Germline mutations in ABL1 cause an autosomal dominant syndrome characterized by congenital heart defects and skeletal malformations

Xia Wang1,2,11, Wu-Lin Charng1,11, Chun-An Chen1,3,11, Jill A Rosenfeld1, Aisha Al Shamsi4, Lihadh Al-Gazali5, Marianne McGuire2, Nicholas Ah Mew6, Georgianne L. Arnold7, Chunjing Qu2, Yan Ding8, Donna M Muzny8, Richard A Gibbs1,8, Christine M Eng1,2, Magdalena Walkiewicz1,2, Fan Xia1,2, Sharon E Plon1,9,10, James R Lupski1,8,9, Christian P Schaaf1,3,12 & Yaping Yang1,2,12

ABL1 is a proto-oncogene well known as part of the fusion gene BCR-ABL1 in the Philadelphia chromosome of leukemia cancer cells1. Inherited germline ABL1 changes have not been associated with genetic disorders. Here we report ABL1 germline variants cosegregating with an autosomal dominant disorder characterized by congenital heart disease, skeletal abnormalities, and failure to thrive. The variant c.734A>G (p.Tyr245Cys) was found to occur de novo or cosegregate with disease in five individuals (families 1–3). Additionally, a de novo c.1066G>A (p.Ala356Thr) variant was identified in a sixth individual (family 4). We overexpressed the mutant constructs in HEK 293T cells and observed increased tyrosine phosphorylation, suggesting increased ABL1 kinase activities associated with both the p.Tyr245Cys and p.Ala356Thr substitutions. Our clinical and experimental findings, together with previously reported teratogenic effects of selective BCR-ABL1 inhibitors in humans2,3 and developmental defects in ABL1 knockout mice4,5, suggest that ABL1 has an important role during organizational development.

Somatic and, more rarely, germline variants leading to activation or misexpression of proto-oncogenes have long been identified as important cancer drivers. Given the many important functions of proto-oncogenes, it is not surprising that germline variants in these genes can also cause human developmental disorders6–12.

We identified germline variants in ABL1 by exome sequencing and Sanger sequencing in six affected individuals who shared similar clinical features, including dysmorphic facial features (6/6), congenital heart disease (CHD; 6/6), skeletal abnormalities (6/6), joint problems (5/6), failure to thrive (5/6), gastrointestinal problems (5/6), and male genital/sexual abnormalities (3/4) (Supplementary Table 1 and Supplementary Note). In younger children, dysmorphic features included a broad forehead, small nose, deep-set eyes, and small chin. In older individuals, the face appeared elongated, with a narrow maxilla, long and narrow nose, and pointed chin (Fig. 1a). Common skeletal abnormalities included pectus excavatum, scoliosis, finger/toe deformities, in particular hindfoot deformity, and finger contractures (Fig. 1b). The CHD observed included atrial and ventricular septal defects and, in older individuals, aortic root dilation. Three patients had joint hyperextensibility/laxity. Most individuals had failure to thrive during infancy and early childhood (Supplementary Fig. 1). Hypospadias/hypongonadism was reported in three of four male patients. On the basis of cardiac and skeletal manifestations, differential diagnoses commonly included connective tissue disorders, for example, Marfan syndrome in family 3.

Clinical exome sequencing was performed on the four probands, as previously described (Supplementary Table 2)13,14. Exome sequencing and Sanger fill-in of regions poorly covered by exome sequencing did not identify any causal variants in known disease-relevant genes (Supplementary Tables 3 and 4). Instead, all probands were found to carry novel heterozygous nonsynonymous variants in ABL1 (NM_007313.2, isoform 1b). One variant, c.734A>G (p.Tyr245Cys), was recurrent in families 1–3, while the other variant, c.1066G>A (p.Ala356Thr), was heterozygous in a single proband from family 4 (Figs. 2 and 3, Table 1 and Supplementary Fig. 2). Sanger sequencing of the asymptomatic parents of the probands in families 1 and 2 was negative for p.Tyr245Cys, indicating that the variant arose de novo in the probands. The similarly affected daughter of subject 1 was found to have inherited the heterozygous p.Tyr245Cys substitution. In family 3, p.Tyr245Cys was found to be inherited...
from the similarly affected father. In family 4, p.Ala356Thr was found to be de novo in the proband.

Unlike many recurrent variants that occur at CpG dinucleotides, the recurrent c.734A>G variant is not located within a CpG site. This variant has been detected in four unrelated but similarly affected individuals (three probands in this report and a fourth individual who was not included owing to lack of consent) from approximately 6,900 consecutive and unrelated individuals referred for clinical exome sequencing. There is significant enrichment of this variant in our affected patient cohort as compared to 0/60,700 in the database from the Exome Aggregation Consortium (ExAC; accessed July 2016; P = 3 × 10⁻⁹, χ² test; Supplementary Data 1).

ABL1 has two isoforms, 1a (NM_005157.5) and 1b (NM_007313.2), as a result of an alternatively spliced first exon. Isoform 1b has 19 additional N-terminal residues that are absent in isoform 1a; myristoylation of these residues has a role in autoinhibition of the kinase activity. The two missense variants in our patients are expected to result in increased overall phosphotyrosine levels and increased phosphorylation of specific ABL1 substrates (STAT5) when compared to cells expressing wild-type protein (Fig. 4 and Supplementary Figs. 3–5). These results indicate that both the c.734A>G (p.Tyr245Cys) and c.1066G>A (p.Ala356Thr) variants cause increased phosphorylation, suggesting increased ABL1 kinase activity.

Of note, Tyr245 and Ala356 are both known as key residues regulating ABL1 kinase activity. The Tyr245 residue is one of the two tyrosine residues required for autophosphorylation-induced activation of ABL1 intrinsic kinase activity. It was shown that, while autophosphorylation causes an 18-fold increase in the activity of wild-type ABL1, introduction of p.Tyr245Phe, which affects the same residue as p.Tyr245Cys in families 1–3, results in inhibition of such activation by 50% (ref. 15). This is in contrast to the increased phosphorylation associated with the p.Tyr245Cys substitution found in this study. To exclude the possibility of systematic differences between our experimental settings and those reported previously, we overexpressed the reported mutant construct. Consistent with the previous report, we observed decreased overall phosphotyrosine for Tyr245Phe when compared to the wild-type protein for both isoforms (Supplementary Figs. 6 and 7). These results support the contention that the p.Tyr245Phe and p.Tyr245Cys substitutions have opposite effects (gain of function versus loss of function) on ABL1 kinase activities.

The Ala356 residue is located in the myristoyl-binding site of the ABL1 kinase domain, which intramolecularly binds the N-terminal myristoyl group and forms an autoinhibitory conformation. It was shown that p.Ala356Asn, which affects the same amino acid as p.Ala356Thr identified in family 4, resulted in higher kinase activity than that seen for wild-type protein16. Our findings on the Ala356Thr mutant suggest that these two changes affecting Ala356, that is, to asparagine versus threonine, have similar effects. Of note, residue Ala356 in isoform 1b is located in the binding site for the N-terminal myristoyl group, which is unique to isoform 1b. However, our data showed that the substitution of this residue in isoform 1a, p.Ala337Thr, also caused increased phosphorylation (Supplementary Fig. 5).

To test whether mutant ABL1 transcript is expressed in the affected individuals, we conducted semiquantitative RT–PCR on lymphoblast-derived cell lines from affected subjects 1 and 2 (F1:II1 and F1:III1) and an unaffected control (F1:II2) (cell lines from other individuals not available). Sanger sequencing of the RT–PCR products showed

**Figure 1** The facial and skeletal features of subjects 1 (F1:II1), 2 (F1:II1), 3 (F2:II1), and 6 (F4:II1). (a) Facial features. From left to right: subjects 1, 2, 3, and 6. Note long face with narrow maxilla and pointed chin in subjects 1 and 3. Younger individuals (subjects 2 and 6) manifest broad forehead, deep-set eyes, small nose, and small chin. (b) Skeletal abnormalities. From left to right: subjects 1, 2, 3, and 6. Abnormalities include pectus excavatum, scoliosis, hindfoot deformity causing pes planus, and finger contractures. Informed consent was obtained from subject 1 and from the parents/legal guardians of subjects 2, 3, and 6 to publish these photographs.

**Figure 2** Identification of ABL1 variants in affected families. Pedigrees are shown for the four families. The genotype is shown below each individual in the pedigrees, with “+” representing the reference allele and “M” representing the mutant allele. Individuals without genotype symbols did not have samples available for genotyping.

Endogenous levels of overall phosphotyrosine (p-Tyr) and phosphorylation of specific ABL1 substrates (STAT5) were measured by immunoblotting. Overexpression of the mutant constructs for both isoforms resulted in increased overall phosphotyrosine levels and increased phosphorylation of specific ABL1 substrates (STAT5) when compared to cells expressing wild-type protein (Fig. 4 and Supplementary Fig. 5). These results indicate that both the c.734A>G (p.Tyr245Cys) and c.1066G>A (p.Ala356Thr) variants cause increased phosphorylation, suggesting increased ABL1 kinase activity.

**Figure 2** Identification of ABL1 variants in affected families. Pedigrees are shown for the four families. The genotype is shown below each individual in the pedigrees, with “+” representing the reference allele and “M” representing the mutant allele. Individuals without genotype symbols did not have samples available for genotyping.

Endogenous levels of overall phosphotyrosine (p-Tyr) and phosphorylation of specific ABL1 substrates (STAT5) were measured by immunoblotting. Overexpression of the mutant constructs for both isoforms resulted in increased overall phosphotyrosine levels and increased phosphorylation of specific ABL1 substrates (STAT5) when compared to cells expressing wild-type protein (Fig. 4 and Supplementary Fig. 5). These results indicate that both the c.734A>G (p.Tyr245Cys) and c.1066G>A (p.Ala356Thr) variants cause increased phosphorylation, suggesting increased ABL1 kinase activity.

Of note, Tyr245 and Ala356 are both known as key residues regulating ABL1 kinase activity. The Tyr245 residue is one of the two tyrosine residues required for autophosphorylation-induced activation of ABL1 intrinsic kinase activity. It was shown that, while autophosphorylation causes an 18-fold increase in the activity of wild-type ABL1, introduction of p.Tyr245Phe, which affects the same residue as p.Tyr245Cys in families 1–3, results in inhibition of such activation by 50% (ref. 15). This is in contrast to the increased phosphorylation associated with the p.Tyr245Cys substitution found in this study. To exclude the possibility of systematic differences between our experimental settings and those reported previously, we overexpressed the reported mutant construct. Consistent with the previous report, we observed decreased overall phosphotyrosine for Tyr245Phe when compared to the wild-type protein for both isoforms (Supplementary Figs. 6 and 7). These results support the contention that the p.Tyr245Phe and p.Tyr245Cys substitutions have opposite effects (gain of function versus loss of function) on ABL1 kinase activities.

The Ala356 residue is located in the myristoyl-binding site of the ABL1 kinase domain, which intramolecularly binds the N-terminal myristoyl group and forms an autoinhibitory conformation. It was shown that p.Ala356Asn, which affects the same amino acid as p.Ala356Thr identified in family 4, resulted in higher kinase activity than that seen for wild-type protein16. Our findings on the Ala356Thr mutant suggest that these two changes affecting Ala356, that is, to asparagine versus threonine, have similar effects. Of note, residue Ala356 in isoform 1b is located in the binding site for the N-terminal myristoyl group, which is unique to isoform 1b. However, our data showed that the substitution of this residue in isoform 1a, p.Ala337Thr, also caused increased phosphorylation (Supplementary Fig. 5).

To test whether mutant ABL1 transcript is expressed in the affected individuals, we conducted semiquantitative RT–PCR on lymphoblast-derived cell lines from affected subjects 1 and 2 (F1:II1 and F1:III1) and an unaffected control (F1:II2) (cell lines from other individuals not available). Sanger sequencing of the RT–PCR products showed
and osteoporosis. We reevaluated the ventricular wall thickness of knockin mice have growth delay, cardiac hyperplasia, and skeletal anomalies, including scoliosis and vertebral defects, CHD (2), hypospadias (2), and pyloric stenosis (2). On the basis of the small number of cases, no conclusive definitions of the potential teratogenicity of imatinib can be drawn, especially as some of the cases were also exposed to additional teratogens, such as Coumadin/warfarin or hydroxyurea. Others have also reported potential teratogenic effects of imatinib, including CHD, skeletal anomalies, hypospadias, intestinal malrotation, and imperforate anus. This spectrum of phenotypes overlaps with what was found in our patients, including scoliosis (3/6), CHD (6/6), hypospadias/hypospadiasism (3/4), pyloric muscle thickening (1/6), and imperforate anus (1/6). These similarities between fetuses exposed to imatinib during pregnancy and human patients with constitutional ABL1 variants, along with the studies in mice, suggest that ABL1 function needs to be tightly regulated during development and that dysregulation can cause congenital malformations.

Our data suggest a novel genetic syndrome caused by constitutional ABL1 variants, which affects growth and the cardiovascular and skeletal systems. The differential diagnosis of the newly described syndrome includes Shprintzen–Goldberg syndrome (dolichostenomelia, arachnodactyly, pectus deformity, scoliosis, aortic root enlargement, and high-arched palate) and Loeser–Dietz syndrome (long face, high-arched palate, microretrognathia, pectus deformity, scoliosis, arachnodactyly, joint laxity, and aortic root aneurysm with risk of dissection). Given the evidence of increased transforming growth

Table 1 Summary of ABL1 variants identified in the subjects

| Family | Coordinate (hg19) Location | Nucleotide change (amino acid change)a | In silico predictions20-22 | NGS reads in proband (mutant: normal) | Functional consequences in this study | Previous functional studies on Try245 and Ala356 |
|--------|-----------------------------|--------------------------------------|--------------------------|----------------------------------------|-------------------------------------|-----------------------------------------------|
| 1      | Chr9: 133738277 Exon 4      | c.734A>G (p.Tyr245Cys)               | Damaging by SIFT; probably damaging by PolyPhen-2; disease causing by MutationTaster | 75:69 De novo in proband, transmitted to affected daughter | Increase in tyrosine phosphorylation | Tyrosine-to-phenylalanine substitution at the same Try245 residue leads to inhibition of autophosphorylation-induced activation of ABL1 intrinsic kinase activity15 |
| 2      | Chr9: 133738277 Exon 4      | c.734A>G (p.Tyr245Cys)               | 82:71 De novo in proband |
| 3      | Chr9: 133738277 Exon 4      | c.734A>G (p.Tyr245Cys)               | 310:255 Proband inherited the variant from affected father |
| 4      | Chr9: 133748348 Exon 6      | c.1066G>A (p.Ala356Thr)              | Damaging by SIFT; probably damaging by PolyPhen-2; disease causing by MutationTaster | 46:68 De novo in proband | Increase in tyrosine phosphorylation | Alanine-to-asparagine substitution at the same Ala356 residue leads to higher ABL1 kinase activity16 |

The two variants are absent in the dbSNP, ESP, ExAC, and COSMIC databases. NGS, next-generation sequencing.

aThe GenBank transcript ID used for the nucleotide and amino acid changes in isoform 1b was NM_007313.2.

Figure 3 In silico analysis of the two ABL1 variants identified in this study. (a) The Tyr245 and Ala356 residues are conserved from human to zebrafish (prepared on the basis of Ensembl browser genomic alignments). (b) Schematic view of the ABL1 1b protein isoform and its domains. The Tyr245 residue is located in the linker region between the SH2 and kinase domains, while the Ala356 residue is located in the kinase domain (based on structural data from Nagar et al., prepared on the basis of UniProt database domains, P00519). (c) 3D structure of the ABL1 1b protein isoform, its N-terminal domains, and the location of the two mutated ABL1 residues (based on structural data from Nagar et al., prepared on the basis of Ensembl browser genomic alignments).
Malignancies have not been detected in any of the affected individuals with germline ABL1 variants, but, as these variants cause increased phosphorylation and possibly increased kinase activity of this proto-oncogene, it may be prudent to perform systematic clinical screening for ABL1-associated cancers in the newly identified individuals.

In conclusion, we report that germline variants in ABL1 cause a syndrome characterized by CHD, skeletal abnormalities, and failure to thrive. ABL1 joins the growing list of genes that are implicated in both cancer and human developmental disorders.

**METHODS**

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

**ACKNOWLEDGMENTS**

We thank the families for their participation and collaboration. This work was funded in part by US National Human Genome Research Institute (NHGRI)/National Heart, Lung, and Blood Institute (NHLBI) grant U54HG006542 to the Baylor-Hopkins Center for Mendelian Genomics (BH-CMG) and the NIH Common Fund, through the Office of Strategic Coordination/Office of the NIH Director under award U01HG007970. W.-L.C. was supported by Cancer Prevention Research Institute of Texas (CPRIT) training program RP140102. The work of N.A.M. is supported by the Rashid Family Fund.

**AUTHOR CONTRIBUTIONS**

X.W., Y.Y., W.-L.C., C.P.S., C.-A.C., and J.R.L. designed the studies. X.W., Y.Y., C.P.S., W.-L.C., J.R.L., S.E.P., J.A.R., and F.X. participated in the writing of the manuscript. X.W., C.Q., Y.D., D.M.M., R.A.G., C.M.E., M.W., F.X., and Y.Y. performed sequencing data analysis. W.-L.C. and C.-A.C. performed the kinase activity assays. X.W., J.A.R., A.A.S., L.A.-G., M.M., N.A.M., G.A., C.P.S., and Y.Y. recruited patients and gathered detailed clinical information for the study. Y.Y., C.P.S., and J.R.L. supervised the studies.

**COMPETING FINANCIAL INTERESTS**

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.
11. Aoki, Y. et al. Germline mutations in HRAS proto-oncogene cause Costello syndrome. Nat. Genet. 37, 1038–1040 (2005).
12. Lahiry, P., Torkamani, A., Schork, N.J. & Hegele, R.A. Kinase mutations in human disease: interpreting genotype-phenotype relationships. Nat. Rev. Genet. 11, 60–74 (2010).
13. Yang, Y. et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N. Engl. J. Med. 369, 1502–1511 (2013).
14. Yang, Y. et al. Molecular findings among patients referred for clinical whole-exome sequencing. J. Am. Med. Assoc. 312, 1870–1879 (2014).
15. Brasher, B.B. & Van Etten, R.A. c-Abl has high intrinsic tyrosine kinase activity that is stimulated by mutation of the Src homology 3 domain and by autophosphorylation at two distinct regulatory tyrosines. J. Biol. Chem. 275, 35631–35637 (2000).
16. Hantschel, O. et al. A myristoyl/phosphotyrosine switch regulates c-Abl. Cell 112, 845–857 (2003).
17. Hardin, J.D., Boast, S., Mendelsohn, M., de los Santos, K. & Goff, S.P. Transgenes encoding both type I and type IV c-abl proteins rescue the lethality of c-abl mutant mice. Oncogene 12, 2669–2677 (1996).
18. Nagar, B. et al. Structural basis for the autoinhibition of c-Abl tyrosine kinase. Cell 112, 859–871 (2003).
19. Guex, N. & Peitsch, M.C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18, 2714–2723 (1997).
20. Kumar, P., Henikoff, S. & Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 4, 1073–1081 (2009).
21. Adzhubei, I.A. et al. A method and server for predicting damaging missense mutations. Nat. Methods 7, 248–249 (2010).
22. Schwarz, J.M., Cooper, D.N., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat. Methods 11, 361–362 (2014).
Exome sequencing. Exome sequencing was performed as previously described. Briefly, genomic DNA samples were fragmented, ligated to Illumina multiplexing paired-end adaptors, amplified, and hybridized to a solution-based exome capture reagent (Roche NimbleGen). Paired-end sequencing (2 × 100 bp) was performed on the Illumina HiSeq 2500 platform to provide a mean sequence coverage of about 120×, with about 97% of the target bases having at least 20× coverage (Supplementary Table 2).

Sequencing data analysis and annotation. Exome sequencing data were processed and variants were annotated as previously described. Briefly, the output data from the Illumina HiSeq 2500 were converted from a bcl file to a FastQ file by Illumina Consensus Assessment of Sequence and Variation software version 1.8.3 and mapped to the human reference genome using the BWA program. Variants were called by Atlas-SNP and Atlas-indel software version 1.8.3 and mapped to the human reference genome using the BW A program. Variants were called by Atlas-SNP and Atlas-indel. An in-house software program, CASSANDRA, was used for variant filtering and annotation.

cDNA constructs. A cDNA clone for isoform 1b of ABL1 (MHS6278-211687872) was purchased from GE Dharmacon (MGC cDNA collection). The cDNA for isoform 1a of ABL1 was obtained by PCR of the isoform 1b cDNA using primers specific to the 1a isoform. All point mutations were introduced using primers specific to the 1a isoform. All constructs were verified by Sanger sequencing (Supplementary Fig. 3).

Wild-type and mutant human ABL1 cDNA constructs were cloned into pcDNA3.1/V5-His A (encoding C-terminal V5 and His tags; Clontech) using KpnI and XhoI restriction enzyme sites. Residues Tyr245 and Ala356 in isoform 1b correspond to residues Tyr226 and Ala337 in isoform 1a, respectively. Restriction enzymes were from NEB, and DNA purification kits were from Invitrogen and Qiagen.

Cell culture and immunoblotting. The HEK 293T cell line (a gift from H. Zoghbi at Baylor College of Medicine; no testing for mycoplasma) was cultured in DMEM (Corning, 45000-312) supplemented with 10% FBS, 1-glutamine, and antibiotics–antimycotics at 37 °C under 5% CO2. Transfection was performed using Lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacturer’s instructions. Cells were collected 24 h after transfection and lysed in lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5% glycerol, 0.5% Triton X-100, protease inhibitor cocktail and phosphatase inhibitor Xpert (Gendepot)). The lysate supernatant was boiled in Laemmli buffer. Polyacrylamide gel electrophoresis, transfer, and immunoblotting were performed according to standard protocols. Primary antibodies were used at the following dilutions: mouse anti-c-Ab1 (1:100 dilution; EMD Millipore, OP-20), mouse anti-phosphotyrosine (4G10) (1:20,000 dilution; EMD Millipore, 05-321), PathScan Bcr/Abl antibody mixture (1:400 dilution; Cell Signaling Technology, 5300), mouse anti-V5 (1:10,000 dilution; Thermo Fisher Scientific, 460705), mouse anti-GAPDH (1:20,000 dilution; Advanced ImmunoChemical, 2-RGM2), and rabbit anti-pSmad2/3 (1:1,000 dilution; Cell Signaling Technology, 8828). Goat anti-rabbit HRP-conjugated secondary antibody (Bio-Rad, 170-5046) was used at a 1:10,000 dilution, and donkey anti-mouse HRP-conjugated secondary antibody (Jackson ImmunoResearch, 715-035-150) was used at a 1:10,000 dilution.

The immunoblotting analysis was performed in three technical replicates from separate transfections of the same cell culture. P values were calculated using one-way ANOVA with Tukey’s post hoc analysis. The exact numbers used for the calculations are provided in Supplementary Data 1. The immunoblotting data shown in Figure 4 and Supplementary Figures 4–7 were replicated three times in the laboratory.

Semi quantitative RT–PCR on lymphoblast cell lines. Total RNA was extracted from human lymphoblast cell lines derived from subjects 1 and 2, and the control with the miNeasy Mini kit (Qiagen) following the manufacturer’s instructions. RNA was quantified with a NanoDrop 1000 (Thermo Fisher). Three micrograms of total RNA was reverse transcribed to cDNA using a reverse transcription kit (Qiagen). cDNA from each sample was then analyzed by RT–PCR followed by Sanger sequencing. Semi quantitative PCR cycle numbers were as follows: ABL1 (exon 3–exon 5), 35 cycles; ABL1 (exon 3–exon 6), 35 cycles; GAPDH, 25 cycles. PCR primer sequences are available in Supplementary Table 5.

Data availability. The c.734A>G (p.Tyr245Cys) and c.1066G>A (p.Ala356Thr) variants have been submitted to the ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/) with accession codes SCV000485092 and SCV000485097, respectively. The raw whole-exome sequencing data that support the findings of this study are available on request from the corresponding author (Y.Y.); they are not publicly available as they contain information that could compromise research participant privacy. All other data generated or analyzed during this study are included in this published article (and its supplementary information files).

23. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589–595 (2010).
24. Shen, Y. et al. A SNP discovery method to assess variant allele probability from next-generation resequencing data. Genome Res. 20, 273–280 (2010).