OPEN ACCESS

Original research

Diagnosis complexity of dentinogenesis imperfecta involving *DSPP* genetic variants

Gaétan Caravello,[1,2] Alexandra Jiménez-Armijo ,[1,3] Marzena Kawczynski,[1,3] Tristan Rey,[1,2] on behalf of O-Rares Consortium, Manuela Antin,[2] Alison Roth,[2] Jean-Baptiste Lamouche,[2] Aurélie Gouronc,[2] Abdelmalek Alioua,[4] Caroline Schluth-Bolard ,[2,5] Agnès Bloch-Zupan[1,3,6,7]

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/jmg-2025-111345).

For numbered affiliations see end of article.

Correspondence to
Professor Agnès Bloch-Zupan; agnes.bloch-zupan@unistra.fr

GC and AJ-A are joint first authors.

Received 4 November 2025
Accepted 13 February 2026

ABSTRACT

Background Variants in the dentin sialophosphoprotein (*DSPP*) gene are associated with dentin dysplasia type II (DD-II; OMIM # 125420) and dentinogenesis imperfecta (DI) types II (OMIM # 125490) and III (OMIM # 125500). *DSPP* encodes a precursor protein cleaved into three dentin matrix proteins: dentin sialoprotein (DSP), dentin phosphoprotein/phosphophoryn (DPP) and dentin glycoprotein (DGP). Exon 5 contains over 200 tandem 9-base pair repeats (DSS domain), complicating sequencing with standard methods.

Materials and methods We studied 112 individuals (42 index cases and 70 relatives) with clinical signs of DI or DD. DNA extracted from saliva was analysed using the GenoDENT next-generation sequencing panel. For inconclusive cases, long-range PCR and Oxford Nanopore Technology (ONT) long-read sequencing were used to overcome limitations in analysing the repetitive *DSPP* region.

Results Pathogenic or likely pathogenic *DSPP* variants were identified in 41 families, including 8 known and 14 novel variants. Most were in exon 5, causing frameshifts resulting in a −1 reading-frame shift with a hydrophobic C-terminal extension and termination at a downstream stop codon. ONT sequencing enabled detection in cases where short-read methods failed. Several variants showed familial segregation and variable expressivity.

Conclusion This study demonstrates the value of long-read sequencing to resolve complex *DSPP* regions and expands the variant spectrum. The variability in clinical presentation suggests the influence of modifier factors, warranting further genotype–phenotype studies.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ The *DSPP* gene is responsible for inherited dentinogenesis imperfecta (DI) and dentin dysplasia (DD). Its highly repetitive exon 5 region poses major challenges for standard sequencing methods, often resulting in missed or inconclusive variant detection.

WHAT THIS STUDY ADDS

⇒ Using a combination of short-read and long-read Oxford Nanopore sequencing, we identified 22 pathogenic variants, including 14 novel ones. Long-read sequencing outperformed conventional methods in detecting variants.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ These results promote the use of long-read sequencing in diagnostics, especially when targeting repetitive genes like *DSPP*. Clinically, it provides better genetic counselling and fosters further research into phenotype modifiers and precision diagnostics.

INTRODUCTION

Dentin is a mineralised tissue formed by ectomesenchymal cells called odontoblasts, with an extracellular matrix similar to bone, sharing type I collagen as their major organic matrix component, although they differ in their principal non-collagenous proteins: osteocalcin (BGLAP) for bone and dentin sialophosphoprotein (DSPP) for dentin. It presents dentinal tubules packed closely together containing the odontoblast cytoplasmic extensions. The odontoblast cell bodies are situated in the periphery of the pulp tissue. Dentin is composed of inorganic (70%) and organic matrix (20%) and water (10%). Several non-collagenous proteins are present in the organic matrix. The major non-collagenous dentin matrix proteins, dentin phosphoprotein/phosphophoryn (DPP), dentin sialoprotein (DSP) and dentin glycoprotein (DGP), are encoded by the *DSPP* gene.[1] Variants in the *DSPP* gene are responsible for dentin dysplasia type II (DD-II; OMIM # 125420; ORPHA 1653) and dentinogenesis imperfecta (DI) types II (OMIM # 125490; ORPHA 166260) and III (OMIM # 125500; ORPHA 166265).[1 2] In patients with DI, both dentitions (primary and permanent) are affected, showing amber-brown discolouration, attrition and short, tapered roots with obliterated pulp chambers.[3–6] While the primary teeth in DD-II show a similar phenotype to DI, the permanent dentition usually exhibits a nearly normal clinical appearance with thistle-tube deformity of the pulp chambers and pulp stones in radiological findings.[3 6]

The *DSPP* gene, expressed predominantly in odontoblasts but also at lower levels in bones and ductal epithelial cells,[7–10] is part of a tooth/bone gene cluster on chromosome 4. It comprises five exons, with the first exon being non-coding. *DSPP* stands as the largest member within the small integrin-binding ligand N-linked glycoprotein

Check for updates

To cite: Caravello G, Jiménez-Armijo A, Kawczynski M, *et al*. *J Med Genet* Epub ahead of print: [please include Day Month Year]. doi:10.1136/jmg-2025-111345

(SIBLING) gene family, characterised by the presence of the integrin-binding tripeptide, arginylglycylaspartic acid (RGD).[11] A distinctive feature of *DSPP* compared with other SIBLING family members is the presence of over 200 tandem copies of a nominal 9-basepair (bp) repeat encoded within exon 5. This repeat comprises a series of aspartic acid (D) and phosphoserine (S) repeats (DSS). DSPP is highly negatively charged, conferring it the ability to bind calcium ions (Ca^{+2}) and thus form hydroxyapatite that participates in dentin mineralisation.[12 13]

Sequencing genes characterised by highly repeated sequences is challenging due to the inherent complexities involved in deciphering repetitive elements. The presence of numerous identical or nearly identical repetitions within these sequences hampers the accurate determination of their specific arrangement and length. Traditional sequencing techniques, such as Sanger sequencing and short-read GenoDENT next-generation sequencing (NGS), often struggle to accurately resolve highly repetitive regions, leading to ambiguities in the obtained sequences. These challenges are exacerbated when attempting to sequence long stretches of DNA, as errors or omissions in repetitive regions may compromise the fidelity of the entire sequence. Although technological advancements have enhanced our ability to analyse repetitive sequences, accurately reconstructing these regions remains a formidable task.[14 15]

DSPP conforms to this pattern, exhibiting extensively repeated sequences within a region widely recognised for harbouring pathogenic variants. This characteristic adds complexity to patient diagnosis, posing challenges in discerning and interpreting the genetic information precisely. Sequencing across the repetitive region of *DSPP* is imperative for elucidating the genetic underpinnings of inherited dentin defects and unravelling their pathogenic mechanisms. The conventional approach involved sequencing the DPP repetitive region by cloning PCR amplification products and using the resultant plasmid DNA as a template.[12] More recently, the use of single-molecule real-time (SMRT) DNA sequencing showed good accuracy to resolve these complex regions, allowing the characterisation of novel frameshift variants involved in DI-II.[5 16] Alternatively, Oxford Nanopore Technology (ONT) long-read sequencing has been used. Although ONT exhibits a higher per-base error rate than SMRT sequencing, it provides flexible and ultra-long read lengths, rapid and simplified library preparation and scalable, cost-effective throughput that is well suited for routine diagnostic use.

In this study, we report genetic results for 112 individuals, including 42 index cases presenting with dentin defects and 70 associated relatives. Overall, a molecular diagnosis was established in 41 families. Using the GenoDENT next-generation short-read sequencing panel,[17 18] diagnoses were obtained for seven families by NGS alone, while variants were confirmed in additional families by complementary approaches: classic Sanger sequencing (8 families), long-range PCR followed by Sanger sequencing (6 families) or ONT long-read sequencing (10 families). For cases in which NGS results were inconclusive due to limited read depth and high mismatch rates or negative, ONT long-read sequencing was applied and enabled diagnosis in 10 additional families. Altogether, ONT sequencing contributed to the reliable characterisation of *DSPP* variants in 20 families (figure 1).

Based on these results, we propose a diagnostic workflow (figure 3E) in which ONT sequencing serves as an effective second-line strategy for resolving *DSPP* variants in cases where short-read sequencing is inconclusive, thereby improving the molecular diagnosis of DI and DD.

METHODS

Patient's phenotype

Patients presenting symptoms of DI or DD were recruited and examined between 2016 and 2024 in the Reference Center (CRMR) for rare oral and dental diseases in Strasbourg, France, or in one of the affiliated Centers of the French network O-Rares, Filière TETECOU or directly by their treating medical practitioner. When possible, parents and relatives were also included in the study. Oral phenotype was documented using the D[4]/phenodent registry protocol, previously described in Bloch-Zupan *et al*.[19]

Each patient presenting with DI or DD, as well as their non-affected family member, gave written informed consent in accordance with the Declaration of Helsinki, both for the registry and for genetic analyses performed on salivary samples. The terms used to describe dental anomalies were defined in de La Dure-Molla *et al*.[20]

Sample collection and DNA extraction

Saliva samples were sent to the genetic diagnostic laboratory at the University Hospitals of Strasbourg for genetic investigation. All the samples were collected using the Oragene DNA kit (DNA Genotek, Canada kit). Genomic DNA was extracted using two different protocols depending on the volume of saliva: manually for volumes <1 mL or using an automated extraction protocol for volumes ≥ 1 mL. The manual extraction procedure is described in Rey *et al* (2019).[18] Automated extraction was performed using the QIAsymphony SP automate and the QIAsymphony DSP DNA Midi Kit (#937255, QIAGEN). The setup of the QIAsymphony SP was done according to the manufacturer's instructions. A customised protocol for DNA extraction was applied following the Oragene DNA Genotek protocol (ID372, available at https://www.dnagenotek.com/ROW/pdf/MK-AN-00029.pdf).

Extracted DNA was directly quantified using the Qubit V.2.0 and the dsDNA BR Assay (#Q32850) or dsDNA HS Assay (#Q32851) Kit (Thermo Fisher Scientific) depending on the DNA sensitivity. The dsDNA HS Assay Kit provides an accurate and selective method for the quantification of sensitive DNA samples and is designed to be accurate for initial DNA sample concentrations from 0.005 to 120 ng/µL, while the dsDNA BR Assay Kit is designed to be accurate for initial DNA sample concentrations from 0.2 to 2000 ng/µL.

Genetic testing

NGS panel GenoDENT

Samples of patients exhibiting clinical features characteristic of DI/DD were first analysed on the GenoDENT panel, which has been updated over the years since its development.[17] The most recent version (V.8.0) explores 653 genes (online supplemental table S1). The GenoDENT panel searches for two gene categories, including a diagnostic panel (248 genes identified as responsible for rare diseases with orodental manifestations in humans) and a discovery panel (405 candidate genes reported in tooth development or orodental anomalies in in vitro/in vivo research models). Probe design was obtained through the Agilent SureDesign portal (https://erray.chem.agilent.com/suredesign, Agilent, USA) with the aim of capturing the exonic sequence along with 25 bases of their flanking intronic sequences by complementarity. To improve sequencing quality for the *DSPP* gene, a specific design was implemented by adding additional probes targeting its repetitive regions. The Agilent SureSelect QXT protocol was used to prepare the libraries, and sequencing was

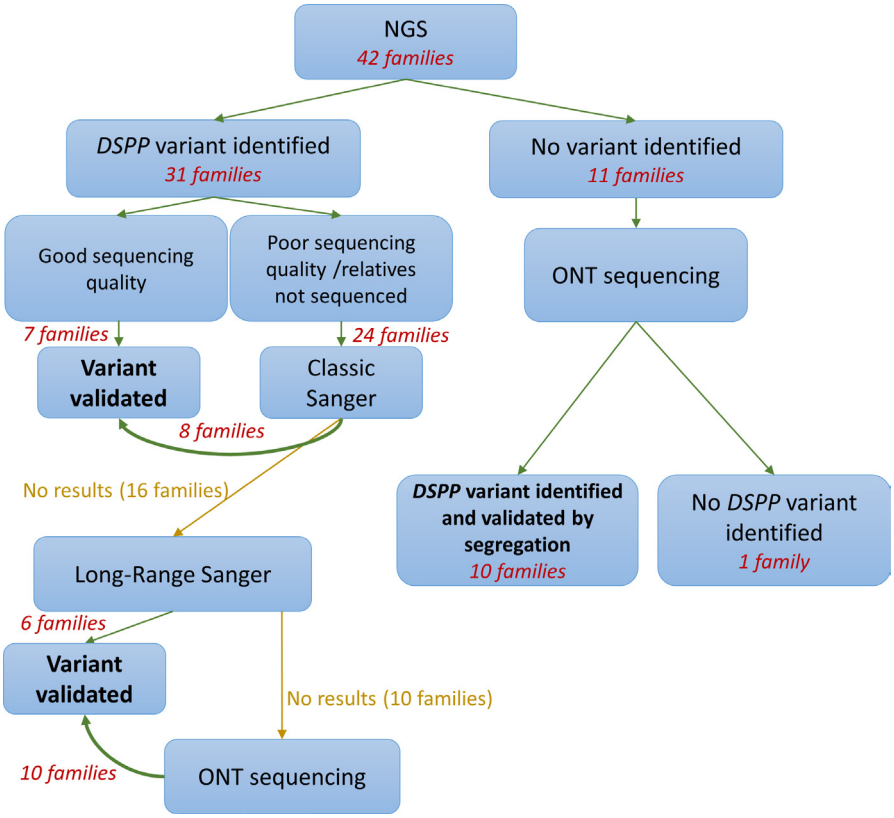

Figure 1 Diagnostic workflow and validation strategy for *DSPP* variant detection. Forty-two families were analysed using an NGS panel. *DSPP* variants were initially identified in 31 families, while no candidate variant was detected in 11 families. When sequencing quality and segregation data were sufficient, variants were directly validated (seven families). In cases of insufficient sequencing quality or unavailable relatives (24 families), classic Sanger sequencing was performed, leading to variant validation in 8 families. For unresolved cases, long-range Sanger sequencing was applied, validating *DSPP* variants in an additional six families. Remaining negative cases underwent ONT sequencing, resulting in validation in 10 families. Among families without variants identified by the initial NGS panel, ONT sequencing identified and validated *DSPP* variants in 10 families, while no *DSPP* variant was found in 1 family. DSPP, dentin sialophosphoprotein; NGS, next-generation sequencing; ONT, Oxford Nanopore Technology.

done on a NextSeq 550 (Illumina, San Diego, USA). GenoDENT has been enforced in a diagnostic setting since 2018 allowing the availability of the results for the corresponding doctor/dentist and genetic counselling.

For families 11 and 14, the probands had been sequenced by NGS several years earlier, and individual parental DNA samples were no longer available in sufficient quantity for separate analysis. Parental DNA samples from the same sequencing run were pooled by parent (mothers and fathers separately) and sequenced as single samples to assess segregation of the variant identified in the proband. Inconclusive results correspond to patients sequenced by NGS who either had a variant identified with poor sequencing quality (<10×, only one allele represented and/or no variants called) or for whom no pathogenic mutation was identified (negative results). In both cases, additional experiments (Sanger sequencing, ONT or both) were required.

NGS exome sequencing

For patients with DI of family 5 (online supplemental figure S1, online supplemental table S4) who tested negative with the GenoDENT panel, exome sequencing was performed directly in the genetic diagnostic laboratory at University Hospitals of Strasbourg. Exons of DNA samples were captured using[19] in-solution enrichment methodology (Nimblegen Seqcap EZ MedExome Target Enrichment Kit, Roche) with the company's biotinylated oligonucleotide probe library (Kappa HyperPrep Library Kit) and sequenced with a NextSeq 550 (Illumina, San Diego, USA) as paired-end 75 bp reads, resulting in an average coverage of 50×.

Long-range PCR

To amplify the repetitive regions of the *DSPP* gene, a long-range PCR was performed using the TaKaRa LA Taq (#RR002M), following the manufacturer's instructions. The primers DSPP-E5-Forward: CCCCTCCATCTATTGATGCTAGCACA and DSPP-E5-Reverse: CCCATAGCCCAGGTAGGTTGAATAGT were used to amplify an amplicon of 3796 bp, which was tested (1) with several primers (online supplemental table S2) for Sanger sequencing or (2) directly sequenced with the MinION.

Sanger sequencing and segregation

The Amplifix V.1.5.4 software was used to design the primers that were eventually ordered from Eurofins MWG (online supplemental table S3). A first PCR was done to generate the amplicons, and the sizes were checked by electrophoresis on the TapeStation system (Agilent).

Following enzymatic cleanup with the Illustra Exoprostar kit (Sigma-Aldrich) to eliminate residual deoxynucleotide triphosphates (dNTPs) and salts, sequencing reactions were set up using the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems, Thermo Fisher Scientific). The resulting products were subsequently purified with the BigDye Xterminator Purification Kit (Applied Biosystems, Thermo Fisher Scientific). Purified

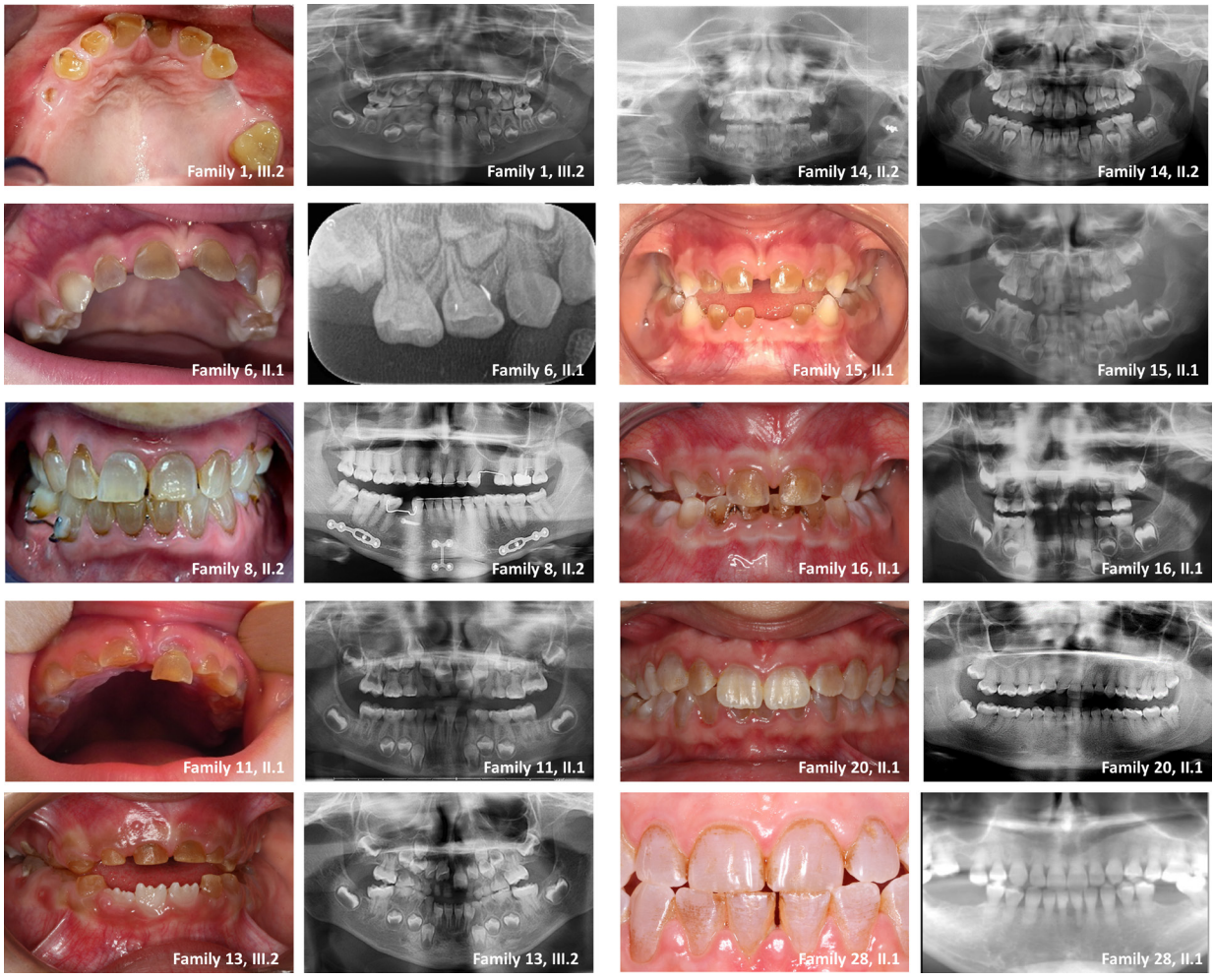

Figure 2 Dental phenotype in patients carrying dentin sialophosphoprotein variants. Both permanent and primary dentition of probands display the hallmarks of DI, with transparent amber-brown discolouration of the dentin, enamel loss and excessive tooth attrition. Clinical variability is observed among families. Panoramic radiographs show obliterated pulp and root canals, with absence of the pulp chamber. The roots appear short and narrow, and a bulbous-shaped crown with a marked cervical constriction is observed. Families 6, 14 and 15 exhibit clinical features of dentin dysplasia, presenting a similar phenotype to DI in primary teeth, while the permanent dentition appears nearly normal but shows deformity of the pulp chambers and pulp stones on radiographic examination. Families are displayed in ascending order according to the location of the pathogenic variant (see online supplemental table S4). DI, dentinogenesis imperfecta.

samples were then analysed on an ABI 3500/3500 XL capillary sequencer (Applied Biosystems, Thermo Fisher Scientific) and sequence data were processed using the SEQUENCE Pilot software (JSI Medical Systems).

Oxford nanopore long-read sequencing

The sequencing libraries were prepared with the Ligation Sequencing Kit V9 (#SQK-LSK109, Oxford Nanopore Technologies) and multiplexed with the Native Barcoding Expansion 1–12 kit V4 (#EXP-NBD104, Oxford Nanopore Technologies) following the manufacturer's instructions. Prepared libraries were loaded on a MinION flow cell (#FLO-MIN106, Oxford Nanopore Technologies) and sequenced with the MinION MK1B device (Oxford Nanopore Technologies). ONT-only variants were validated through segregation analysis.

Variant classification

The American College of Medical Genetics (ACMG) classification[21 22] was followed to classify genetic variants. Class 1 and 2 correspond to benign and likely benign polymorphisms, respectively. Class 3 or variants of uncertain significance are not recommended for clinical decision-making. Variants of class 4 (probably pathogenic) or 5 (pathogenic) identified by NGS were checked by Sanger sequencing. Long-range sequencing was performed when sequencing artefacts or poor-quality reads were encountered due to the sequencing difficulties of some areas of the *DSPP* gene. A detailed report describing the genetic variants found in the patient was written and sent to the geneticist.

Bioinformatics analysis

Short-read sequencing (NGS) data analysis

The STARK (Stellar Tools from raw sequencing data Analysis to variant RanKing) bioinformatics pipeline was used for NGS data analysis. Analysis was performed as previously described in Bloch-Zupan *et al*.[19] Sequences were aligned to the Genome Reference Consortium Human Build 37 (GRCh37). A sequencing depth of ≥30× was achieved for 95% of the coding regions.

Long-read sequencing (ONT) data analysis

We implemented an in-house analysis pipeline using ONT recommended tools. Basecalling was performed with Dorado,

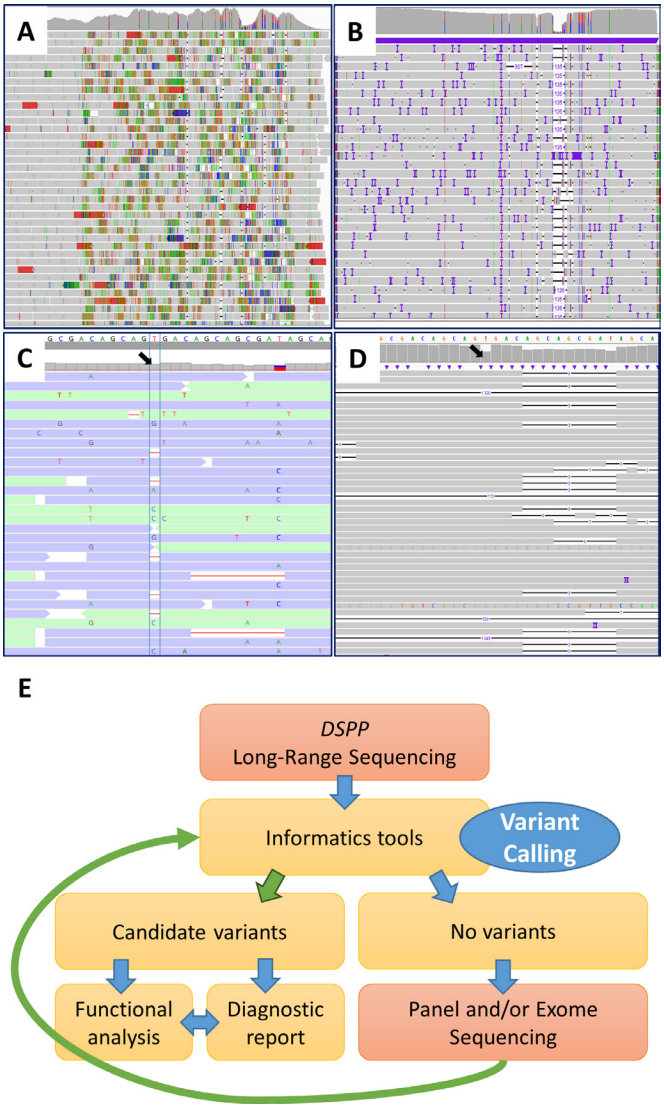

Figure 3 Sequencing results and revised diagnostic strategy for patients with dentinogenesis imperfecta or dentin dysplasia. (A) Next-generation short-read sequencing shows lower accuracy and higher error rates compared with (B) ONT long-read sequencing. (C) Short-read sequencing poorly detected a nucleotide deletion compared with (D) ONT long-read sequencing. The black arrow indicates the c.3480del variant. (E) Updated diagnostic workflow: initial analysis begins with long-range PCR of *DSPP* exon 5 followed by ONT sequencing, alignment and variant calling. If variants are detected, variants of uncertain significance undergo functional analysis, while pathogenic variants directly support diagnostic reporting. If no variants are identified, short-read gene panel or exome sequencing is subsequently performed. DSPP, dentin sialophosphoprotein; ONT, Oxford Nanopore Technology.

ONT's latest basecaller, and reads with a mean quality score below 8 were discarded. Filtered reads were aligned to the reference genome using minimap2 with default parameters. Single-nucleotide variants and insertions/deletions were identified using DeepVariant. Sequences have been aligned to either Genome Reference Consortium Human Build 37 (GRCh37) or the Telomere-to-Telomere assembly of the CHM13 cell line (T2T-CHM13v2.0) and analysed using the Integrative Genomics Viewer V.2.12.2 programme. Only samples with a minimum coverage of 20× were analysed.

Control population databases

All the variants described in the publication were checked within the following databases: Genome Aggregation Database (gnomAD),[23] 1000 Genomes Project[24] and Exome Aggregation Consortium.[25]

RESULTS

For patients in whom we identified a pathogenic variant in *DSPP* using our standard validation protocol that involves next-generation short-read sequencing,[17 18] we managed to confirm the variant using two strategies: either by performing Sanger sequencing or by amplifying the whole exon 5 of *DSPP* with a long-range PCR, followed by Sanger sequencing. However, in certain cases, the conventional approach detected variants in *DSPP* within repetitive sequences, but without sufficient quality to pass genetic diagnosis (online supplemental table S4), such as low coverage and high rates of sequencing mismatches in these regions (figure 3A–D). In this context, we chose the ONT approach for these patients, as well as those who were tested negative by NGS.

Clinical descriptions of dentin defects by family

DI type II manifests as discoloured amber-brown translucent teeth. Both primary and permanent teeth can be involved. Rapid enamel wear and attrition are typically found, exposing defective dentin underneath (families 1–5, 8–13, 16–41). Frequent infections of the pulp chamber may occur. In radiographic examinations, a marked cervical constriction and globular or bulbous crown are usually observed. The roots are short and conical. Pulp obliteration is often present, with both the pulp chamber and root canals affected (figure 2, online supplemental figure S1).

DD affects both primary and permanent teeth. Primary teeth show a similar phenotype to DI (families 6, 7, 14 and 15). Permanent teeth have a normal clinical appearance. X-rays show bulbous crowns, cervical constrictions and wide pulp chambers. Other signs, such as flame-shaped pulp chambers and calcifications, can also be part of the tooth anomalies (family 14; see figure 2).

DSPP variants identified by NGS

Using our routine strategy, we identified 14 different *DSPP* variants: 1 in exon 3 and 13 in exon 5 (online supplemental table S4). Among them, the heterozygous pathogenic variant c.52G>T; p.(Val18Phe), previously described by Xiao *et al*,[26] was found in families 1, 2 and 3 with isolated DI (figure 2, online supplemental figure S1). Familial segregation confirmed maternal transmission in families 1 and 3 (online supplemental table S4 and figure S1).

Novel variants found using ONT

Variants in repetitive regions required long-range PCR for detection.

Eight families (21–28) initially showed a non-pathogenic insertion (c.2656_2657insGTGAAGCA) using NGS. The use of ONT coupled with the most recent reference genome (T2T), which has been shown to better solve repetitive repeats,[27 28] revealed a pathogenic variant (c.2652del; p.Asp884Glufs*430) instead. Multigenerational inheritance was confirmed in most cases, although parental samples were unavailable for some.

Moreover, ONT sequencing also revealed additional novel pathogenic variants, c.1887del, c.1937del, c.2483del, c.2827del

and c.3180del, in families 13, 15, 17, 29 and 32, respectively (online supplemental table S4).

Several frameshift variants were detected and classified as probably pathogenic or pathogenic based on their absence from population databases and clinical features consistent with DI or DD.

Familial segregation supported dominant inheritance and affected individuals commonly exhibited enamel loss, amber to brown discolouration, obliterated pulp chambers, short or conical roots, bulbous crowns and marked cervical constrictions (figure 2, online supplemental figure S1).

All *DSPP*-associated entities represent a single genetic disease spectrum, in which clinical phenotypes reflect varying degrees of severity rather than distinct disorders.

DISCUSSION

The present study provides significant insights into the genetic landscape of DI and DD-II by investigating *DSPP* gene variants in a cohort of 112 individuals. Our findings confirm that *DSPP* variants harbour a high degree of phenotypic variability, particularly within exon 5, which is known for its repetitive nature. Our previous routine strategy, based on next-generation short-read sequencing, enabled the identification of *DSPP* variants in 24 individuals, including probands and their relatives. However, in several cases, sequencing artefacts and low-quality reads led to inconclusive results. To overcome these limitations, we revised our diagnostic workflow to begin directly with long-read *DSPP* exon 5 sequencing using ONT (figure 3E). This approach has successfully resolved previously ambiguous cases, allowed for accurate characterisation of *DSPP* variants and led to the discovery of novel pathogenic mutations. These findings highlight the value of integrating long-read sequencing into routine diagnostics for hereditary dentin disorders. Notably, *DSPP* variants were identified in all index cases (41) except one presenting DI or DD clinical features, suggesting that the penetrance of *DSPP*-related dentin disorders is complete.

One of the major challenges in *DSPP* sequencing is the presence of extensive tandem repeats within exon 5, which complicates traditional sequencing methods. Innovative strategies and bioinformatics tools are continually being developed to enhance the precision and reliability of sequencing efforts in these intricate genomic regions. Short-read NGS variant calling and/or conventional Sanger sequencing failed to detect several pathogenic variants due to the repetitive nature of these sequences. Long-range PCR followed by Sanger sequencing provided partial resolution, but ONT-based sequencing emerged as the most reliable approach for characterising these regions. Our results highlight the critical role of third-generation sequencing in overcoming these diagnostic hurdles and establishing a more comprehensive variant spectrum for *DSPP*.

ONT sequencing started to be widely used in the 2010s. It is a cutting-edge technology in the field of DNA sequencing that offers a unique and promising approach. Unlike traditional sequencing methods that rely on the detection of fluorescent signals, nanopore sequencing involves passing a single DNA strand through a nanoscale pore. As the DNA molecule traverses the pore, it causes disruptions in the electrical current, allowing for real-time analysis of the sequence. This method is advantageous for several reasons, including its ability to sequence long stretches of DNA, detect various types of modifications and provide rapid results.[29 30] This technology has already proved its efficiency: it was used to pinpoint variants in giant protein titin (TTN) repeated domains and improved the diagnosis of patients with myopathies,[31] allowed the identification of variants for patients who were in diagnostic wandering[32] and enabled the characterisation of structural deletions that were unsolved with exome sequencing.[33] In our study, the use of ONT allowed the identification of *DSPP* variants classified as likely pathogenic or pathogenic in 20 families. Of these, variants in half of these families had already been detected using the targeted NGS panel, while the remaining half represented novel discoveries made possible by long-read sequencing. Although long-read sequencing significantly reduced diagnostic wavering by improving the detection of variants in the repetitive regions of *DSPP* (figure 3), these unresolved cases suggest that other mechanisms—such as deep intronic or regulatory variants, or RNA-level alterations—may contribute to the phenotype and remain undetected with our current approach. Additionally, we cannot exclude that variants in other genes, not covered by our panel, may be responsible for the clinical features observed in these individuals.

The identified *DSPP* variants predominantly consisted of 3′-*DSPP* −1 frameshift mutations. In a mouse model, these variations have been demonstrated to impact odontoblasts through polarity loss, autophagy and junction abnormalities, which result in delayed dentin deposition, ectopic mineralisation along collagen and missing processes.[1]

Notably, the clinical variability observed among patients carrying the same pathogenic variant suggests the influence of additional genetic, epigenetic or environmental factors modulating the phenotypic expression of *DSPP* variants. The heterogeneous expressivity is evident in families where the same variant is inherited across generations, yet affected individuals exhibit differing severities of dentin anomalies. A key observation from our study was the recurrent inheritance pattern of specific *DSPP* variants. For example, the c.52G>T (p.Val18Phe) variant was identified in different families and consistently associated with isolated DI. Similarly, novel variants such as c.1729_1924del (p.Ser577Alafs672) and c.1814del (p.Ser605Metfs709) exhibited strong pathogenicity and familial segregation patterns. These findings further validate the classification of these variants as disease-causing variants and emphasise the importance of genetic counselling for affected families.

Furthermore, our study underscores the need for continuous improvements in genome assembly references and draws attention to the potential influence of *DSPP* haplotype diversity on variant interpretation within the DPP repetitive region. The initial misinterpretation of the c.2656_2657insGTGAAGCA variant as disease-associated—later clarified through alignment with the updated T2T reference genome—may, at least in part, be related to differences between patient-specific *DSPP* haplotypes and the reference sequence, as previously reported.[5 16] This observation illustrates how haplotype-related indel configurations in highly repetitive regions can complicate short-read alignment and variant calling and highlights the value of improved reference genomes and long-read sequencing approaches for accurate *DSPP* variant classification.

The identification of novel pathogenic variants and the confirmation of known ones strengthen the genetic basis of DI and DD-II by leveraging advanced sequencing methodologies. It is important to perform genetic testing in individuals with dentin defects. Aside from isolated DI and DD-II, this dental phenotype might be associated with a mild form of osteogenesis imperfecta or an undiagnosed skeletal disorder that needs specific care and support/management of the patient, for instance, related to *COL1A1* and *COL1A2* pathogenic variants.[34]

In conclusion, our study enhances the understanding of *DSPP*-related dentin disorders by leveraging advanced sequencing methodologies. We identified 22 pathogenic or likely pathogenic *DSPP* variants, including 8 (36%) previously reported and 14 (64%), novel variants thereby expanding the known mutational spectrum. Our findings advocate for the integration of long-read sequencing technologies into routine diagnostic workflows, ensuring more accurate genetic diagnosis and personalised management strategies for affected individuals. Future research should focus on expanding cohort studies, investigating potential genotype–phenotype correlations and exploring therapeutic interventions targeting *DSPP*-associated dentin defects.

Author affiliations
[1]Pôle de Médecine et Chirurgie Bucco-dentaires, Hôpital Civil, Centre de Référence des Maladies Rares Orales et Dentaires, O-Rares, Filière Santé Maladies Rares TETECOU, European Reference Network ERN CRANIO, Les Hopitaux Universitaires de Strasbourg, Strasbourg, France
[2]Laboratoire de diagnostic génétique, Nouvel Hôpital Civil, Les Hôpitaux Universitaires de Strasbourg, Strasbourg, France
[3]Université de Strasbourg, INSERM U1258, CNRS-UMR7104, Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France
[4]CNRS, Université de Strasbourg, Institut de Biologie Moléculaire des Plantes, Strasbourg, France
[5]Laboratoire de Génétique Médicale, INSERM URMS_1112, Institut de Génétique Médicale d'Alsace, Centre de Recherche en Biomédecine de Strasbourg, Université de Strasbourg, Strasbourg, France
[6]Faculté de Chirurgie Dentaire, Université de Strasbourg, Strasbourg, France
[7]Institut d'études avancées (USIAS), Université de Strasbourg, Strasbourg, France

Acknowledgements We are grateful to the families and individuals for their participation and invaluable contribution to this work. We thank health professionals for their dedication and commitment to taking care of the patients (Bernard Daclin, Renaud Rinkenbach, Nuh Kalkan, Philippe Magnani, Orlando Pimenta Ridrigues, Stéphanie Etienne Weimann, François Becmeur, Tatiana Wendling-Guivante, Aymeric Courval, Daniel Jasiak, Florence Demurger, Rabisoa W Razafindrakoto, Saholy N Ralison, Jean-Claude Randrianaivo, Georgette Ralison, Rija O Andriamasinoro, Rija H Ramanampamaharana, Solofomanantsoa E Randrianazary, Louise H Ralimanana). We thank Patrick Reilly for critical reading and improvement of English language for this manuscript.

Collaborators O-Rares Consortium: French Reference and Competence Centers, healthcare network for rare oral and dental diseases CRMR/CCMR O-Rares, French Filière 'TETECOU' and associated paediatric/genetic teams. Elsa Garot (CHU de Bordeaux—Hôpital Pellegrin); James Lespinasse (Centre Hospitalier Métropole Savoie, Chambéry); Jean-Jacques Morrier, Béatrice Thivichon-Prince (Hospices Civils de Lyon, Groupement Centre, Pôle d'Activités Médicales d'Odontologie, Centre de soins Dentaires); Corinne Tardieu (APHM, Hôpital de la Timone Enfants, Marseille); Clara Joseph-Beaudin, Valentin Ruault (CHU de Montpellier); Serena Lopez (CHU de Nantes—Hôtel Dieu); Christine Voha (CHU de Nice); Benjamin Fournier (Hôpital Rothschild, APHP, Paris); Gwenaël Le Guyader (CHU de Poitiers); Marie Paule Gelle (CHU Reims, Hôpital Maison Blanche, Pôle Odontologie); Agnès Bloch-Zupan, François Clauss, Nhat Minh Do, Alexandra Jimenez-Armijo, Marzena Kawczynski, Marie-Cécile Manière, Yves Alembik, Salima El Chehadeh, Elise Schaefer (Hôpitaux Universitaires de Strasbourg); Isabelle Bailleul-Forestier, Nicolas Chassaing (CHU de Toulouse—Hôpital Rangueil); Annick Toutain (CHU de Tours).

Contributors Conceptualisation: AB-Z; data curation: GC, AJ-A, MK, O-Rares C, TR; formal analysis: GC, AJ-A, AG, CS-B, AB-Z; funding acquisition: AB-Z; investigation: GC, TR, MA, AR, J-BL, AG; methodology: AA, CS-B, AB-Z; project administration: AB-Z; resources: CS-B, AB-Z; software: J-BL, AA; supervision: AB-Z; validation: CS-B, AB-Z; visualisation: GC, AJ-A; writing—original draft: GC, AJ-A; writing—review and editing: GC, AJ-A, MK, TR, O-Rares C, MA, AR, J-BL, AG, AA, CS-B, AB-Z. AB-Z is the guarantor of this article.

Funding This work was supported by the following grants: PHRC 2008–2014 N°4266 Amelogenesis imperfecta, the University Hospitals of Strasbourg (HUS); HUS, API, 2009–2012, 'Oral cavity development: from gene to clinical phenotype in humans'; INTERREG IV Upper Rhine Offensive Sciences A27 'Oral manifestations of rare diseases: diagnostic and therapeutic perspectives'; No. 1. 7 'RARENET: A trinational network for education, research and management of complex and rare disorders in the Upper Rhine' 2016–2019 cofinanced by the European Union's European Regional Development Fund (ERDF) as part of the INTERREG V program; the e-GenoDENT project financed by the Fonds d'Intervention Régionale (FIR) of the Agence Régionale de Santé Grand Est (2019–2022); the Fondation Maladies Rares e-health co-design workshop (2019); the AMI Economie numérique Grand-Est i-Dent (2020–2021); Impulsion recherche Filière TETECOU, 2020 and 2022; Bpifrance as part of the digital health acceleration strategy of the Délégation ministérielle à la santé numérique (DNS) and the Secrétariat général pour l'investissement (SGPI), as well as support for the Health Data Hub (HDH) as part of the 'France 2030' national strategy (Data Challenge, Health Data Hub, BPI 2021–2023); Fondation Force DIAGNODENT 2023-2026; ANR 2023 'Agence Nationale de la Recherche' grant, 3DBioDENT 2023–2026 ANR-23-CE17-0048-01; CREGEMES (Centre Regional de GEnetique MEdicale de Strasbourg) grant 2023. This work is also part of the IMCBio Interdisciplinary Thematic Institute, part of the ITI 2021–2028 programme of the University of Strasbourg, CNRS and Inserm, was supported by IdEx Unistra (ANR-10-IDEX-0002), and by the SFRI-STRAT'US project (ANR 20-SFRI-0012) and the IMCBio EUR (ANR-17-EURE-0023) as part of the French Investissements d'Avenir program, and finally the MIG F04 dedicated to rare disease reference centers (CRMR), DGOS, French Ministry of Health and Prevention.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval The patients presenting with DI or DD and the non-affected family members gave written informed consents in accordance with the Declaration of Helsinki, both for the D[4]/phenodent registry and for genetic analyses performed on salivary samples. This clinical study is registered at https://clinicaltrials.gov: NCT02397824, and in the MESR (French Ministry of Higher Education and Research) Bioethics Commission as a biological collection 'Orodental Manifestations of Rare Diseases' DC-2012-1,677 within DC-2012-1,002 and was acknowledged by the CPP (person protection committee) Est IV 11 December 2012. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information. The variants were submitted in ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), a freely available, public archive of human genetic variants and interpretations of their significance to disease, maintained at the National Institutes of Health. Their accession numbers are: SCV006323536, SCV006323537, SCV006323538, SCV006323539, SCV006323540, SCV006323541, SCV006323543, SCV006323544, SCV006323545, SCV006323546, SCV006323549, SCV006323550, SCV006323551, SCV006323553, SCV006323554, SCV006323555, SCV006323556, SCV006323557, SCV006323558, SCV006323559, SCV006323560 and SCV006550886.

ORCID iDs
Alexandra Jiménez-Armijo https://orcid.org/0000-0002-3967-3996
Caroline Schluth-Bolard https://orcid.org/0000-0002-2279-4273

REFERENCES

1 Liang T, Smith CE, Hu Y, *et al*. Dentin defects caused by a Dspp^{-1} frameshift mutation are associated with the activation of autophagy. *Sci Rep* 2023;13:6393.
2 Lim D, Wu K-C, Lee A, *et al*. DSPP dosage affects tooth development and dentin mineralization. *PLoS ONE* 2021;16:e0250429.
3 de La Dure-Molla M, Philippe Fournier B, Berdal A. Isolated dentinogenesis imperfecta and dentin dysplasia: revision of the classification. *Eur J Hum Genet* 2015;23:445–51.
4 Li F, Liu Y, Liu H, *et al*. Phenotype and genotype analyses in seven families with dentinogenesis imperfecta or dentin dysplasia. *Oral Dis* 2017;23:360–6.
5 Simmer JP, Zhang H, Moon SJH, *et al*. The Modified Shields Classification and 12 Families with Defined DSPP Mutations. *Genes (Basel)* 2022;13:858.
6 Yuan M, Zheng X, Xue Y, *et al*. A novel DSPP frameshift mutation causing dentin dysplasia type 2 and disease management strategies. *Oral Dis* 2023;29:2394–400.
7 Qin C, Brunn JC, Cadena E, *et al*. The expression of dentin sialophosphoprotein gene in bone. *J Dent Res* 2002;81:392–4.

8 Ogbureke KUE, Fisher LW. Expression of SIBLINGs and their partner MMPs in salivary glands. *J Dent Res* 2004;83:664–70.
9 Ogbureke KUE, Fisher LW. Renal expression of SIBLING proteins and their partner matrix metalloproteinases (MMPs). *Kidney Int* 2005;68:155–66.
10 Ogbureke KUE, Fisher LW. SIBLING expression patterns in duct epithelia reflect the degree of metabolic activity. *J Histochem Cytochem* 2007;55:403–9.
11 Fisher LW, Fedarko NS. Six genes expressed in bones and teeth encode the current members of the SIBLING family of proteins. *Connect Tissue Res* 2003;44 Suppl 1:33–40.
12 McKnight DA, Suzanne Hart P, Hart TC, *et al*. A comprehensive analysis of normal variation and disease-causing mutations in the human DSPP gene. *Hum Mutat* 2008;29:1392–404.
13 Prasad M, Butler WT, Qin C. Dentin sialophosphoprotein in biomineralization. *Connect Tissue Res* 2010;51:404–17.
14 Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. *Genomics* 2016;107:1–8.
15 Mardis ER. DNA sequencing technologies: 2006-2016. *Nat Protoc* 2017;12:213–8.
16 Yang J, Kawasaki K, Lee M, *et al*. The dentin phosphoprotein repeat region and inherited defects of dentin. *Mol Genet Genomic Med* 2016;4:28–38.
17 Prasad MK, Geoffroy V, Vicaire S, *et al*. A targeted next-generation sequencing assay for the molecular diagnosis of genetic disorders with orodental involvement. *J Med Genet* 2016;53:98–110.
18 Rey T, Tarabeux J, Gerard B, *et al*. Protocol GenoDENT: Implementation of a New NGS Panel for Molecular Diagnosis of Genetic Disorders with Orodental Involvement. *Methods Mol Biol* 2019;1922:407–52.
19 Bloch-Zupan A, Rey T, Jimenez-Armijo A, *et al*. *Amelogenesis imperfecta*: Next-generation sequencing sheds light on Witkop's classification. *Front Physiol* 2023;14:1130175.
20 de La Dure-Molla M, Fournier BP, Manzanares MC, *et al*. Elements of morphology: Standard terminology for the teeth and classifying genetic dental disorders. *Am J Med Genet A* 2019;179:1913–81.
21 Richards S, Aziz N, Bale S, *et al*. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015;17:405–24.
22 Harrison SM, Biesecker LG, Rehm HL. Overview of Specifications to the ACMG/AMP Variant Interpretation Guidelines. *Curr Protoc Hum Genet* 2019;103:e93.
23 Chen S, Francioli LC, Goodrich JK, *et al*. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature New Biol* 2024;625:92–100.
24 Siva N. 1000 Genomes project. *Nat Biotechnol* 2008;26:256.
25 Lek M, Karczewski KJ, Minikel EV, *et al*. Analysis of protein-coding genetic variation in 60,706 humans. *Nature New Biol* 2016;536:285–91.
26 Xiao S, Yu C, Chou X, *et al*. Dentinogenesis imperfecta 1 with or without progressive hearing loss is associated with distinct mutations in DSPP. *Nat Genet* 2001;27:201–4.
27 Aganezov S, Yan SM, Soto DC, *et al*. A complete reference genome improves analysis of human genetic variation. *Science* 2022;376:eabl3533.
28 Nurk S, Koren S, Rhie A, *et al*. The complete sequence of a human genome. *Science* 2022;376:44–53.
29 Oxford Nanopore Technologies. Company history. 2021. Available: https://nanoporetech.com/about-us/history [Accessed 12 Feb 2024].
30 MacKenzie M, Argyropoulos C. An Introduction to Nanopore Sequencing: Past, Present, and Future Considerations. *Micromachines (Basel)* 2023;14:459.
31 Bonnet C, Pellerin D, Roth V, *et al*. Optimized testing strategy for the diagnosis of GAA-FGF14 ataxia/spinocerebellar ataxia 27B. *Sci Rep* 2023;13:9737.
32 Miller DE, Sulovari A, Wang T, *et al*. Targeted long-read sequencing identifies missing disease-causing variation. *Am J Hum Genet* 2021;108:1436–49.
33 McClinton B, Crinnion LA, McKibbin M, *et al*. Targeted nanopore sequencing enables complete characterisation of structural deletions initially identified using exon-based short-read sequencing strategies. *Mol Genet Genomic Med* 2023;11:e2164.
34 Yamaguti PM, de La Dure-Molla M, Monnot S, *et al*. Unequal Impact of COL1A1 and COL1A2 Variants on Dentinogenesis Imperfecta. *J Dent Res Juin* 2023;102:616–25.