV-6, inherited chromosomally integrated human herpesvirus 6; IQR, interquartile range.

mothers in families without iciHHV-6 were 27.6% (8/29), 10.3% (3/29), and 14.8% (87/589), respectively (P = .012). Because recurrent pregnancy loss is defined as 2 or more spontaneous abortions, the number of mothers with a history of 2 or more abortions was compared among the 3 groups. Of 8 mothers with iciHHV-6, 3 (37.5%) had repeated spontaneous abortions. None (0%) of the mothers without/fathers with iciHHV-6 had repeated spontaneous abortions and 18 of 285 (6.3%) mothers in the control group had repeated spontaneous abortions. There was a statistically significant difference among the 3 groups (P = .002). Maternal age at the most recent pregnancy was not significantly different among the 3 groups (P = .46).

Risk Factors for Spontaneous Abortion

To elucidate potential risk factors for 2 or more spontaneous abortions, we examined characteristics associated with iciHHV-6 status and maternal age at the most recent pregnancy (Table 3). Univariable analyses indicated that iciHHV-6 family status did not affect the risk of spontaneous abortion (P = .20), but mother with iciHHV-6 and maternal age at the most recent

pregnancy of 40 years or older were associated with 2 or more spontaneous abortions (P = .012 and P = .021, respectively). Among 23 families with iciHHV-6, there was no correlation between spontaneous abortion and HHV-6 species (P = 1.00) and HHV-6 integration site (P = 1.00). Multivariable analyses revealed that mother with iciHHV-6 (odds ratio [OR], 6.41; 95% confidence interval [CI], 1.10–37.4; P = .001) and maternal age at the most recent pregnancy of 40 years or older (OR, 3.91; 95% CI, 1.30–11.8; P = .016) were significantly associated with 2 or more spontaneous abortions (Table 3).

DISCUSSION

Similar to our previous study [7], there was no sex difference in index cases with iciHHV-6, but iciHHV-6 of maternal origin was lower than expected (32% instead of 50%) in this study. In contrast to our studies, Hall et al [25] (who first demonstrated that HHV-6 can be genetically transmitted from parent to child) reported that 60% (18/30) of children with iciHHV-6 inherited it from their mother. An insufficient number of cases may be the main reason for this

Table 3. Risk Factors for 2 or More Spontaneous Abortions

Risk Factor	Univariable		Multivariable	
	Odds Ratio (95% CI)	P Value	Odds Ratio (95% CI)	P Value
iciHHV-6 family				
No (n = 285)	Ref			
Yes (n = 23)	2.23 (.60–8.20)	.20		
iciHHV-6 species				
iciHHV-6A (n = 11)	Ref			
iciHHV-6B (n = 12)	2.00 (.16–25.8)	1.00		
iciHHV-6 integration site				
Other (n = 9)	Ref			
22q (n = 12)	2.00 (.16–25.8)	1.00		
Mother with iciHHV-6				
No (n = 300)	Ref		Ref	
Yes (n = 8)	9.40 (2.08–42.5)	**.012**	6.41 (1.10–37.4)	**.040**
Maternal age at most recent pregnancy, y				
<35 (n = 194)	Ref		Ref	
35–39 (n = 74)	1.05 (.32–3.46)	1.00	1.11 (.33–3.68)	.87
≥40 (n = 35)	3.81 (1.29–11.3)	**.021**	3.91 (1.30–11.8)	**.016**

Bold values are statistically significant.

Abbreviations: CI, confidence interval; iciHHV-6, inherited chromosomally integrated human herpesvirus 6; Ref, reference.

discrepancy, but the exact reason is unknown. Common viral genome integration sites are different between Japan and Western countries; 22q is the most common site in Japan and 17p is the most common site in Western countries [7, 26]. Nine of 10 index case of iciHHV-6A had the 22q integration site in this study. This pattern has been found in the majority of Japanese cases and our previous study suggested that iciHH-6A in 22q likely derived from a common ancestor [27]. Furthermore, genetic diversity of the viral genome in each region has been demonstrated by phylogenetic tree analysis of integrated viral genomes from different regions [28]. Thus, it is possible that differences in chromosome of integration and genetic diversity of the integrated viral genome may be associated with different patterns of inheritance between Japanese and the Western populations.

Spontaneous abortion rates in mothers without/fathers with iciHHV-6 and mothers in families without iciHHV-6 were 10.3% and 14.8%, respectively, which is consistent with findings from previous studies [29, 30]. The spontaneous abortion rate and number of mothers with repeated spontaneous abortions were significantly higher in mothers with iciHHV-6 than in the other 2 groups ($P = .012$ and $P = .002$). In addition, univariable analyses showed that iciHHV-6 family status, integration in chromosome 22q, and HHV-6 species did not increase the risk of spontaneous abortion, but maternal age at the most recent pregnancy of 40 years or older and mother with iciHHV-6 were the only risk factors for 2 or more spontaneous abortions identified by multivariable analyses. Because older maternal age is an expected risk factor for spontaneous abortion [31], the reliability of this study is considered to be high. Furthermore, it is well known that older maternal age, genetic abnormalities, selected maternal autoantibodies, endocrine dysfunction, and uterine abnormalities are risk factors for spontaneous abortion [32]. However, the cause of approximately 50% of recurrent spontaneous abortions is unknown. Therefore, the present findings may help explain the cause of some unexplained recurrent pregnancy loss.

It is necessary to elucidate mechanisms by which iciHHV-6 increases the risk of spontaneous abortion. This study suggests that mother with iciHHV-6 can be a risk factor for spontaneous abortion but fetal iciHHV-6 is not, because in families with iciHHV-6, father with iciHHV-6 was not a risk factor. Thus, maternal factors may play an important role in iciHHV-6–related spontaneous abortion. Results of previous studies on the association between maternal HHV-6 infection and spontaneous abortion may be helpful for future research. It has been suggested that active HHV-6 infection during pregnancy may cause spontaneous abortion [33, 34] as well as premature delivery with neonatal hypotonia [35]. Furthermore, a series of Italian studies has demonstrated that some infertile women have HHV-6A infection in endometrial epithelial cells accompanied by alterations in the maternal immune system [36–38]. Their in vitro study demonstrated that endometrial cells infected with HHV-6A are less permissive to the attachment of trophoblast cells and microRNA expression patterns of endometrial cells were altered by HHV-6A infection [39].

As described above, HHV-6 reactivation from integrated viral genomes has been demonstrated in both in vivo and in vitro studies [16–19]. Recently, HHV-6 transcripts have been detected in placentas of pregnant women with iciHHV-6 and preeclampsia, and the risk of preeclampsia was significantly higher in mothers with iciHHV-6 than control mothers [40]. Detection of HHV-6 mRNA in placentas with iciHHV-6 suggests viral reactivation in placental tissue. Taken together with the results of this study, HHV-6 reactivation in pregnant women with iciHHV-6 may cause spontaneous abortion or preeclampsia. There are 2 possible mechanisms by which reactivation of HHV-6 from an integrated viral genome can affect the host cells: direct cellular damage induced by viral infection and bystander effects induced by viral reactivation. Recent studies have suggested that iciHHV-6 can exacerbate acute graft-versus-host disease via upregulation of inflammatory cytokines [41]. Furthermore, Kumata et al reported that host transcripts in the certain tissues are modified by iciHHV-6 [42]. Although a prospective study might be difficult because of the limited number of cases, a study following pregnant women with iciHHV-6 is necessary to explore the mechanisms underlying iciHHV-6–related spontaneous abortion. In addition, although no notable clinical manifestations were observed in neonates with iciHHV-6 delivered from mothers with iciHHV-6 in this study, more neonates with iciHHV-6 should be examined and followed for longer periods to understand the effects of iciHHV-6 during the embryonic period.

In conclusion, although there was no sex difference among iciHHV-6 index cases, the frequency of iciHHV-6 inherited from the mother was lower than expected in this study. In addition to maternal age at the most recent pregnancy of 40 years or older, mother with iciHHV-6 was a risk factor for 2 or more spontaneous abortions. Therefore, careful management may be required for pregnant women with iciHHV-6, who may be predisposed to spontaneous abortion.

Notes

editors consider relevant to the content of the manuscript have been disclosed.

References

1. Dominguez G, Dambaugh TR, Stamey FR, Dewhurst S, Inoue N, Pellett PE. Human herpesvirus 6B genome sequence: coding content and comparison with human herpesvirus 6A. J Virol **1999**; 73:8040–52.

2. Isegawa Y, Mukai T, Nakano K, et al. Comparison of the complete DNA sequences of human herpesvirus 6 variants A and B. J Virol **1999**; 73:8053–63.

3. Yamanishi K, Okuno T, Shiraki K, et al. Identification of human herpesvirus-6 as a causal agent for exanthem subitum. Lancet **1988**; 1(8594):1065–7.

4. Asano Y, Suga S, Yoshikawa T, Urisu A, Yazaki T. Human herpesvirus type 6 infection (exanthem subitum) without fever. J Pediatr **1989**; 115:264–5.

5. Bates M, Monze M, Bima H, et al. Predominant human herpesvirus 6 variant A infant infections in an HIV-1 endemic region of Sub-Saharan Africa. J Med Virol **2009**; 81:779–89.

6. Tanaka-Taya K, Sashihara J, Kurahashi H, et al. Human herpesvirus 6 (HHV-6) is transmitted from parent to child in an integrated form and characterization of cases with chromosomally integrated HHV-6 DNA. J Med Virol **2004**; 73:465–73.

7. Miura H, Kawamura Y, Hattori F, et al. Chromosomally integrated human herpesvirus 6 in the Japanese population. J Med Virol **2018**; 90:1636–42.

8. Lee SO, Brown RA, Razonable RR. Chromosomally integrated human herpesvirus-6 in transplant recipients. Transpl Infect Dis **2012**; 14:346–54.

9. Leong HN, Tuke PW, Tedder RS, et al. The prevalence of chromosomally integrated human herpesvirus 6 genomes in the blood of UK blood donors. J Med Virol **2007**; 79:45–51.

10. Ward KN, Thiruchelvam AD, Couto-Parada X. Unexpected occasional persistence of high levels of HHV-6 DNA in sera: detection of variants A and B. J Med Virol **2005**; 76:563–70.

11. Troy SB, Blackburn BG, Yeom K, Caulfield AK, Bhangoo MS, Montoya JG. Severe encephalomyelitis in an immunocompetent adult with chromosomally integrated human herpesvirus 6 and clinical response to treatment with foscarnet plus ganciclovir. Clin Infect Dis **2008**; 47:e93–6.

12. Wittekindt B, Berger A, Porto L, et al. Human herpes virus-6 DNA in cerebrospinal fluid of children undergoing therapy for acute leukaemia. Br J Haematol **2009**; 145:542–5.

13. Montoya JG, Neely MN, Gupta S, et al. Antiviral therapy of two patients with chromosomally-integrated human herpesvirus-6A presenting with cognitive dysfunction. J Clin Virol **2012**; 55:40–5.

14. Pantry SN, Medveczky MM, Arbuckle JH, et al. Persistent human herpesvirus-6 infection in patients with an inherited form of the virus. J Med Virol **2013**; 85:1940–6.

15. Gravel A, Dubuc I, Morissette G, Sedlak RH, Jerome KR, Flamand L. Inherited chromosomally integrated human herpesvirus 6 as a predisposing risk factor for the development of angina pectoris. Proc Natl Acad Sci U S A **2015**; 112:8058–63.

16. Endo A, Watanabe K, Ohye T, et al. Molecular and virological evidence of viral activation from chromosomally integrated human herpesvirus 6A in a patient with X-linked severe combined immunodeficiency. Clin Infect Dis **2014**; 59:545–8.

17. Hall CB, Caserta MT, Schnabel KC, et al. Transplacental congenital human herpesvirus 6 infection caused by maternal chromosomally integrated virus. J Infect Dis **2010**; 201:505–7.

18. Gravel A, Hall CB, Flamand L. Sequence analysis of transplacentally acquired human herpesvirus 6 DNA is consistent with transmission of a chromosomally integrated reactivated virus. J Infect Dis **2013**; 207:1585–9.

19. Arbuckle JH, Medveczky MM, Luka J, et al. The latent human herpesvirus-6A genome specifically integrates in telomeres of human chromosomes in vivo and in vitro. Proc Natl Acad Sci U S A **2010**; 107:5563–8.

20. Hill JA, Magaret AS, Hall-Sedlak R, et al. Outcomes of hematopoietic cell transplantation using donors or recipients with inherited chromosomally integrated HHV-6. Blood **2017**; 130:1062–9.

21. Miura H, Kawamura Y, Kudo K, et al. Virological analysis of inherited chromosomally integrated human herpesvirus-6 in three hematopoietic stem cell transplant patients. Transpl Infect Dis **2015**; 17:728–31.

22. Tanaka N, Kimura H, Hoshino Y, et al. Monitoring four herpesviruses in unrelated cord blood transplantation. Bone Marrow Transplant **2000**; 26:1193–7.

23. Ihira M, Ohta A, Sugata K, Suga S, Asano Y, Yoshikawa T. Loop-mediated isothermal amplification for discriminating between human herpesvirus 6 A and B. J Virol Methods **2008**; 154:223–5.

24. Ohye T, Kawamura Y, Inagaki H, et al. A simple cytogenetic method to detect chromosomally integrated human herpesvirus-6. J Virol Methods **2016**; 228:74–8.

25. Hall CB, Caserta MT, Schnabel K, et al. Chromosomal integration of human herpesvirus 6 is the major mode of congenital human herpesvirus 6 infection. Pediatrics **2008**; 122:513–20.

26. Tweedy J, Spyrou MA, Pearson M, Lassner D, Kuhl U, Gompels UA. Complete genome sequence of germline chromosomally integrated human herpesvirus 6A and analyses integration sites define a new human endogenous virus with potential to reactivate as an emerging infection. Viruses **2016**; 8:19.

27. Kawamura Y, Ohye T, Miura H, et al. Analysis of the origin of inherited chromosomally integrated human herpesvirus 6 in the Japanese population. J Gen Virol **2017**; 98:1823–30.

28. Greninger AL, Knudsen GM, Roychoudhury P, et al. Comparative genomic, transcriptomic, and proteomic reannotation of human herpesvirus 6. BMC Genomics **2018**; 19:204.

29. Ammon Avalos L, Galindo C, Li DK. A systematic review to calculate background miscarriage rates using life table analysis. Birth Defects Res A Clin Mol Teratol **2012**; 94:417–23.

30. Wilcox AJ, Weinberg CR, O'Connor JF, et al. Incidence of early loss of pregnancy. N Engl J Med **1988**; 319:189–94.

31. Magnus MC, Wilcox AJ, Morken NH, Weinberg CR, Håberg SE. Role of maternal age and pregnancy history in risk of miscarriage: prospective register based study. BMJ **2019**; 364:l869.

32. El Hachem H, Crepaux V, May-Panloup P, Descamps P, Legendre G, Bouet PE. Recurrent pregnancy loss: current perspectives. Int J Womens Health **2017**; 9:331–45.

33. Ando Y, Kakimoto K, Ekuni Y, Ichijo M. HHV-6 infection during pregnancy and spontaneous abortion. Lancet **1992**; 340:1289.

34. Revest M, Minjolle S, Veyer D, Lagathu G, Michelet C, Colimon R. Detection of HHV-6 in over a thousand samples: new types of infection revealed by an analysis of positive results. J Clin Virol **2011**; 51:20–4.

35. Drago F, Broccolo F, Zaccaria E, et al. Pregnancy outcome in patients with pityriasis rosea. J Am Acad Dermatol **2008**; 58:S78–83.

36. Marci R, Gentili V, Bortolotti D, et al. Presence of HHV-6A in endometrial epithelial cells from women with primary unexplained infertility. PLoS One **2016**; 11:e0158304.

37. Bortolotti D, Gentili V, Rotola A, et al. HHV-6A infection of endometrial epithelial cells affects immune profile and trophoblast invasion. Am J Reprod Immunol **2019**; 82:e13174.

38. Caselli E, Bortolotti D, Marci R, et al. HHV-6A infection of endometrial epithelial cells induces increased endometrial NK cell-mediated cytotoxicity. Front Microbiol **2017**; 8:2525.

39. Bortolotti D, Soffritti I, D'Accolti M, et al. HHV-6A infection of endometrial epithelial cells affects miRNA expression and trophoblast cell attachment. Reprod Sci **2020**; 27:779–86.

40. Gaccioli F, Lager S, de Goffau MC, et al. Fetal inheritance of chromosomally integrated human herpesvirus 6 predisposes the mother to pre-eclampsia. Nat Microbiol **2020**; 5:901–8.

41. Weschke DP, Leisenring WM, Lawler RL, et al. Inflammatory cytokine profile in individuals with inherited chromosomally integrated human herpesvirus 6. Biol Blood Marrow Transplant **2020**; 26:254–61.

42. Kumata R, Ito J, Sato K. Inherited chromosomally integrated HHV-6 possibly modulates human gene expression. Virus Genes **2020**; 56:386–9.

Unexpected Mutations by CRISPR-Cas9 CTG Repeat Excision in Myotonic Dystrophy and Use of CRISPR Interference as an Alternative Approach

Miki Ikeda,[1] Mariko Taniguchi-Ikeda,[2,3] Takema Kato,[3] Yasuko Shinkai,[3] Sonoko Tanaka,[1] Hiroki Hagiwara,[4] Naomichi Sasaki,[5] Toshihiro Masaki,[5] Kiichiro Matsumura,[1] Masahiro Sonoo,[1] Hiroki Kurahashi,[3] and Fumiaki Saito[1]

[1]Department of Neurology, School of Medicine, Teikyo University, Tokyo 1738606, Japan; [2]Department of Clinical Genetics, Fujita Health University Hospital, Aichi 4701192, Japan; [3]Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, Aichi 4701192, Japan; [4]Department of Medical Science, Teikyo University of Science, Uenohara Campus, Yamanashi 4090193, Japan; [5]Department of Medical Science, Teikyo University of Science, Senju Campus, Tokyo 1200045, Japan

Myotonic dystrophy type 1 is the most common type of adult-onset muscular dystrophy. This is an autosomal dominant disorder and caused by the expansion of the CTG repeat in the 3′ untranslated region of the dystrophia myotonica protein kinase (*DMPK*) gene. Messenger RNAs containing these expanded repeats form aggregates as nuclear RNA foci. Then, RNA binding proteins, including muscleblind-like 1, are sequestered to the RNA foci, leading to systemic abnormal RNA splicing. In this study, we used CRISPR-Cas9 genome editing to excise this CTG repeat. Dual cleavage at the 5′ and 3′ regions of the repeat using a conventional Cas9 nuclease and a double nicking with Cas9 nickase successfully excised the CTG repeat. Subsequently, the formation of the RNA foci was markedly reduced in patient-derived fibroblasts. However, contrary to expectations, a considerable amount of off-target digestions and on-target genomic rearrangements were observed using high-throughput genome-wide translocation sequencing. Finally, the suppression of *DMPK* transcripts using CRISPR interference significantly decreased the intensity of RNA foci. Our results indicate that close attention should be paid to the unintended mutations when double-strand breaks are generated by CRISPR-Cas9 for therapeutic purposes. Alternative approaches independent of double-strand breaks, including CRISPR interference, may be considered.

INTRODUCTION

Myotonic dystrophy type 1 (DM1) is the most common type of adult-onset muscular dystrophy, affecting 1 in 8,000 individuals.[1] DM1 is an autosomal dominant disorder that is characterized by systemic symptoms, including progressive muscular atrophy, muscular weakness, myotonia, cardiac arrhythmia, insulin resistance, gastrointestinal dysfunctions, cataract, and cognitive impairment.[1] It is caused by the expansion of the CTG repeat in the 3′ untranslated region (UTR) of the dystrophia myotonica protein kinase (*DMPK*) gene.[2,3] Healthy subjects have 5–37 CTG repeats, whereas DM1 symptoms

are identified in individuals with more than 50 CTG repeats.[2–4] The larger repeat sizes tend to be associated with worse clinical manifestations, and several thousand repeats are typically observed in severe congenital DM1 patients.[5,6] The expanded CTG repeats exhibit somatic and intergenerational instability with a bias toward expansion.[7,8]

The expanded CTG repeats in the 3′ UTR of *DMPK* are transcribed to mRNA as CUG repeats. It was previously demonstrated that the expanded CUG repeat forms stable hairpin structures that aggregate as RNA foci.[9–11] The intranuclear RNA foci sequester RNA binding proteins, including muscleblind-like 1 (MBNL1), a known splicing regulator.[12–14] This in turn leads to the depletion of soluble MBNL1 with normal regulatory function.[15] In addition, the RNA foci upregulate the activity of another splicing regulator, CUGBP Elav-like family member 1 (CELF1), by activating the protein kinase C pathway and suppressing the expression of specific microRNAs for CELF1.[16,17] The altered function of these splicing regulators results in the abnormal splicing of many genes, including *CLCN1*, *BIN1*, or *IRα*, which account for some aspects of the systemic features of DM1.[18–21]

The CRISPR-Cas9 system was first discovered as a microbial adaptive immune system.[22] It has since been successfully applied in the genome editing of eukaryotic cells and in a variety of research fields.[23–25] The most basic CRISPR-Cas9 system uses Cas9 nuclease derived from *Streptococcus pyogenes* and a single-guide RNA (sgRNA) with a complementary sequence to the target region of interest. These two components form a complex that is able to induce double-strand breaks (DSBs) at the target site. After cleavage, the DSBs are repaired by one of the two major repair pathways, that is,

non-homologous end joining (NHEJ) or homology-directed repair (HDR).[26] This powerful tool has been adapted for medical therapeutics, including DM1. Thus far, several groups have successfully excised the CTG repeat of the *DMPK* gene using the conventional Cas9 nuclease system in cultured cells and model mice.[27–30]

Although CRISPR-Cas9 is an innovative technology, care must be taken to avoid causing undesirable mutations when used for therapeutic purposes.[31] One way to avoid this lies in the use of the double nicking strategy.[32] In this system, Cas9 nickase, a D10A mutant of Cas9, is utilized with a pair of offset sgRNAs complementary to opposite strands of the target site. The nicks of both of the DNA strands lead to a DSB with a 5′ overhang. A large reduction in off-target cutting is expected due to the need for two sgRNAs, since it is unlikely that two off-target nicks will be generated by chance in close proximity.[32,33] Importantly, by dual DSBs, the region encompassed by up to several Mb can be removed and the 5′ and 3′ cut ends can be rejoined using the NHEJ or HDR repair systems.[34] Another candidate is CRISPR interference (CRISPRi), a strategy in which the transcription of any gene is downregulated without inducing DSBs.[35,36] This strategy utilizes catalytically inactive Cas9 (dCas9) fused with a transcription suppressor, KRAB, and sgRNA designed at the vicinity of transcription start sites (TSSs). This DSB-free method is expected to be much safer than DSB-dependent genome editing.

In the present study, we demonstrated that both a conventional Cas9 nuclease and a double nicking strategy using Cas9 nickase successfully excised the CTG repeat tract by designing sgRNAs at the 5′ and 3′ flanking regions. Using these procedures, the formation of RNA foci was markedly inhibited. However, the unbiased detection of genomic alterations using linear amplification-mediated high-throughput genome-wide translocation sequencing (LAM-HTGTS)[37,38] revealed unexpected on- and off-target mutations as a result of using these procedures. Lastly, we showed that the downregulation of *DMPK* transcription by CRISPRi significantly suppressed the formation of RNA foci. Based on these observations, we propose that approaches that are independent of a DSB formation, such as CRISPRi, should be considered when applying the CRISPR-Cas9 technologies for therapeutic purposes in the future.

RESULTS

Excision of *DMPK* CTG Repeat by Cas9 Nuclease

First, we tested whether the conventional CRISPR-Cas9 system using Cas9 nuclease and a pair of sgRNAs designed at the 5′ and 3′ region of the CTG repeat could be used to remove the repeat sequence in HEK293 cells. We confirmed by Sanger sequencing that the strain of HEK293 cells we used contained five CTG repeats (data not shown). As shown in Figure 1A, three sgRNAs were designed on both the 5′ and the 3′ regions of the CTG repeats. Each sgRNA was expressed with Cas9 nuclease in HEK293 cells and showed similar indel frequencies, according to the result of a T7 endonuclease 1 assay (Figure S1). Then, Cas9 nuclease and two sgRNAs, one targeting the

5′ region and another targeting the 3′ region of the repeat, were co-transfected. Three days after transfection, genomic PCR was performed using primers designed at the outer sites of the sgRNA targeting loci (Figure 1A). Upon these co-expressions, lower molecular weight bands were observed (Figure 1B), suggesting the excision of the CTG repeats.

We investigated whether similar events could be observed in the fibroblasts of DM1, denoted as GM03991, which contains 50–80 CTG repeats. Using PCR amplification, the control fibroblasts (mock electroporation) exhibited two distinct bands at 400 and 700 bp, corresponding to the wild-type and mutant allele, respectively. Based on the CRISPR design tool score, we chose two combinations of sgRNA pairs: (1) 5′ guide 2/3′ guide 2 and (2) 5′ guide 3/3′ guide 2. These showed the lowest likelihood of off-target digestion. By electroporation with these sgRNAs, the lower molecular weight bands appeared above 200 bp (Figure 1C). Sanger sequencing confirmed that these bands consisted of amplicons lacking the CTG repeat (Figures 1D, S2A, and S2B). Although several junctional sequences were obtained, they were mostly homogeneous (57.1% for type deletion [del] 2 in HEK293 cells and 83.3% for type del 1 in fibroblasts; Figure S2C). In the following experiments, we used the sgRNA pair 5′ guide 3/3′ guide 2 for Cas9 nuclease cleavage. The T7 endonuclease 1 assay did not reveal any obvious off-target mutations at the top five sites predicted by the CRISPR design tool in fibroblasts (Figure S3).

When designing *in vivo* gene therapy for DM1 using CRISPR-Cas9 technology, it is of particular importance to remove the CTG repeats from as many cells as possible. To this purpose, we investigated whether repeated genome editing may facilitate the excision. The repeated genome editing of HEK293 cells and fibroblasts using Cas9 nuclease increased the intensity of the lower bands gradually, indicating it facilitated repeat excision (Figure 1E).

Excision of *DMPK* CTG Repeat by Cas9 Nickase

In order to adapt the CRISPR-Cas9 technology for clinical usage, it is critical to reduce the potential for off-target effects. For this purpose, we tested whether the CTG repeat was excised using the CRISPR-Cas9 double nicking strategy. We designed seven (Nick 1, 2, 5, 6, 7, 8, and 9) and two (Nick 3 and 4) nicking pairs of sgRNAs on the 5′ and 3′ regions of the CTG repeat, respectively (Figure 2A). Upon transfection of these nicking pairs into HEK293 cells, a T7 endonuclease 1 assay demonstrated indel frequencies ranging from 30% to 40% or more (Figure S4). We then co-transfected HEK293 cells with Cas9 nickase and two pairs of sgRNAs, one targeting the 5′ region and another the 3′ region of the repeat. Genomic PCR showed lower molecular weight bands in all of the combinations of sgRNA pairs (Figure 2B). Among them, we chose the nicking pairs of Nick 1 and Nick 3 for the subsequent experiments, since the lower band was robust and the decrease in molecular weight was small, and the latter may only minimally affect the 3′ UTR structure of the *DMPK* gene. Sanger
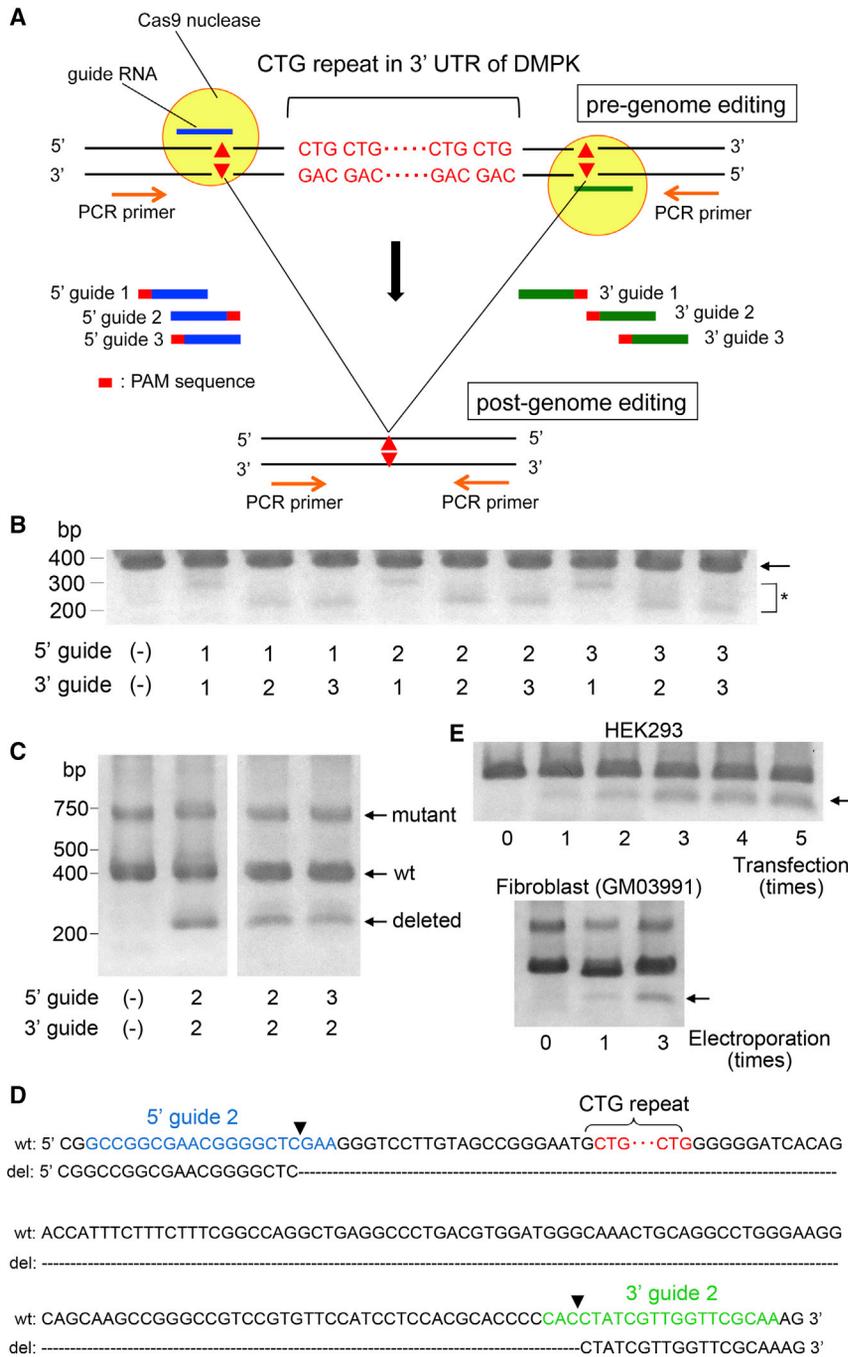
A

Figure 1. Excision of *DMPK* CTG Repeat Using Cas9 Nuclease

(A) Schematic representation of 3′ UTR of the *DMPK* gene is shown. The CTG repeat sequence is indicated in red characters. Three guide RNAs were designed at the 5′ and 3′ region of the CTG repeat (guide 1, 2, and 3). This region was amplified by PCR using the primers flanking the target sequence. (B) HEK293 cells were co-transfected with guide RNAs designed at the 5′ and 3′ regions with Cas9 nuclease, and genomic PCR was performed. By mock transfection, i.e., 5′ guide (–), 3′ guide (–), a PCR product of 400 bp was observed (arrow). Upon transfection with each combination of guide RNA, lower molecular weight bands were observed (asterisk). (C) Genomic PCR of DM1 patient-derived fibroblasts GM03991, harboring 50–80 CTG repeats, showed two distinct bands with an apparent molecular mass of 400 and 700 bp, corresponding to the wild-type and the mutant allele with an expanded CTG repeat, respectively. By genome editing using 5′ guide 2/3′ guide 2 and 5′ guide 3/3′ guide 2, the bands with a lower molecular mass above 200 bp were observed. (D) Sanger sequencing revealed that the lower molecular weight bands consisted of amplicons in which the CTG repeat region was excised. A typical junctional sequence is shown here. The blue, green, and red characters represent the position of the 5′ guide 2, 3′ guide 2, and CTG repeat, respectively. Arrowheads indicate the position of the expected DSBs. (E) Genome editing was repeated five times with an interval of 3 days to HEK293 cells (upper panel) and three times with an interval of 7 days to DM1-derived fibroblasts (lower panel). The CTG repeat excision was facilitated by repeating the genome editing (arrow).

lower bands, indicating that it facilitated the excision of the repeat (Figures 2D and 2E).

Suppression of RNA Foci Formation by CRISPR-Cas9 Genome Editing

As described above, the fibroblast GM03991, harboring 50–80 CTG repeats, was found to contain a wild-type allele at 400 bp and a mutant allele at 700 bp, according to the PCR results. In contrast, the fibroblast GM05163, harboring 400 repeats, showed only a wild-type allele, suggesting that the PCR failed to amplify the large mutant allele (Figure 3A, mock). Upon the genome editing of these fibroblasts using conventional Cas9 nuclease or double nicking with Cas9 nickase, the lower molecular weight bands appeared above 200 bp, indicating the successful excision of the CTG repeat (Figure 3A).

RNA foci were not detected in GM03991 using the RNA-fluorescence *in situ* hybridization (FISH) assay. As such, we studied GM05163 to evaluate the formation of RNA foci, one of the major pathological hallmarks of DM1 (Figure S6). To determine whether
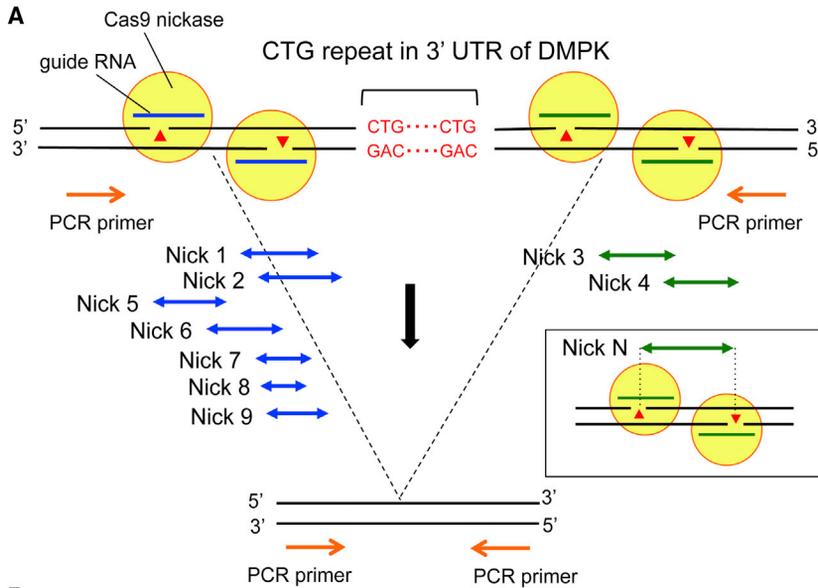
sequencing demonstrated that the lower band consisted of amplicons lacking the CTG repeat (Figures 2C, S5A, and S5B). By aligning the sequence data, it was evident that the junctional sequences were heterogeneous, compared to those observed using the conventional Cas9 nuclease (Figures S2B and S5B). No identical junctional sequence was observed among the clones we tested (Figure S5C). Repeated genome editing of HEK293 cells using the double nicking strategy gradually increased the intensity of the

Figure 2. Excision of *DMPK* CTG Repeat Using Cas9 Nickase

(A) Schematic representation of the 3′ UTR of the *DMPK* gene. The CTG repeat sequence is indicated in red characters. Seven (Nick 1, 2, 5, 6, 7, 8, and 9) and two (Nick 1 and 2) nicking pairs of guide RNAs were designed at the 5′ and 3′ regions of the CTG repeat, respectively. This region was amplified by PCR using the primers flanking the target sequence. (B) HEK293 cells were co-transfected with two pairs of guide RNAs, one against the 5′ region and another against the 3′ region of the repeat together with Cas9 nickase. Genomic PCR showed a PCR product of 400 bp by mock transfection i.e., 5′ Nick (–), 3′ Nick (–) (arrow). Upon transfection with each combination of nicking pairs of guide RNAs, lower molecular weight bands were observed (asterisk). (C) Sanger sequencing of the lower molecular weight band observed in Figure 3B by 5′ Nick 1 and 3′ Nick 3 revealed that it consisted of amplicons lacking the CTG repeat. A few junctional sequences are shown here. The blue, green, and red characters represent the position of Nick 1, Nick 3, and the CTG repeat, respectively. Arrowheads indicate the position of expected nicks. (D) Genome editing with Cas9 nickase was repeated five times with an interval of 3 days to HEK293 cells. The CTG repeat excision was facilitated by repeating the genome editing. (E) Quantitative densitometric analyses of the lower bands in (B) confirmed that the repeated genome editing increases the efficiency of repeat excision. The results are expressed as the mean ± SEM (n = 3).
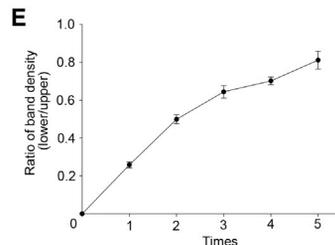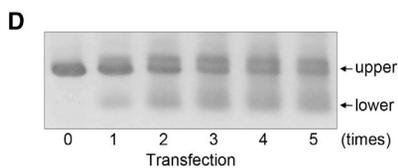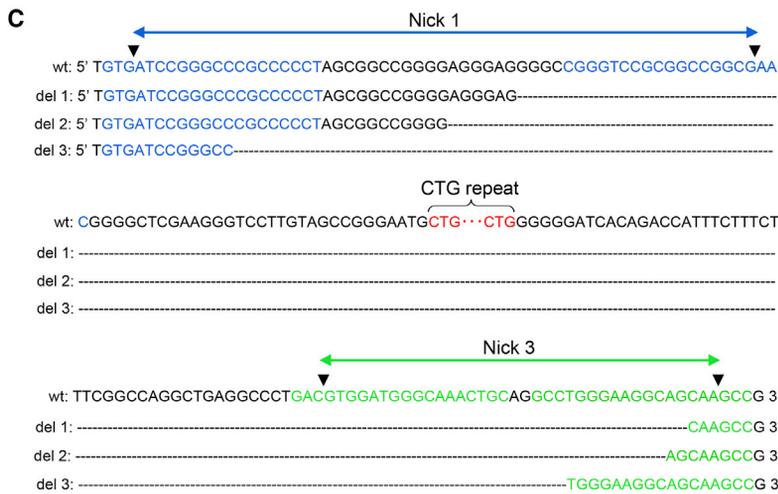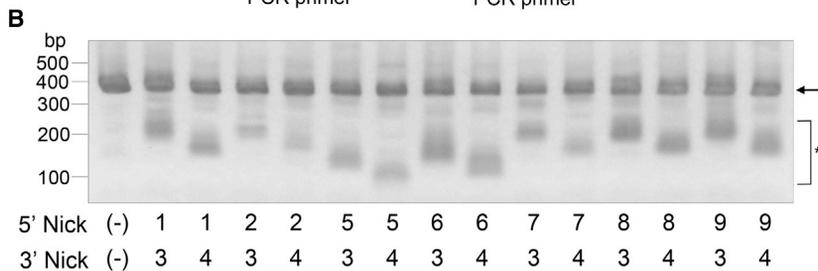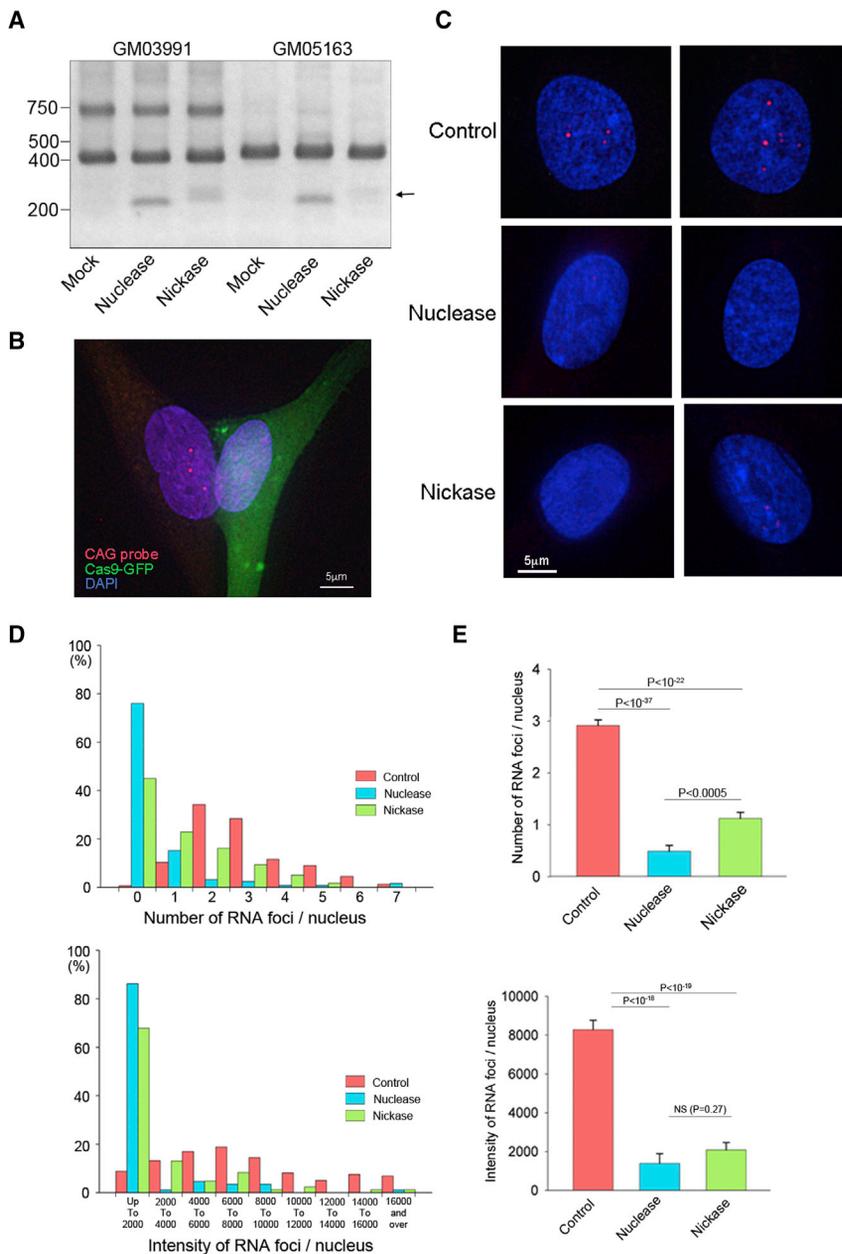
Figure 3. Suppression of RNA Foci Formation by the Repeat Excision

(A) Fibroblast GM03991, harboring 50–80 CTG repeats, was found to consist of a wild-type allele at 400 bp and a mutant allele at 700 bp by PCR. However, fibroblast GM05163, with 400 CTG repeats, exhibited only the wild-type allele, presumably because the PCR failed to amplify the highly repetitive sequence of the mutant allele. Genome editing was performed to these cells using conventional Cas9 nuclease and two sgRNAs (5′ guide 3 and 3′ guide 2) or the double nicking strategy with Cas9 nickase and four sgRNAs (Nick 1 and Nick 3 pairs). Using both of the procedures, lower molecular weight bands were observed above 200 bp (arrow), indicating the successful excision of the CTG repeat. (B) Fibroblast GM05163 was co-transfected with GFP-tagged Cas9 nuclease or Cas9 nickase together with sgRNAs. The RNA-FISH image shows intense RNA foci in the GFP-negative fibroblast (left) but not in the GFP-positive Cas9 nuclease-expressing cell (right). (C) In the nuclei of the control fibroblasts (no plasmids), several RNA foci were consistently observed. However, the RNA foci were mostly undetectable in fibroblasts expressing Cas9 nuclease or nickase. (D) Quantitative analyses of the number and the total intensity of intranuclear RNA foci were performed. Histograms of both the number (upper panel) and the intensity (lower panel) with a left-side skew were obtained for Cas9 nuclease and nickase. (E) The average number of RNA foci was significantly reduced by both Cas9 nuclease and nickase. When compared between Cas9 nuclease and nickase, the nuclease-treated cells exhibited significantly fewer foci (upper panel). The average intensity of RNA foci was significantly decreased by both Cas9 nuclease and nickase. There was no significant difference between the nuclease and nickase by this parameter (lower panel). The results are expressed as the mean ± SEM.

nuclease and nickase (Figure 3C). Furthermore, to ascertain the sequestration of MBNL1 to the RNA foci, an RNA-FISH assay was performed, followed by immunofluorescent analysis using anti-MBNL1 antibody. As a result, we found that MBNL1 colocalized with RNA foci in the DM1 patient-derived fibroblasts. Upon genome editing using Cas9 nuclease, both of the signals of the RNA foci and MBNL1 were abolished, indicating that the trapped MBNL1 was released from the foci (Figure S7).

Subsequently, we obtained more than 100 images of the nuclei from each group to quantitatively analyze the formation of the RNA foci. The number of RNA foci and the total intensity of RNA foci in each nucleus were measured. Histograms of both the number (Figure 3D, upper panel) and the total intensity (Figure 3D, lower panel) showed a left-sided skew resulting from Cas9 nuclease and nickase compared to the control. These data demonstrated that both genome editing strategies suppressed the formation of RNA foci.

the removal of the CTG repeat affected the formation of foci, fibroblasts were co-transfected with GFP-tagged Cas9 nuclease or Cas9 nickase together with sgRNAs targeting the 5′ and 3′ regions flanking the CTG repeat. A typical RNA-FISH image 3 days after genome editing is provided in Figure 3B. In the nucleus of the GFP-negative fibroblasts (Figure 3B, left), several intense dot-shaped RNA foci were detected, which were not observed in GFP-positive Cas9-expressing fibroblasts (Figure 3B, right). In the nuclei of the control fibroblasts (mock electroporation), several RNA foci were consistently observed. However, the RNA foci were mostly undetectable in the nuclei of fibroblasts expressing Cas9

Furthermore, the average number of RNA foci was significantly reduced by both Cas9 nuclease and nickase. When comparing Cas9 nuclease and nickase, the nuclease-treated cells exhibited significantly fewer foci (Figure 3E, upper panel). The average intensity of the RNA foci was also significantly decreased by these procedures. There was no significant difference between the nuclease and nickase with regard to this parameter (Figure 3E, lower panel).

Unbiased Genome-wide Detection of On- and Off-Target Mutations

To evaluate the unexpected mutations caused by CRISPR-Cas9, we performed LAM-HTGTS. Generally, DSBs generated by genome editing are rejoined via the classic NHEJ pathway, with occasional indels at the break site, and rarely by HDR. However, some DSBs, which are not rejoined immediately at the original loci, are fused with the separate cut ends when the other DSBs occur due to off-target cleavage. This leads to translocations to other chromosomes or intra-chromosomal deletions. By designing the sequence-specific primers at the fixed "bait" (DMPK locus in this case), LAM-HTGTS analyzes the genome-wide off-target "prey" DSBs captured by the on-target bait DSBs using next-generation sequencing (Figure 4A).

First, as a positive control, we induced a DSB using Cas9 nuclease and one sgRNA at the *RAG1* locus in HEK293 cells. Using this single DSB, the translocation hotspots were identified in chromosomes 7, 12, 15, and 19, as shown in a Circos plot (Figure 4B). This pattern of hotspots was consistent with that reported in a previous study.[37] Next, we performed genome editing to the *DMPK* 3′ UTR using Cas9 nuclease and two sgRNAs to excise the CTG repeat, and detected translocations using *DMPK* as a bait. Although no hotspots were found in the negative control (no editing) (Figure 4C), six translocation hotspots were observed in chromosomes 1, 14, 15, 17, 19, and X by the dual DSBs (Figure 4D). Furthermore, we attempted the double nicking strategy using Cas9 nickase and four sgRNAs to generate dual DSBs in the *DMPK* locus. We anticipated that off-target effects would be strikingly reduced by this procedure. However, contrary to our expectation, as many as 25 translocation hotspots were identified using double nicking genome editing (Figures 4E and 4F). The frequency of translocation to each hotspot is shown in Table S1. The total frequency of these translocations in total mapped reads by a DSB at the *RAG1* locus with Cas9 nuclease was 0.002625%. Comparably, the total frequencies by DSBs at the *DMPK* locus using Cas9 nuclease and Cas9 nickase were 0.001592% and 0.002169%, respectively.

Furthermore, to evaluate the on-target mutations neighboring the cutting sites, we mapped the paired reads obtained by LAM-HTGTS using integrative genomics viewer software.[39] Upon generation of a single DSB at the *RAG1* locus, the left-sided reads of some of the paired reads were mapped several kb apart from the breakpoint (Figure S8A). The non-edited samples using *DMPK* as a bait, which were used as negative controls for Cas9 nuclease and Cas9 nickase, exhibited only minimal gaps between the paired reads (Figures S8B and S8C). However, dual DSBs, generated by both Cas9 nuclease and Cas9 nickase, resulted in much larger gaps between the paired
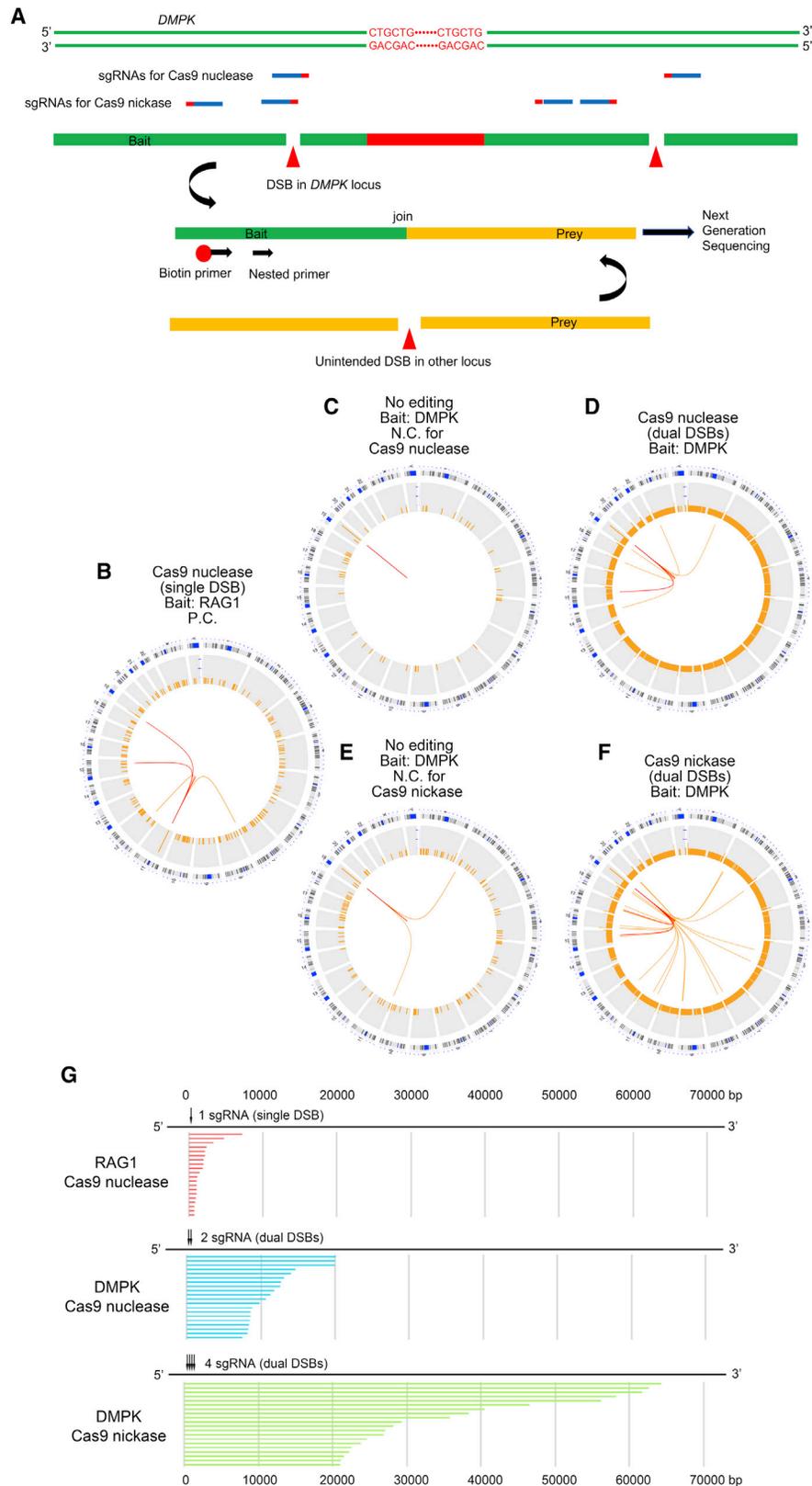
reads. These large gaps indicate that genomic rearrangements occurred including large deletions. Among them, paired reads with opposite orientations facing each other and with inferred insert sizes larger than expected, such as several hundred bp or larger, represent possible deletions. We extracted the paired reads with these possible deletions and ranked the top 20 according to their size (Figure 4G; Table S2). Using *RAG1* cleavage with a single sgRNA, potential on-target deletions ranging from several hundred bp to several kb were observed. However, the dual digestion at *DMPK* by Cas9 nuclease and nickase led to the generation of much larger deletions spanning beyond 20 and 60 kb, respectively.

Suppression of RNA Foci by CRISPRi

The generation of unexpected mutagenesis is of great concern when DSBs are generated using CRISPR-Cas9, as described above. Therefore, we examined whether the RNA foci were suppressed by the downregulation of *DMPK* transcription using CRISPRi, a DSB-free method. We searched the TSS of the human *DMPK* gene using the FANTOM5 database and identified the most enriched peak at chromosome 19 (ch19), 46,285,748 (hg19). Then, three sgRNAs neighboring the peak were designed (Figure 5A). Three days after the transfection of HEK293 cells with dCas9-KRAB and each sgRNA, the RNA was extracted from the whole cell population without selection. Quantitative RT-PCR revealed a significant reduction in the level of *DMPK* mRNA by guide 2 (Figure 5B). Next, to evaluate the formation of RNA foci, an RNA-FISH assay was performed using the fibroblast GM05163. Several RNA foci were clearly observed in the control fibroblasts, whereas RNA foci were mostly inconspicuous in the transfected cells with the guide 2, according to the red fluorescent protein (RFP) fluorescence of EF1-RFP-U6-gRNA plasmid (Figure 5C). Finally, the number and total intensity of the RNA foci were quantitatively analyzed. Histograms of both the number (Figure 5D, upper panel) and the total intensity (Figure 5D, lower panel) of RNA foci with a left-sided skew were obtained using CRISPRi. Although the average foci number was not significantly different (Figure 5E, upper panel), the average intensity of foci was significantly decreased by CRISPRi (Figure 5E, lower panel).

DISCUSSION

Thus far, no therapeutic strategy has been successfully applied in clinical use for the treatment of DM1, although many experimental approaches have been attempted in the past two decades. These include small molecular therapeutic strategies and posttranscriptional silencing using nucleotide sequences. For example, small molecules, such as erythromycin, and several designer small molecular compounds have been reported to block the interaction between CUG repeat and MBNL1.[40,41] It was also demonstrated that RNase-H active gapmer antisense oligonucleotides (ASOs) modified by 2′-O-methoxyethyl and 2′-4′-constrained ethyl effectively corrected the phenotype of DM1 in model mice.[42,43] Note that a gapmer ASO developed by Ionis Pharmaceuticals (IONIS-DMPKRX) entered a phase I/IIa trial, but enough concentration of the drug was not achieved in skeletal muscle to provide a therapeutic benefit (ClinicalTrials.gov: NCT02312011). Recently, genome editing technologies

A *DMPK*

B Cas9 nuclease (single DSB) Bait: RAG1 P.C.

C No editing Bait: DMPK N.C. for Cas9 nuclease

D Cas9 nuclease (dual DSBs) Bait: DMPK

E No editing Bait: DMPK N.C. for Cas9 nickase

F Cas9 nickase (dual DSBs) Bait: DMPK

G

(legend on next page)

based on CRISPR-Cas9 have been found to have a robust applicability in a variety of research fields, including medical therapeutics. Several groups have reported successful excision of the CTG repeat in DM1 with conventional Cas9 nuclease by designing sgRNAs at the 5′ and 3′ region flanking the repeats.[27–30] In alternative approaches, polyadenylation signals were inserted in the 3′ UTR upstream of the CTG repeat using CRISPR-Cas9 or transcription activator-like effector nuclease (TALEN).[44,45] This insertion led to the premature termination of transcription and the reversal of aberrant splicing. In terms of the delivery method for *in vivo* gene targeting, one of the most attractive vectors is the adeno-associated virus (AAV).[46] AAV does not integrate into the genome and produces high levels of long-term gene expression. A variety of serotypes are available that provide increased delivery efficiencies for the specific cell/tissue types. However, AAV has some disadvantages, including (1) a prolonged and uncontrollable expression that potentially increases the off-target mutagenesis, (2) a production of antibodies against AAV that may reduce the therapeutic effectiveness, and (3) a limited packing capacity that sometimes requires separate vector systems. Alternatively, CRISPR-Cas9 components can be delivered as a ribonucleoprotein (RNP) complex, which is generally considered to be safe since it is rapidly degraded. Although a local injection of RNP itself was shown to successfully induce genome editing,[47] many non-viral delivery systems are currently under development for more efficient RNP delivery. These include lipid nanoparticles,[48] gold nanoparticles,[49] cell-penetrating peptides,[50] and extracellular nanovesicles.[51]

In this study, we tried to validate the excision of expanded CTG repeats using a conventional CRISPR-Cas9 system. By designing sgRNAs at the 5′ and 3′ region of the repeats, Cas9 nuclease successfully removed the targeted region. Upon this procedure, RNA foci were suppressed, and trapped MBNL1 was released from the foci. Although we have not examined the correction of the mis-splicing of mRNAs, including *CLCN1*, *BIN1*, or *IRα*, it is presumed to be difficult to validate this using fibroblasts.[21] CRISPR-Cas9 is an innovative tool that has potential applications in clinical use in the future; however, it can lead to unintended mutagenesis. To reduce this possibility,

we excised the CTG repeats using the double nicking strategy, by which the off-target cuttings are considered to be strictly inhibited. We found that the repeat sequence was successfully deleted in HEK293 cells and DM1 fibroblasts using this procedure. Note that on-target junctional sequences were found to be heterogeneous using Sanger sequencing after double nicking. The breakpoints spanned widely, from the upstream of 5′ nicking sgRNAs to the downstream of 3′ sgRNAs. This is in contrast to Cas9 nuclease, which mostly generated the expected junctions. It is difficult to predict the exact on-target cutting sites when the double nicking strategy is employed, which is a potential disadvantage of this procedure. In addition to the accuracy, a degree of effectiveness is required to apply the genome editing to clinical use, especially for disorders in which the pathogenic mechanism is based on the gain of toxic functions, including DM1. We found that the excision of CTG repeats was facilitated by genome editing. Thus, several rounds of repeating genome editing therapies could be used in severe cases to obtain an additive efficacy, such as in chemotherapy for cancer or immunosuppressive therapies for autoimmune-related disorders.[52]

Although both the conventional Cas9 nuclease and the double nicking strategy with Cas9 nickase significantly reduced the RNA foci, Cas9 nuclease seems to have a subtle advantage over Cas9 nickase in terms of its efficiency. One possible reason for this difference may be that four sgRNAs are necessary to exert an effect simultaneously at each targeting loci in the double nicking strategy, in contrast to Cas9 nuclease, which requires two sgRNAs. Alternatively, the efficiency of the DSB formation by paired nickase may be lower than that by Cas9 nuclease. The former is more likely since D10A Cas9 nickase used in this study was recently reported to have a higher cutting efficiency than Cas9 nuclease.[53]

So far, many methodologies for the detection of unintended mutations by genome editing have been reported, and they are grouped into two categories: "biased" methods and "unbiased" methods.[54] In the biased methods, potential off-target sites are predicted using *in silico* homology searches, and the resulting sites are analyzed by

---

Figure 4. Unbiased Genome-wide Detection of On- and Off-Target Mutations

(A) A diagram of LAM-HTGTS for the detection of unbiased genome-wide off-target effects is shown. The broken end of the *DMPK* locus generated by on-target DSB works as a "bait" and captures genome-wide "prey" induced by unintended DSBs in other loci. Following PCR using a biotinylated primer, enrichment with avidin beads, and nested PCR, the bait-prey hybrid sequences are read using next-generation sequencing. (B) Using Cas9 nuclease and one sgRNA targeting the *RAG1* site, a single DSB was generated in the HEK293 cells as a positive control. With *RAG1* as a bait, LAM-HTGTS was used to reveal several translocation hotspots. The chromosomal location is indicated in the outermost part of the Circos plot. Split reads binned into 50-Mb regions (orange bars) have been plotted on a log scale. Colored lines (orange and red) connect the *RAG1* bait site on chromosome 11 to the genome-wide prey hotspots binned into 100-bp regions. The orange and red line colors indicate a split-reads frequency of 3–9 and of more than 10, respectively. (C) LAM-HTGTS was performed without genome editing using *DMPK* as a bait, a negative control for Cas9 nuclease. No translocation hotspot was detected. (D) Using Cas9 nuclease and two sgRNAs targeting the *DMPK* locus, dual DSBs were generated. Using *DMPK* as a bait, six translocation hotspots were identified. (E) LAM-HTGTS was performed without genome editing using *DMPK* as a bait, a negative control for the double nicking strategy with Cas9 nickase. (E) differs from (C) in the filtering process of split reads since the deduced on-target breakpoint is different between them (details are found in Materials and Methods). (F) Using Cas9 nickase and four sgRNAs, dual DSBs were generated. Using *DMPK* as a bait, as many as 24 translocation hotspots were identified. (G) Split reads obtained by LAM-HTGTS were mapped in integrative genomics viewer. Possible deletions, jugged from the orientation of the paired reads and their intervals, were extracted and the top 20 were sorted according to their size. By generating a single DSB on the *RAG1* locus using Cas9 nuclease, possible on-target deletions ranging from several hundred bp to several kb were observed (upper panel). By generating dual DSBs on the *DMPK* locus using Cas9 nuclease and two sgRNAs, larger deletions with an estimated size ranging from several to 20 kb were observed (middle panel). By generating dual DSBs on the *DMPK* locus using Cas9 nickase and four sgRNA, larger deletions with an estimated size ranging from 20 to 60 kb were observed (lower panel). Arrows indicate the position of sgRNA.
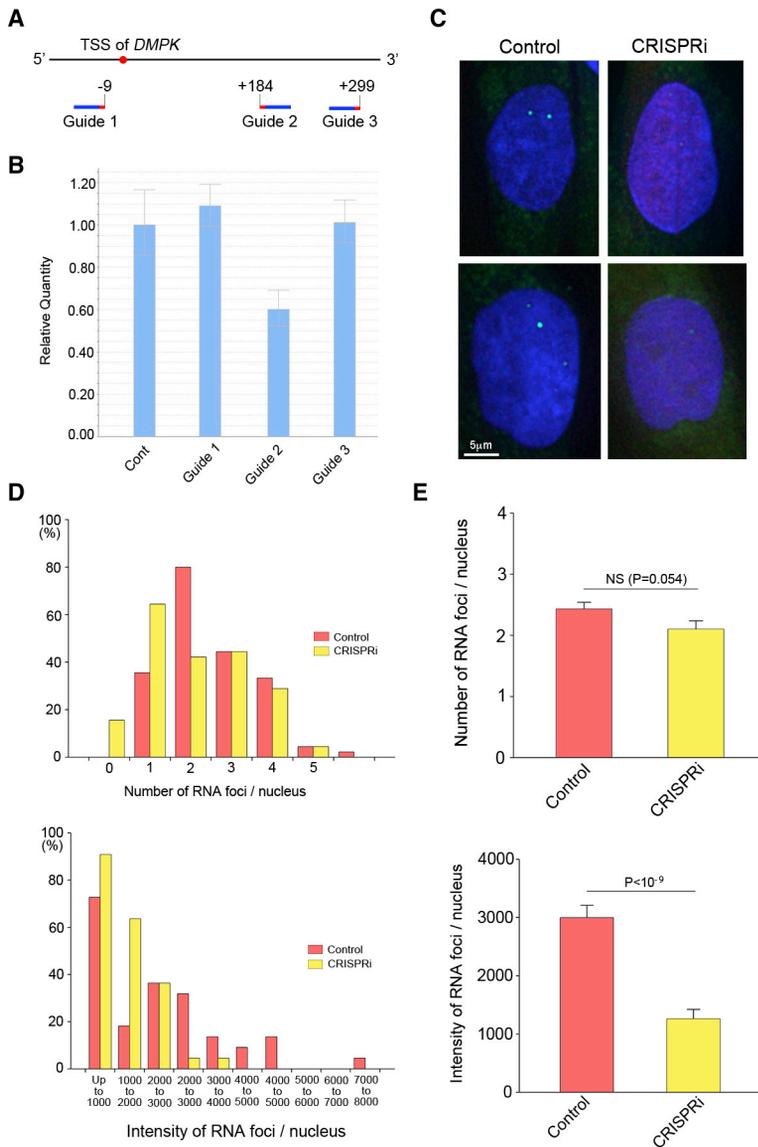
Figure 5. Suppression of RNA Foci by CRISPRi

(A) To suppress the transcription of the *DMPK* gene by CRISPRi, three sgRNAs, one upstream and two downstream of the TSS, were designed. (B) After the transfection of the HEK293 cells, total RNA was extracted from the whole cell population without selection. Quantitative RT-PCR revealed a significant reduction in the level of *DMPK* mRNA by guide 2. The results are expressed as the mean ± SEM (N = 3). (C) An RNA-FISH assay was performed using fibroblast GM05163 after CRISPRi using guide 2. Several RNA foci were clearly observed in the control fibroblasts, but they were mostly inconspicuous in the transfected cells. (D) Quantitative analysis of the number and total intensity of the intranuclear RNA foci was performed. Histograms of both the number (upper panel) and the total intensity (lower panel) shifted toward the left side by CRISPRi compared to control. (E) Although the fibroblasts transfected for CRISPRi tended to show fewer foci, the difference in the average foci number was not statistically significant (upper panel). The average intensity of the RNA foci was significantly decreased by CRISPRi (lower panel). The results are expressed as the mean ± SEM.

tively frequent homologous sequences along the genome, thereby generating DSBs individually without pairing. Another possibility is that the single-strand breaks (SSBs; nicks) generated by Cas9 nickase led to the formation of DSBs during the replication of the chromosome. SSBs are mainly repaired via the base excision repair pathway with a high degree of fidelity.[56] However, if they are left unrepaired, replicative polymerases encounter these SSBs, which can result in the collapse of the replication fork, and subsequently in the formation of DSBs.[57–59] In the presence of Cas9 nickase, the cycles of nick formation and their repair will be continuously repeated until the enzyme is inactivated. This may result in the formation of frequent DSBs, especially in the actively dividing cells, such as HEK293 cells.

Apart from the off-target effects, LAM-HTGTS revealed on-target genomic rearrangements, including possible large deletions. The largest deletion size estimated was several kb, resulting from a single DSB at the *RAG1* locus. However, this was exceeded by 20 and 60 kb by dual DSBs at the *DMPK* locus as a result of Cas9 nuclease and Cas9 nickase, respectively. Recently, on-target large deletions caused by CRISPR-Cas9 have been reported by several research groups. They have found deletions ranging from several hundred bp up to several kb in mouse embryonic stem cells (ESCs) or cultured cell lines using long PCR and Sanger sequencing.[60–63] However, larger deletions, such as those identified in our study, are difficult to detect using these strategies, since PCR primer binding sites are lost. In another report, an unbiased mutation detection methodology, UDiTaS (uni-directional targeted sequencing), was used to detect on-target mutations.[64] They generated dual DSBs using Cas9 nuclease at the *DMD* locus and found large deletions of up to several kb in mice.[65] The difference in the deletion size between these studies and our own study (several kb versus

a T7 endonuclease 1 assay or deep sequencing.[26] However, these methods have only a limited capability to detect off-target mutations and cannot reveal unpredicted genomic alterations. In contrast, the unbiased methods directly detect mutations at the genome-wide level independent of *in silico* prediction tools.[54] Therefore, to develop medical therapeutics for human diseases using genome editing, on- and off-target mutations need to be thoroughly assessed using the unbiased method, although most studies in the context of disease treatment do not use them. Using LAM-HTGTS, we clearly demonstrated a significant translocation hotspot, even in the case of the double nicking strategy, against our expectations.[37] One possible explanation for the off-target cleavages is that the Cas9 nickase may exert residual nuclease activity. Indeed, D10A Cas9 nickase with a single sgRNA (not a double nicking) exhibited low levels of on-target indels.[24,55] The sgRNAs we used for the double nicking strategy may have rela-

several 10s of kb) may be attributed to the differences in the targeting species, locus, or the procedures used. In our study, the dual DSBs rather than the single DSB, and the double nicking rather than the conventional Cas9 nuclease, induced larger deletions. However, no general conclusions can be currently drawn due to the limited number of experiments. As such, further investigations will be required to confirm our findings.

As discussed above, DSBs can exert potential deleterious effects on genomic structure, even when the double nicking strategy is used. In addition, recent reports found that DSBs induced by CRISPR-Cas9 trigger a p53-mediated DNA damage response and cell cycle arrest.[66,67] Prompted by these observations, we sought to clarify the effects of CRISPRi, a methodology independent of a DSB formation, using dCas9. Since it was demonstrated that the FANTOM5 promoter atlas represented the most reliable source of TSS annotations, we used this online database for the prediction of TSSs.[68] Although the position of sgRNA spanning −50 to +150 relative to the TSS is generally recommended, its functionality also depends on the chromatin accessibility of the target site.[68] In our case, using the sgRNA designed at +184 relative to the *DMPK* TSS, the transcription and the formation of RNA foci were successfully suppressed. Although this procedure inhibits the transcription of both the normal and mutant allele of *DMPK*, serious undesired effects were not presumed since a targeted deletion of *DMPK* exhibited no significant phenotypic alteration in mice.[69] So far, two studies have reported on the use of dCas9 to inhibit the pathogenic pathway of DM1. The strategies reported in these studies differ from our own in that the CTG or CUG repeat sequence on the genome or mRNA was directly targeted by sgRNAs.[70,71] These approaches allow for the selective reduction of mutant *DMPK* mRNA, while non-selectively affecting the CTG repeat tracts found in the form of microsatellite sequences on the human genome.

In conclusion, we demonstrated that both a conventional Cas9 nuclease method and a double nicking strategy using Cas9 nickase can be used to successfully excise CTG repeats in the DM1 locus and suppress the formation of RNA foci. However, contrary to our expectations, off-target cleavage and on-target genomic rearrangements were observed as a result of using the double nicking strategy, to a comparable degree as that observed when using Cas9 nuclease excision. By reducing the transcription of *DMPK* using CRISPRi, a DSB-free procedure, the formation of RNA foci was significantly inhibited. We propose that this alternative approach should be used for the development of a safer therapeutic strategy for the treatment of DM1 in the future.

MATERIALS AND METHODS
Plasmid Construction
The expression plasmids for Cas9 nuclease (pSpCas9(BB)-2A-GFP) and Cas9 nickase, a D10A mutant version of Cas9 nuclease (pSpCas9n(BB)-2A-GFP), were gifts from Dr. Feng Zhang (Addgene, plasmid nos. 48138 [http://addgene.org/48138] and 48140 [http://addgene.org/48140]). The target sites of sgRNA were selected using

the online CRISPR design tool (http://zlab.bio/guide-design-resources). Based on the input of the sequences flanking the 5′ and 3′ regions of the CTG repeat of the *DMPK* gene, several targeting sites were chosen to generate the fewest number of off-target sites as close to the CTG repeat as possible. The sgRNAs were cloned into the plasmids as described previously.[26] Briefly, the top and bottom strands of oligonucleotides with BbsI restriction sites on their 5′ termini were chemically synthesized at FASMAC (Atsugi, Japan). The oligonucleotides were phosphorylated and annealed in a thermal cycler using the following program: 37°C for 30 min; 95°C for 5 min; ramp down to 25°C at 5°C/min. Then, pSpCas9(BB)-2A-GFP and pSpCas9n(BB)-2A-GFP were digested by BbsI (NEB, Ipswich, MA, USA) and ligated with each sgRNA. The insertions of the sgRNAs were confirmed by Sanger sequencing. For the CRISPRi experiments, AAV CMV-dSaCas9-KRAB-bGHpA, gifted by Dr. Charles Gersbach (Addgene, plasmid no. 106219 [http://addgene.org/106219]) and EF1-RFP-U6-gRNA (System Biosciences, Palo Alto, CA, USA) were used to express dCas9 and sgRNA, respectively.[72] The TSS of the human *DMPK* gene was identified using the online database FANTOM5 (https://fantom.gsc.riken.jp/5/).[73] sgRNAs with a PAM sequence of NNGRRT or NNGRR were designed in the vicinity using Benchling software (https://www.benchling.com). The corresponding oligonucleotides were synthesized, annealed, and ligated to the EF1-RFP-U6-gRNA plasmid according to the manufacturer's instructions. The sgRNA sequences used are listed in Table S3.

Cell Culture, Transfection, and Electroporation
HEK293 cells and DM1 patient-derived fibroblasts were obtained from RIKEN BioResource Research Center (BRC) (Tsukuba, Japan) and the Coriell Institute (Camden, NJ, USA), respectively. The fibroblast GM03991, harboring 50–80 CTG repeats, was obtained from a mildly affected patient, while the fibroblast GM05163, with 400 CTG repeats, was obtained from a moderately affected individual. HEK293 cells were plated on plastic culture dishes or six-well plates (Corning Life Sciences, Oneonta, NY, USA) and grown in the Dulbecco's modified Eagle's medium (Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 10% fetal bovine serum and antibiotic-antimycotic (Thermo Fisher Scientific, Waltham, MA, USA). Fibroblasts were plated on six-well plates and grown in minimal essential medium (Thermo Fisher Scientific, Waltham, MA, USA) with 10% fetal bovine serum and antibiotic-antimycotic. All cells were grown in a humidified 37°C incubator with 5% CO_2 and 95% air. The HEK293 cells were transfected with the expression plasmids using Effectene (QIAGEN, Hilden, Germany). The fibroblasts were transfected using electroporation with an Amaxa 4D-Nucleofector (Lonza, Basel, Switzerland) by condition CA-137 according to the manufacturer's instructions. The cells were harvested 3 days after the transfection.

PCR Amplification and T7 Endonuclease 1 Assay
Genomic DNA was extracted from the cultured cells using a DNeasy Blood & Tissue kit (QIAGEN, Hilden, Germany) and quantified using a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). The PCR amplification of the on-target *DMPK* locus was

performed using the extracted genomic DAN as a template with KOD FX (Toyobo, Osaka, Japan). The PCR primers used are indicated in Table S3. The following program was used: an initial denaturation at 94°C for 2 min; 35 cycles at 98°C for 10 s, 68°C for 3 min. After the separation of the PCR products in 2% agarose gel, the DNA was extracted using a QIAquick gel extraction kit (QIAGEN, Hilden, Germany) and analyzed by Sanger sequencing. A T7 endonuclease 1 assay was performed using a Surveyor mutation detection kit (Integrated DNA Technologies, Coralville, IA, USA), according to the manufacturer's protocol. Briefly, after genome editing with Cas9 nuclease or Cas9 nickase, the genomic DNA was extracted from HEK293 cells or fibroblasts. On- and off-target loci were amplified by PCR using LA Taq DNA polymerase (Takara Bio, Shiga, Japan) and the primers listed in Table S3. The PCR products were heat-denatured and then re-annealed according to the following program in a thermal cycler: 95°C for 10 min; ramp down to 25°C at 1°C/15 s; hold at 4°C. Heteroduplex PCR products comprised of wild-type and mutant alleles were digested by incubating with Surveyor nuclease at 42°C for 60 min. The PCR products were separated by 4%–20% gradient polyacrylamide gel electrophoresis, and the band intensities were measured using ImageJ software. Indel occurrence (indel %) was estimated as previously described.[26]

Quantitative RT-PCR

The RNA was extracted from cultured cells using an RNeasy mini kit (QIAGEN, Hilden, Germany) and quantified using a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). Then, the cDNA was reverse transcribed using ReverTra Ace (Toyobo, Osaka, Japan) according to the manufacturer's instructions. Quantitative RT-PCR was conducted with a 7500 Fast real-time PCR system (Thermo Fisher Scientific, Waltham, MA, USA) using the cDNAs as templates. The gene expression was quantitatively analyzed by TaqMan assay using TaqMan probes Hs01094329_m1 and Hs02786624_g1 for DMPK and GAPDH, where the latter was used as an internal control.

RNA-FISH and Quantitative Analysis of RNA Foci

For the RNA-FISH assay, a DNA/locked nucleic acid (LNA) chimeric oligonucleotide labeled with Cy3 or Alexa Fluor 488 was synthesized as follows (GeneDesign, Osaka, Japan): 5′-Cy3/Alexa 488-CAGCAG CAGCAGCAGCAGCA-3′, where underlining denotes LNAs. The fibroblasts were grown on non-coated coverslips. The cells were fixed with cold 4% paraformaldehyde for 30 min at 4°C and permeabilized for 5 min at 4°C with 2% acetone in PBS pre-chilled at −20°C. Hybridization was performed with 2 ng/μL of DNA/LNA probe in hybridization buffer (30% formamide, 2× SSC, 200 ng/mL single-stranded DNA [ssDNA], 0.02% BSA, 10% dextran sulfate, and 2 mM vanadyl ribonucleoside) overnight at 37°C. Then, the cells were washed with 30% formamide in 2× SSC for 30 min at 45°C and 30% formamide in 1× SSC for 30 min at 37°C. For the immuno-fluorescence detection of the transfected cells, fibroblasts were blocked with 3% BSA in PBS for 1 h and incubated with anti-GFP (Thermo Fisher Scientific, Waltham, MA, USA) or anti-RFP (MBL International, Nagoya, Japan) antibody. After washing, the coverslips

were mounted with ProLong Diamond antifade mountant with DAPI (Thermo Fisher Scientific, Waltham, MA, USA), and fluorescent images were taken using an FSX100 fluorescence microscope (Olympus, Tokyo, Japan). For the quantitative analysis of the RNA foci, images of more than 100 nuclei were captured from each group. The numbers of RNA foci in the nuclei were counted by visual observation, and the intensities of the RNA foci in the nuclei were measured using ImageJ software. Statistical differences were evaluated using Student's t test. p values <0.05 were considered statistically significant. For the detection of MBNL1, the RNA-FISH assay was followed by immunofluorescent staining with anti-MBNL1 MB1a (Developmental Studies Hybridoma Bank, Iowa City, IA, USA), according to the standard protocol.

Unbiased Genome-wide Detection of Off-target DSBs

LAM-HTGTS was performed as previously reported with some modifications.[37,38] Briefly, HEK293 cells (6×10^6 cells for each group) were transfected with Cas9 nuclease or Cas9 nickase with the respective sgRNAs (5′ guide 3 and 3′ guide 2 for Cas9 nuclease, Nick 1 and Nick 3 for Cas9 nickase). As a positive control of the off-target effects, the cells were transfected with Cas9 nuclease and sgRNA targeting RAG1 gene site A (RAG1A) (Table S3).[37] Three days after the transfection, genomic DNA was extracted from the cells with a DNeasy Blood & Tissue kit (QIAGEN, Hilden, Germany) and incubated overnight at 56°C. Then, the genomic DNA was sheared using a Covaris Focused-ultrasonicator (Covaris, Woburn, MA, USA). After purification using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, USA), the sheared DNA was amplified by LAM-PCR using KOD FX (Toyobo, Osaka, Japan) and sequence-specific biotinylated primers for DMPK and RAG1 (Table S4) under the following conditions: an initial denaturation at 98°C for 2 min; 80 cycles at 95°C for 30 s, 58°C for 30 s, 72°C for 90 s; a final extension at 72°C for 2 min. The PCR products were incubated with Dynabeads MyONE streptavidin C1 beads (Thermo Fisher Scientific, Waltham, MA, USA) for 4 h. The DNA-beads complex was then captured by the magnetic stand. After annealing the upper and lower strands of the oligonucleotides for the bridge adaptor (Table S4), this was ligated to the DNA-beads complex using T4 DNA ligase (Promega, Madison, WI, USA) by incubating at 25°C for 1 h, 92°C for 2 h, and 16°C for 1 h. The on-beads ligation products were again captured by the magnetic stand and washed twice. Nested PCR was performed using the on-beads ligation products as templates with the primers indicated in Table S4. The PCR conditions were as follows: an initial denaturation at 95°C for 5 min; 15 cycles at 95°C for 60 s, 60°C for 30 s, 72°C for 60 s; a final extension at 72°C for 6 min. The PCR products were centrifuged at 15,000 × g for 5 min and the supernatants were concentrated using a QIAquick gel extraction kit (QIAGEN, Hilden, Germany). The DNA samples were barcoded for multiplexing using dual index primer sets (Illumina, San Diego, CA, USA) (Table S4) using the following conditions: an initial denaturation at 95°C for 3 min; 16 cycles at 95°C for 30 s, 62°C for 30 s, 72°C for 60 s; a final extension at 72°C for 6 min. The PCR products were run on a 1% agarose gel and the DNA was extracted from an area ranging from 500 to 1,000 bp using a QIAquick gel extraction kit. After barcoding, the samples were pooled in equal mass ratios. A denatured and diluted

library pool of 600 μL of PhiX control was sequenced with the 300-bp paired-end sequencing reactions on the MiSeq sequencer using the a V3 600 cycle kit (Illumina, San Diego, CA, USA). After trimming the adaptor sequences, the sequence reads were mapped to the hg19 reference genome, and any PCR duplicates were removed. The reads, which contain 20 bp of an adjacent to on-target breakpoint locus and at least 10 bp of uncertain nucleotides beyond the breakpoint, were specifically extracted for subsequent analysis. Off-target digestion and genomic rearrangements were detected as split reads, which mapped more than two different genomic loci in the same reads. Split reads were counted in a series of 100-bp equal-sized bins of entire genome. Off-target sites and the number of split reads were drawn using a Circos plot.[74]

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.omtm.2020.05.024.

AUTHOR CONTRIBUTIONS

CONFLICTS OF INTEREST

The authors declare no competing interests.

ACKNOWLEDGMENTS

REFERENCES

1. Turner, C., and Hilton-Jones, D. (2010). The myotonic dystrophies: diagnosis and management. J. Neurol. Neurosurg. Psychiatry 81, 358–367.

2. Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barceló, J., O'Hoy, K., et al. (1992). Myotonic dystrophy mutation: an unstable CTG repeat in the 3′ untranslated region of the gene. Science 255, 1253–1255.

3. Fu, Y.H., Pizzuti, A., Fenwick, R.G., Jr., King, J., Rajnarayan, S., Dunne, P.W., Dubel, J., Nasser, G.A., Ashizawa, T., de Jong, P., et al. (1992). An unstable triplet repeat in a gene related to myotonic muscular dystrophy. Science 255, 1256–1258.

4. Zerylnick, C., Torroni, A., Sherman, S.L., and Warren, S.T. (1995). Normal variation at the myotonic dystrophy locus in global human populations. Am. J. Hum. Genet. 56, 123–130.

5. Harley, H.G., Rundle, S.A., MacMillan, J.C., Myring, J., Brook, J.D., Crow, S., Reardon, W., Fenton, I., Shaw, D.J., and Harper, P.S. (1993). Size of the unstable CTG repeat sequence in relation to phenotype and parental transmission in myotonic dystrophy. Am. J. Hum. Genet. 52, 1164–1174.

6. Tsilfidis, C., MacKenzie, A.E., Mettler, G., Barceló, J., and Korneluk, R.G. (1992). Correlation between CTG trinucleotide repeat length and frequency of severe congenital myotonic dystrophy. Nat. Genet. 1, 192–195.

7. Lavedan, C., Hofmann-Radvanyi, H., Shelbourne, P., Rabes, J.P., Duros, C., Savoy, D., Dehaupas, I., Luce, S., Johnson, K., and Junien, C. (1993). Myotonic dystrophy: size- and sex-dependent dynamics of CTG meiotic instability, and somatic mosaicism. Am. J. Hum. Genet. 52, 875–883.

8. Ashizawa, T., Anvret, M., Baiget, M., Barceló, J.M., Brunner, H., Cobo, A.M., Dallapiccola, B., Fenwick, R.G., Jr., Grandell, U., Harley, H., et al. (1994). Characteristics of intergenerational contractions of the CTG repeat in myotonic dystrophy. Am. J. Hum. Genet. 54, 414–423.

9. Davis, B.M., McCurrach, M.E., Taneja, K.L., Singer, R.H., and Housman, D.E. (1997). Expansion of a CUG trinucleotide repeat in the 3′ untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. Proc. Natl. Acad. Sci. USA 94, 7388–7393.

10. Michalowski, S., Miller, J.W., Urbinati, C.R., Paliouras, M., Swanson, M.S., and Griffith, J. (1999). Visualization of double-stranded RNAs from the myotonic dystrophy protein kinase gene and interactions with CUG-binding protein. Nucleic Acids Res. 27, 3534–3542.

11. Mooers, B.H., Logue, J.S., and Berglund, J.A. (2005). The structural basis of myotonic dystrophy from the crystal structure of CUG repeats. Proc. Natl. Acad. Sci. USA 102, 16626–16631.

12. Miller, J.W., Urbinati, C.R., Teng-Umnuay, P., Stenberg, M.G., Byrne, B.J., Thornton, C.A., and Swanson, M.S. (2000). Recruitment of human muscleblind proteins to (CUG)$_n$ expansions associated with myotonic dystrophy. EMBO J. 19, 4439–4448.

13. Mankodi, A., Urbinati, C.R., Yuan, Q.P., Moxley, R.T., Sansone, V., Krym, M., Henderson, D., Schalling, M., Swanson, M.S., and Thornton, C.A. (2001). Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. Hum. Mol. Genet. 10, 2165–2170.

14. Fardaei, M., Rogers, M.T., Thorpe, H.M., Larkin, K., Hamshere, M.G., Harper, P.S., and Brook, J.D. (2002). Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. Hum. Mol. Genet. 11, 805–814.

15. Kanadia, R.N., Johnstone, K.A., Mankodi, A., Lungu, C., Thornton, C.A., Esson, D., Timmers, A.M., Hauswirth, W.W., and Swanson, M.S. (2003). A muscleblind knockout model for myotonic dystrophy. Science 302, 1978–1980.

16. Kuyumcu-Martinez, N.M., Wang, G.S., and Cooper, T.A. (2007). Increased steady-state levels of CUGBP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation. Mol. Cell 28, 68–78.

17. Kalsotra, A., Singh, R.K., Gurha, P., Ward, A.J., Creighton, C.J., and Cooper, T.A. (2014). The Mef2 transcription network is disrupted in myotonic dystrophy heart tissue, dramatically altering miRNA and mRNA expression. Cell Rep. 6, 336–345.

18. Mankodi, A., Takahashi, M.P., Jiang, H., Beck, C.L., Bowers, W.J., Moxley, R.T., Cannon, S.C., and Thornton, C.A. (2002). Expanded CUG repeats trigger aberrant splicing of ClC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy. Mol. Cell 10, 35–44.

19. Charlet-B, N., Savkur, R.S., Singh, G., Philips, A.V., Grice, E.A., and Cooper, T.A. (2002). Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. Mol. Cell 10, 45–53.

20. Fugier, C., Klein, A.F., Hammer, C., Vassilopoulos, S., Ivarsson, Y., Toussaint, A., Tosch, V., Vignaud, A., Ferry, A., Messaddeq, N., et al. (2011). Misregulated alternative splicing of BIN1 is associated with T tubule alterations and muscle weakness in myotonic dystrophy. Nat. Med. 17, 720–725.

21. Savkur, R.S., Philips, A.V., and Cooper, T.A. (2001). Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. Nat. Genet. 29, 40–47.

22. Garneau, J.E., Dupuis, M.È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468, 67–71.

23. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823.

24. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science 339, 823–826.

25. Komor, A.C., Badran, A.H., and Liu, D.R. (2017). CRISPR-based technologies for the manipulation of eukaryotic genomes. Cell 168, 20–36.

26. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. 8, 2281–2308.

27. van Agtmaal, E.L., André, L.M., Willemse, M., Cumming, S.A., van Kessel, I.D.G., van den Broek, W.J.A.A., Gourdon, G., Furling, D., Mouly, V., Monckton, D.G., et al. (2017). CRISPR/Cas9-induced (CTG·CAG)n repeat instability in the myotonic dystrophy type 1 locus: implications for therapeutic genome editing. Mol. Ther. 25, 24–43.

28. Provenzano, C., Cappella, M., Valaperta, R., Cardani, R., Meola, G., Martelli, F., Cardinali, B., and Falcone, G. (2017). CRISPR/Cas9-mediated deletion of CTG expansions recovers normal phenotype in myogenic cells derived from myotonic dystrophy 1 patients. Mol. Ther. Nucleic Acids 9, 337–348.

29. Dastidar, S., Ardui, S., Singh, K., Majumdar, D., Nair, N., Fu, Y., Reyon, D., Samara, E., Gerli, M.F.M., Klein, A.F., et al. (2018). Efficient CRISPR/Cas9-mediated editing of trinucleotide repeat expansion in myotonic dystrophy patient-derived iPS and myogenic cells. Nucleic Acids Res. 46, 8275–8298.

30. Lo Scrudato, M., Poulard, K., Sourd, C., Tomé, S., Klein, A.F., Corre, G., Huguet, A., Furling, D., Gourdon, G., and Buj-Bello, A. (2019). Genome editing of expanded CTG repeats within the human DMPK gene reduces nuclear RNA foci in the muscle of DM1 mice. Mol. Ther. 27, 1372–1388.

31. Zhang, X.H., Tee, L.Y., Wang, X.G., Huang, Q.S., and Yang, S.H. (2015). Off-target effects in CRISPR/Cas9-mediated genome engineering. Mol. Ther. Nucleic Acids 4, e264.

32. Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., and Zhang, F. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 154, 1380–1389.

33. Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., Wang, L., Hodgkins, A., Iyer, V., Huang, X., and Skarnes, W.C. (2014). Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. Nat. Methods 11, 399–402.

34. Canver, M.C., Bauer, D.E., Dass, A., Yien, Y.Y., Chung, J., Masuda, T., Maeda, T., Paw, B.H., and Orkin, S.H. (2014). Characterization of genomic deletion efficiency mediated by clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. J. Biol. Chem. 289, 21312–21324.

35. Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152, 1173–1183.

36. Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 154, 442–451.

37. Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.J., Kii, E., and Alt, F.W. (2015). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. Nat. Biotechnol. 33, 179–186.

38. Hu, J., Meyers, R.M., Dong, J., Panchakshari, R.A., Alt, F.W., and Frock, R.L. (2016). Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. Nat. Protoc. 11, 853–871.

39. 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.

40. Nakamori, M., Taylor, K., Mochizuki, H., Sobczak, K., and Takahashi, M.P. (2015). Oral administration of erythromycin decreases RNA toxicity in myotonic dystrophy. Ann. Clin. Transl. Neurol. 3, 42–54.

41. Rzuczek, S.G., Colgan, L.A., Nakai, Y., Cameron, M.D., Furling, D., Yasuda, R., and Disney, M.D. (2017). Precise small-molecule recognition of a toxic CUG RNA repeat expansion. Nat. Chem. Biol. 13, 188–193.

42. Wheeler, T.M., Leger, A.J., Pandey, S.K., MacLeod, A.R., Nakamori, M., Cheng, S.H., Wentworth, B.M., Bennett, C.F., and Thornton, C.A. (2012). Targeting nuclear RNA for in vivo correction of myotonic dystrophy. Nature 488, 111–115.

43. Jauvin, D., Chrétien, J., Pandey, S.K., Martineau, L., Revillod, L., Bassez, G., Lachon, A., MacLeod, A.R., Gourdon, G., Wheeler, T.M., et al. (2017). Targeting DMPK with antisense oligonucleotide improves muscle strength in myotonic dystrophy type 1 mice. Mol. Ther. Nucleic Acids 7, 465–474.

44. Gao, Y., Guo, X., Santostefano, K., Wang, Y., Reid, T., Zeng, D., Terada, N., Ashizawa, T., and Xia, G. (2016). Genome therapy of myotonic dystrophy type 1 iPS cells for development of autologous stem cell therapy. Mol. Ther. 24, 1378–1387.

45. Wang, Y., Hao, L., Wang, H., Santostefano, K., Thapa, A., Cleary, J., Li, H., Guo, X., Terada, N., Ashizawa, T., and Xia, G. (2018). Therapeutic genome editing for myotonic dystrophy type 1 using CRISPR/Cas9. Mol. Ther. 26, 2617–2630.

46. Wang, D., Zhang, F., and Gao, G. (2020). CRISPR-based therapeutic genome editing: strategies and in vivo delivery by AAV vectors. Cell 181, 136–150.

47. Staahl, B.T., Benekareddy, M., Coulon-Bainier, C., Banfal, A.A., Floor, S.N., Sabo, J.K., Urnes, C., Munares, G.A., Ghosh, A., and Doudna, J.A. (2017). Efficient genome editing in the mouse brain by local delivery of engineered Cas9 ribonucleoprotein complexes. Nat. Biotechnol. 35, 431–434.

48. Finn, J.D., Smith, A.R., Patel, M.C., Shaw, L., Youniss, M.R., van Heteren, J., Dirstine, T., Ciullo, C., Lescarbeau, R., Seitzer, J., et al. (2018). A single administration of CRISPR/Cas9 lipid nanoparticles achieves robust and persistent in vivo genome editing. Cell Rep. 22, 2227–2235.

49. Lee, K., Conboy, M., Park, H.M., Jiang, F., Kim, H.J., Dewitt, M.A., Mackley, V.A., Chang, K., Rao, A., Skinner, C., et al. (2017). Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. Nat. Biomed. Eng. 1, 889–901.

50. Ramakrishna, S., Kwaku Dad, A.B., Beloor, J., Gopalappa, R., Lee, S.K., and Kim, H. (2014). Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. Genome Res. 24, 1020–1027.

51. Gee, P., Lung, M.S.Y., Okuzaki, Y., Sasakawa, N., Iguchi, T., Makita, Y., Hozumi, H., Miura, Y., Yang, L.F., Iwasaki, M., et al. (2020). Extracellular nanovesicles for packaging of CRISPR-Cas9 protein and sgRNA to induce therapeutic exon skipping. Nat. Commun. 11, 1334.

52. Schwab, I., and Nimmerjahn, F. (2013). Intravenous immunoglobulin therapy: how does IgG modulate the immune system? Nat. Rev. Immunol. 13, 176–189.

53. Gopalappa, R., Suresh, B., Ramakrishna, S., and Kim, H.H. (2018). Paired D10A Cas9 nickases are sometimes more efficient than individual nucleases for gene disruption. Nucleic Acids Res. 46, e71.

54. Yee, J.K. (2016). Off-target effects of engineered nucleases. FEBS J. 283, 3239–3248.

55. Alateeq, S., Ovchinnikov, D., Tracey, T., Whitworth, D., Al-Rubaish, A., Al-Ali, A., and Wolvetang, E. (2018). Identification of on-target mutagenesis during correction of a β-thalassemia splice mutation in iPS cells with optimised CRISPR/Cas9-double nickase reveals potential safety concerns. APL Bioeng. 2, 046103.

56. Caldecott, K.W. (2008). Single-strand break repair and genetic disease. Nat. Rev. Genet. 9, 619–631.

57. Kuzminov, A. (2001). Single-strand interruptions in replicating chromosomes cause double-strand breaks. Proc. Natl. Acad. Sci. USA 98, 8241–8246.

58. Ensminger, M., Iloff, L., Ebel, C., Nikolova, T., Kaina, B., and Löbrich, M. (2014). DNA breaks and chromosomal aberrations arise when replication meets base excision repair. J. Cell Biol. 206, 29–43.

59. Cannan, W.J., and Pederson, D.S. (2016). Mechanisms and consequences of double-strand DNA break formation in chromatin. J. Cell. Physiol. 231, 3–14.

60. Adikusuma, F., Piltz, S., Corbett, M.A., Turvey, M., McColl, S.R., Helbig, K.J., Beard, M.R., Hughes, J., Pomerantz, R.T., and Thomas, P.Q. (2018). Large deletions induced by Cas9 cleavage. Nature 560, E8–E9.

61. Shin, H.Y., Wang, C., Lee, H.K., Yoo, K.H., Zeng, X., Kuhns, T., Yang, C.M., Mohr, T., Liu, C., and Hennighausen, L. (2017). CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. Nat. Commun. 8, 15464.

62. Kosicki, M., Tomberg, K., and Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. Nat. Biotechnol. 36, 765–771.

63. Owens, D.D.G., Caulder, A., Frontera, V., Harman, J.R., Allan, A.J., Bucakci, A., Greder, L., Codner, G.F., Hublitz, P., McHugh, P.J., et al. (2019). Microhomologies are prevalent at Cas9-induced larger deletions. Nucleic Acids Res. 47, 7402–7417.

64. Giannoukos, G., Ciulla, D.M., Marco, E., Abdulkerim, H.S., Barrera, L.A., Bothmer, A., Dhanapal, V., Gloskowski, S.W., Jayaram, H., Maeder, M.L., et al. (2018). UDiTaS™, a genome editing detection method for indels and genome rearrangements. BMC Genomics 19, 212.

65. Nelson, C.E., Wu, Y., Gemberling, M.P., Oliver, M.L., Waller, M.A., Bohning, J.D., Robinson-Hamm, J.N., Bulaklak, K., Castellanos Rivera, R.M., Collier, J.H., et al. (2019). Long-term evaluation of AAV-CRISPR genome editing for Duchenne muscular dystrophy. Nat. Med. 25, 427–432.

66. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., and Taipale, J. (2018). CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. Nat. Med. 24, 927–930.

67. Ihry, R.J., Worringer, K.A., Salick, M.R., Frias, E., Ho, D., Theriault, K., Kommineni, S., Chen, J., Sondey, M., Ye, C., et al. (2018). p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. Nat. Med. 24, 939–946.

68. Radzisheuskaya, A., Shlyueva, D., Müller, I., and Helin, K. (2016). Optimizing sgRNA position markedly improves the efficiency of CRISPR/dCas9-mediated transcriptional repression. Nucleic Acids Res. 44, e141.

69. Carrell, S.T., Carrell, E.M., Auerbach, D., Pandey, S.K., Bennett, C.F., Dirksen, R.T., and Thornton, C.A. (2016). Dmpk gene deletion or antisense knockdown does not compromise cardiac or skeletal muscle function in mice. Hum. Mol. Genet. 25, 4328–4338.

70. Pinto, B.S., Saxena, T., Oliveira, R., Méndez-Gómez, H.R., Cleary, J.D., Denes, L.T., McConnell, O., Arboleda, J., Xia, G., Swanson, M.S., and Wang, E.T. (2017). Impeding transcription of expanded microsatellite repeats by deactivated Cas9. Mol. Cell 68, 479–490.e5.

71. Batra, R., Nelles, D.A., Pirie, E., Blue, S.M., Marina, R.J., Wang, H., Chaim, I.A., Thomas, J.D., Zhang, N., Nguyen, V., et al. (2017). Elimination of toxic microsatellite repeat expansion RNA by RNA-targeting Cas9. Cell 170, 899–912.e10.

72. Thakore, P.I., Kwon, J.B., Nelson, C.E., Rouse, D.C., Gemberling, M.P., Oliver, M.L., and Gersbach, C.A. (2018). RNA-guided transcriptional silencing in vivo with S. aureus CRISPR-Cas9 repressors. Nat. Commun. 9, 1674.

73. Noguchi, S., Arakawa, T., Fukuda, S., Furuno, M., Hasegawa, A., Hori, F., Ishikawa-Kato, S., Kaida, K., Kaiho, A., Kanamori-Katayama, M., et al. (2017). FANTOM5 CAGE profiles of human and mouse samples. Sci. Data 4, 170112.

74. Hu, Y., Yan, C., Hsu, C.H., Chen, Q.R., Niu, K., Komatsoulis, G.A., and Meerzaman, D. (2014). OmicCircos: a simple-to-use R package for the circular visualization of multidimensional omics data. Cancer Inform. 13, 13–20.