HGG Advances

ARTICLE

Clinical, genetic, epidemiologic, evolutionary, and functional delineation of *TSPEAR*-related autosomal recessive ectodermal dysplasia 14

Adam Jackson,[1,2] Sheng-Jia Lin,[3] Elizabeth A. Jones,[1,2] Kate E. Chandler,[2] David Orr,[2] Celia Moss,[4] Zahra Haider,[4] Gavin Ryan,[5] Simon Holden,[6] Mike Harrison,[7] Nigel Burrows,[8] Wendy D. Jones,[9] Mary Loveless,[3] Cassidy Petree,[3] Helen Stewart,[10] Karen Low,[11] Deirdre Donnelly,[12] Simon Lovell,[1] Konstantina Drosou,[13,14] The Genomics England Research Consortium, Solve-RD consortium, Gaurav K. Varshney,[3,15] and Siddharth Banka[1,2,15,16,*]

Summary

TSPEAR variants cause autosomal recessive ectodermal dysplasia (ARED) 14. The function of TSPEAR is unknown. The clinical features, the mutation spectrum, and the underlying mechanisms of ARED14 are poorly understood. Combining data from new and previously published individuals established that ARED14 is primarily characterized by dental anomalies such as conical tooth cusps and hypodontia, like those seen in individuals with *WNT10A*-related odontoonychodermal dysplasia. AlphaFold-predicted structure-based analysis showed that most of the pathogenic *TSPEAR* missense variants likely destabilize the β-propeller of the protein. Analysis of 100000 Genomes Project (100KGP) data revealed multiple founder *TSPEAR* variants across different populations. Mutational and recombination clock analyses demonstrated that non-Finnish European founder variants likely originated around the end of the last ice age, a period of major climatic transition. Analysis of gnomAD data showed that the non-Finnish European population TSPEAR gene-carrier rate is ~1/140, making it one of the commonest AREDs. Phylogenetic and AlphaFold structural analyses showed that TSPEAR is an ortholog of drosophila *Closca*, an extracellular matrix-dependent signaling regulator. We, therefore, hypothesized that TSPEAR could have a role in enamel knot, a structure that coordinates patterning of developing tooth cusps. Analysis of mouse single-cell RNA sequencing (scRNA-seq) data revealed highly restricted expression of Tspear in clusters representing enamel knots. A *tspeara*$^{-/-}$;*tspearb*$^{-/-}$ double-knockout zebrafish model recapitulated the clinical features of ARED14 and fin regeneration abnormalities of wnt10a knockout fish, thus suggesting interaction between tspear and wnt10a. In summary, we provide insights into the role of TSPEAR in ectodermal development and the evolutionary history, epidemiology, mechanisms, and consequences of its loss of function variants.

Introduction

Cell fate specifications in the three primordial germ layers dictate the development of complex tissues and organs. The ectodermal layer gives rise to structures on the outer surface of the organism including hair, feathers, teeth, skin, and sweat glands. Evolutionary adaptation of ectodermal tissues via the emergence of new features is of central importance in the survival and success of species. For example, humans have a uniquely high density of sweat glands embedded within skin, and our capacity to sweat was critical to the survival and migration of our species.[1] In addition, during the course of human evolution, the size of the jaw, as well as the size and number of teeth, has reduced due to decreased functional requirement[2]—the most cited example being the third molar, which is notable absent in nearly 40% of humans.[3] Even though ectodermal-derived tissues display a high degree of variability across species, the underpinning molecular mechanisms that orchestrate ectodermal development are highly conserved throughout eumetazoans.

Ectodermal dysplasias (EDs) are multi-system disorders with substantial morbidity and healthcare burden.[4,5] Their accurate molecular diagnosis is important for appropriate clinical management, treatment, complication prevention, and risk estimation. They are genetically highly heterogeneous with at least 150 known ED genes.[6] Out of

[1]Division of Evolution, Infection and Genomics, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK; [2]Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation Trust, Health Innovation Manchester, Manchester, UK; [3]Genes and Human Disease Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA; [4]Department of Dermatology, Birmingham Children's Hospital, Birmingham Women's and Children's NHS Foundation Trust, Birmingham, UK; [5]West Midlands Regional Genetics Laboratory, Birmingham Women's and Children's NHS Foundation Trust, Birmingham, UK; [6]Clinical Genetics, Addenbrooke's Hospital, Cambridge, UK; [7]Department of Pediatric Dentistry, Guy's and St Thomas' Dental Institute, London, UK; [8]Department of Dermatology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK; [9]North East Thames Regional Genetics Service, Great Ormond Street Hospital for Children, Great Ormond Street NHS Foundation Trust, London, UK; [10]Oxford Centre for Genomic Medicine, Oxford University Hospitals NHS Foundation Trust, Oxford, UK; [11]Department of Clinical Genetics, St Michael's Hospital, Bristol, UK; [12]Department of Genetic Medicine, Belfast HSC Trust, Lisburn Road, Belfast, UK; [13]Department of Earth and Environmental Sciences, Manchester Institute of Biotechnology, University of Manchester, Manchester, UK; [14]Division of Cell Matrix Biology and Regenerative Medicine, Faculty of Biology, Medicine and Health, University of Manchester, 99 Oxford Road, Manchester, UK
[15]These authors contributed equally
[16]Lead contact
*Correspondence: siddharth.banka@manchester.ac.uk
https://doi.org/10.1016/j.xhgg.2023.100186.

these, variants in approximately 40 genes result in autosomal recessive EDs (AREDs).[7] Variants in *EDAR* (ED 10B, OMIM: 224900), *EDARADD* (ED 11B, OMIM: 614941), and *WNT10A* (odontoonychodermal dysplasia, or Schopf-Schulz-Passarge syndrome, OMIM: 257980 and 224750, respectively) are thought to be the commonest causes for AREDs. Notably, monoallelic and biallelic variants in *WNT10A* are also responsible for non-syndromic tooth agenesis (OMIM: 150400).[8–11] Biallelic *TSPEAR* variants were recently identified as a cause of ARED14 (OMIM: 618180).[12,13] *TSPEAR* encodes a protein of structure and function, which is to date uncharacterized, and the variant spectrum, clinical features, and disease mechanism for ARED14 are poorly understood.

Here, we describe 11 individuals with biallelic deleterious *TSPEAR* variants and establish the major clinical features and the variant spectrum of ARED14. We describe several founder non-Finnish European *TSPEAR* variants, and by investigating the gene-carrier rates across different ancestries, we show that ARED14 is likely to be one of the commonest AREDs. Using AlphaFold structural predictions, we propose a possible mechanism for deleteriousness of missense changes in this gene. Using phylogenetic analysis and AlphaFold structural analyses, we predict a role for TSPEAR in extracellular matrix (ECM)-dependent signaling in ectodermal development. Phylogenetic analysis showed drosophila *Closca* to be a likely ortholog of TSPEAR. Based on known functions of *Closca*, we hypothesized TSPEAR to participate in primary enamel knot-dependent signaling. Using mouse single-cell RNA sequencing (scRNA-seq) data, we demonstrated highly restricted expression of Tspear in enamel knot of developing tooth, thus validating our hypothesis. Based on the clinical similarities between ARED14 and odontoonychodermal dysplasia, we hypothesized an interaction between TSPEAR and WNT10A. Our hypothesis is supported by a zebrafish CRISPR-Cas9-mediated double knockout of *TSPEAR* paralogs, *tspeara/tspearb*, which recapitulates the defective dental phenotype in human and demonstrates fin regeneration abnormalities similar to that seen in *wnt10a* knockout fish.

Materials and methods

Clinical and human genetic studies

A previously described panel-agnostic filtering pipeline[14–17] was applied on trio whole-genome sequencing data from an individual with ED in whom initial panel-led analysis had not yielded causative pathogenic variants. Additional individuals with biallelic pathogenic or likely pathogenic *TSPEAR* variants were identified from the 100,000 Genomes Project (100KGP),[18] DECIPHER,[19] and Solve-RD[20] databases and through personal collaborations. Variants were classified using the American College of Medical Genetics (ACMG) 2015 criteria.[21] Informed consents were obtained, and clinical information was collected using a standardized clinical proforma.

We searched the 100KGP dataset using three strategies.

1. A direct search of the LabKey tiering data. In this study, we used the September 3rd, 2020, release of the 100KGP dataset, and the tiered variants were accessed and filtered via the LabKey application. An initial population allele frequency filtering threshold of 1% was employed where an autosomal recessive mode of inheritance was used. The tiering data were filtered for biallelic *TSPEAR* variants only.
2. We interrogated the aggregated gVCF containing 59,464 samples aligned to GRCh38 specifically for our variants of interest (i.e., recurrent variants). Both heterozygous and homozygous individuals were included.
3. A custom script was used to iterate over the VCF files of all individuals in the rare disease arm (n = 74,008) of the 100KGP, extracting all rare (<0.01 minor-allele frequency [MAF] in the 1000 Genomes Project) protein-coding variants in *TSPEAR*. Both heterozygous and homozygous individuals were included.

We searched the DECIPHER database for any biallelic cases within the research track. For Solve-RD, we searched the Genotype-Phenotype Analysis Platform (GPAP) for biallelic rare (gnomAD, 1,000 Genomes, and internal allele frequencies <0.01) and predicted deleterious variants in *TSPEAR*.

Molecular modeling

Predicted TSPEAR protein structure was downloaded in PDB format from the AlphaFold protein structure database[22] (P59090) and inspected using the 3D Viewer on the AlphaFold webserver (www.alphafold.ebi.ac.uk). Patient variants were mapped to the predicted structure using PyMol. Charge analysis of the protein surface was undertaken using the PyMol APBS Electrostatic plugin, while conservation figures were created using conservation scores calculated by ConSurf.[23] Homology models were subsequently generated using Modeller 9.24.[24] Twenty models were built in each case, and the model with the lowest discrete optimized protein energy score was chosen for visualization and analysis in PyMol.

Haplotype analysis

Pre-computed regions of homozygosity (ROHs) (ROHCaller, Illumina) were accessed in the Genomics England Research Environment (GERE). ROHs were extracted from affected individuals' BED files and exported from the GERE through the Airlock system (Research Registry Project #461). The core haplotype was defined by the ROHs shared between affected individuals carrying a specific variant. The core haplotype SNVs were extracted from VCF files of all heterozygous carriers for that same variant in the 100KGP.

Phasing homozygotes with parental samples

Using the ROHs, we extracted all homozygous SNVs occurring in these regions that were heterozygous in both parents and hence represent haplotype-informative SNVs. We then looked at linkage disequilibrium (r^2) between these SNVs in the larger group of heterozygous carriers within the 100KGP for these variants. We extracted genotypes from the gVCFs of these individuals for these haplotype-informative SNVs and plotted r^2 using LDHeatmap.[25] We next looked at recombination events occurring within this core haplotype by selecting carrier individuals who did not carry all core SNVs and reconstructed all haplotypes present in this cohort defined by continuous allele sharing of these core SNVs.

Phase-free inference of haplotypes in heterozygous carriers

For p.Arg197Ter (n = 66) and p.V576LX*38 (n = 5) there were no homozygotes on the 100KGP available for phasing; hence, we jointly variant called, using Platypus, for a 4 Mb region flanking the variant (2 Mb on either side). We also utilized this method for p.S585I, where using core SNVs derived from the single homozygotes was non-informative, as all of these were carried by all heterozygotes. We then extracted genotypes for high-confidence variants and, using a custom script, defined the haplotype as the minimal region of continuous allele sharing with any other carrier. This method is adapted from Mathieson et al. [26] and scans left and right of the index variant and identifies inconsistent homozygous genotypes between carrier individuals—once an inconsistent homozygous genotype (i.e., one individual is a homozygous alternate while another is a homozygous reference) is identified, we know that allele sharing has stopped, and hence any intervening shared haplotype has ended (Figure S1). This method likely over-estimates the length of the haplotype, as the end of allele sharing likely occurs before inconsistent homozygous genotypes are reached. An over-estimate of haplotype length, however, is likely to result in a reduction in the estimate of the most recent common ancestor (MRCA), and hence subsequent calculations for the MRCA likely reflect lower bounds of the true age of the allele.

MRCA

The MRCA for variant carriers was estimated using discordant SNVs within the shared ROH, if applicable. The results were compared with two published methods for estimating the MRCA from genetic data: the gamma method proposed by Gondolfo et al.[27] (https://shiny.wehi.edu.au/rafehi.h/mutation-dating/) and the Genealogical Estimation of Variant Age (GEVA) database[28] (available at https://human.genome.dating/). These estimates were corroborated by variant calling from BAM files for 1,670 published ancient human genomes ranging from 45,000 BCE to 300 CE and five published archaic hominin genomes ranging from 50,000 to 120,000 ya.

Estimation of MRCA using discordant SNVs

We extracted all high-quality (read depth [DP] > 10 and genotype quality [GQ] > 15, GRCh38) discordant SNVs within the ROH (i.e., SNVs that were not homozygous in all probands) using a jointly called VCF file. We performed joint variant calling from the BAM files of homozygous individuals with parental samples available using Platypus.[29] We then considered only discordant SNVs that were inherited (i.e., present in one of the relevant parents). Discordant SNVs that were not present in a parent were taken to be either *de novo* or sequencing error and hence non-informative (Figure S2). We then filtered for discordant SNVs with a maximum allele frequency in gnomADv.2.1.1 less than that of the variant of interest, which would mean they are unlikely recurrent events. We ensured that these were indeed derived mutations by extracting genotypes for all carriers of the remaining filtered discordant SNVs, in the aggregated gVCF, as well as the *TSPEAR* variant of interest (only SNVs that co-occur with the *TSPEAR* variant of interest can be derived SNVs, assuming no recombination events). We used the number of derived mutations within this ROH to calculate the MRCA using the human germline *de novo* mutation rate reviewed by Acuna-Hidalgo et al. ($1.0–1.8 \times 10^{-8}$ per nucleotide per generation).[30] The equation below shows how MRCA is calculated.

$$\mathrm{MRCA} = \frac{M}{L_{Seg} * R_{DN}},$$

where M is the number of derived mutations, L_{Seg} is the length of the ancestral segment, and R_{DN} is the *de novo* mutation rate for the human genome.

Generation time

While both the Gondolfo et al. webserver and GEVA used generation times of 20 or 25 years as default, recent literature suggests that 25 years may be a lower bound for generation age over the last 250,000 years of human history. Moorjani et al. used the recombination clock specifically looking at Neanderthal introgressed variants in European genomes to determine that the generation interval over the last 45,000 years was likely between 26 and 30 years.[31] Most recently, Wang et al. suggested a generation time of 26.2 years with males consistently older (30.7) and females consistently younger (23.2).[32] Given these data, we chose 30 years as the human generation time.

Variant calling in ancient samples

We downloaded the BAM files for 13 studies[33–45] describing whole-genome sequencing in a total of 1,670 distinct ancient human samples using the European Nucleotide Archive, transecting a time period of the last 45,000 years for modern humans. This cohort included 10 high-coverage (>10×) ancient human genomes, as well as the medium- to high-coverage genomes for the Altai Neanderthal (120,000 years ago [ya], >10×),[46] the Chagyrskaya Neanderthal (>10×),[47] Vindija 33.19 Neanderthal (60,000 ya, >10×),[48] and the Denisovan specimens Denisova 3 (>10×)[49] and Denisova 11 (an F1 Neanderthal/Denisova hybrid, 90,000 ya, 2–3×).[50]

We only included studies where alignment data were available (studies providing only raw FASTQ files were excluded). Sequencing and capture methods are described in the individual studies, but, in brief, the majority of low-coverage samples had been enriched for 1.2 million (1,240,000) human SNVs to minimize microbial contamination while also having full or partial treatment with UDG, which aims to reduce artifact from cytosine deamination causing apparent C→T transitions in ancient DNA. Due to this capture method, we were reliant on off-target coverage of exonic regions in our genes of interest. All studies followed a standard ancient DNA pipeline including adapter removal, duplicate removal, and mapDamage to confirm their authenticity. All available BAM files had been aligned to reference genome GRCh37; hence, this build was used for all further downstream analysis for ancient samples.

BAM files were indexed using samtools, and variants were called within the coding regions of *TSPEAR*, *EDAR*, *EDARADD*, and *WNT10A* using bcftools mpileup with default parameters creating individual BCF files for each sample. Variants were then merged into a single text file and annotated using ensembl VEP (GRCh37). The normal transition-transversion (Ts:Tv) ratio in humans is 2.1:1 for genomes and 3:1 for exomes.[51] We calculated the Ts:Tv ratio, using distinct variants, for this pooled dataset at increasing DP thresholds until this ratio reached between 2:1 and 3:1. We then used this DP to filter all variants. This strategy is likely to yield high-confidence variants; however, is also likely to filter out true variants at low-coverage sites.

For *TSPEAR*, in particular, we searched the entire pooled dataset for any reads supportive of present-day pathogenic variants seen in our cohort. Given this search strategy is relaxed, it is possible that some of these variants could be artifact.

Calculation of gene-carrier rate and comparison to other forms of ARED

All high-quality (PASS) coding variants from gnomADv.2.1.1 (accessed June 24th, 2021) were downloaded for *EDAR*, *EDARADD*, *WNT10A*, and *TSPEAR*. The export also contains ClinVar annotations. The cDNA coordinates for each variant, using the canonical transcript (ENST00000323084/NM_144991), were uploaded to Ensembl VEP for CADD and SpliceAI annotations. The resultant VEP annotations were downloaded as .csv and merged with the gnomAD allele frequency.csv file using a custom python script. Variants were then filtered in Excel based on two stringency thresholds.

1. All ClinVar pathogenic (P) and likely pathogenic (LP) variants, regardless of star rating.
2. All predicted deleterious variants, including the following:
   a. ClinVar P/LP variants.
   b. High-confidence loss-of-function variants (frameshift, nonsense, canonical splice site with no LOFTEE flags).
   c. Any variant with a CADD score >25.
   d. Any variant with a SpliceAI delta score (gain or loss) >0.8, regardless of sequence ontology annotation.

The gene-carrier rate (GCR) was calculated for each ancestry in the gnomAD dataset separately using the following formulae, which have been adapted from Guo et al.[52] First, the variant carrier rate (VCR) was calculated. "AC" stands for the total allele count. "Hom" is the total number of homozygotes for this variant. "AN" is the total allele number for this variant position.

$$\mathrm{VCR} = \frac{\mathrm{AC} - \mathrm{Hom}}{0.5 * \mathrm{AN}}$$

The GCR is then calculated using the below formula. "VCR_i" is the VCR for variant "*I*". "*v*" is the total number of variants of interest. "Π" stands for the product of (1 − VCR).

$$\mathrm{GCR} = 1 - \prod_{i=1}^{v}(1 - \mathrm{VCR}_i)$$

GCRs were then plotted for each ancestry in R using ggplot2.

We also performed variant burden analysis for the known protein domains of TSPEAR. The predicted deleterious coding variants (splice variants were excluded) were mapped, and the variant burden, *V*, was calculated as

$$V = \frac{N}{L},$$

where *N* is the number of separate coding variants and *L* is the length (in amino acids) of the protein domain. Variants were plotted using rTrackViewer.[53]

Phylogenetic analysis

TSPEAR cDNA sequences for 46 diverse species were retrieved using the BLASTN function in Ensembl. Sequences were aligned and a pairwise sequence identity matrix was produced using ClustalO,[54] which was used for input into the R package pHeatmaps.

Protein BLAST searches for Uniprot amino acid sequence for TSPEAR (Q8WU66) and ENAM (Q9NRM1) were performed, and pairwise sequence identity heatmaps were generated for species that were represented in both BLAST searches. Stipulations in Uniprot BLAST included (1) reference proteomes; (2) an E-threshold below 0.0001; (3) an auto matrix; (4) no filtering parameters; (5) gapped sequences; and (6) a limit of 500 returns. As this identifies all sequences with similarity, some species can be represented more than once if a gene has been duplicated, for instance. A further filtering step was applied to keep only the sequence with the highest similarity to the human sequence.

The *TSPEAR* gene tree was produced by downloading the TSPEAR Newick format phylogenetic data from Ensembl. The tree is generated by the Gene Orthology/Paralogy prediction method pipeline in Ensembl. This was loaded into FigTree (http://tree.bio.ed.ac.uk/software/figtree/) for visualization.

Zebrafish functional studies

Zebrafish (*Danio rerio*)-related work was carried out in an Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC)-accredited facility at the Oklahoma Medical Research Foundation (OMRF) under standard conditions using standard methods as described earlier. All experiments were performed as per protocols 21-12 approved by the Institutional Animal Care Committee (IACUC) of OMRF. All zebrafish work was carried out in wild-type strain NHGRI-1.

Zebrafish has two paralogs of human TSPEAR, *tspeara* and *tspearb*. We used CRISPR-Cas9-mediated targeted mutagenesis to target both paralogs separately as described earlier. We identified two independent alleles, an 8 bp insertion and a 1 bp deletion in the *tspeara* gene and a 4 bp deletion in the *tspearb* gene, and selected these alleles for further functional analysis. We performed genotyping of the mutants using fluorescent PCR as described earlier.

RNA sequencing and qRT-PCR

Total RNA from wild-type controls and *tspeara*$^{-/-}$;*tspearb*$^{-/-}$ mutant animals was extracted using the TRIzol Reagent (Thermo Fischer Scientific, Waltham, MA, USA) and purified using a miRNeasy Mini kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. Samples with RNA integrity numbers (RINs) ≥9 were used for cDNA library preparation and further sequencing. After quality filtering, the reads were aligned against the zebrafish genome (GRCz11). The differentially expressed genes (DEGs) between the wild type (WT) and *tspeara*$^{-/-}$;*tspearb*$^{-/-}$ were identified using RNA isolated from 5 days post-fertilization (dpf) larvae and 7-month-old zebrafish teeth. Selected candidate genes were confirmed by qRT-PCR.

For qRT-PCR experiments, cDNA was prepared from mRNA extracted from either whole animals (mutants and controls) at 5 dpf or teeth from 7-month-old animals (mutants and controls) using the iScript cDNA synthesis kit (Bio-Rad, Hercules, CA, USA). qRT-PCR was performed with SYBR Green Supermix according to the manufacturer's instructions (Thermo Fisher Scientific). The relative quantifications were done from three or four independent experimental replicates and three technical replicates per amplification and then normalized to the housekeeping gene *18S*. The primer sequences for qRT-PCR are described in Table S2. The cycle threshold (Ct) value data were imported into Microsoft Excel for the relative gene expression analysis, and quantification was carried out based on the $2^{(-\Delta\Delta CT)}$ method using corresponding controls as calibrators.

Alizarin red S staining of larvae and adults

Tspear double mutants and control animals at larval and adult stages were stained with alizarin red S staining to visualize teeth development using the well-established protocol as described by

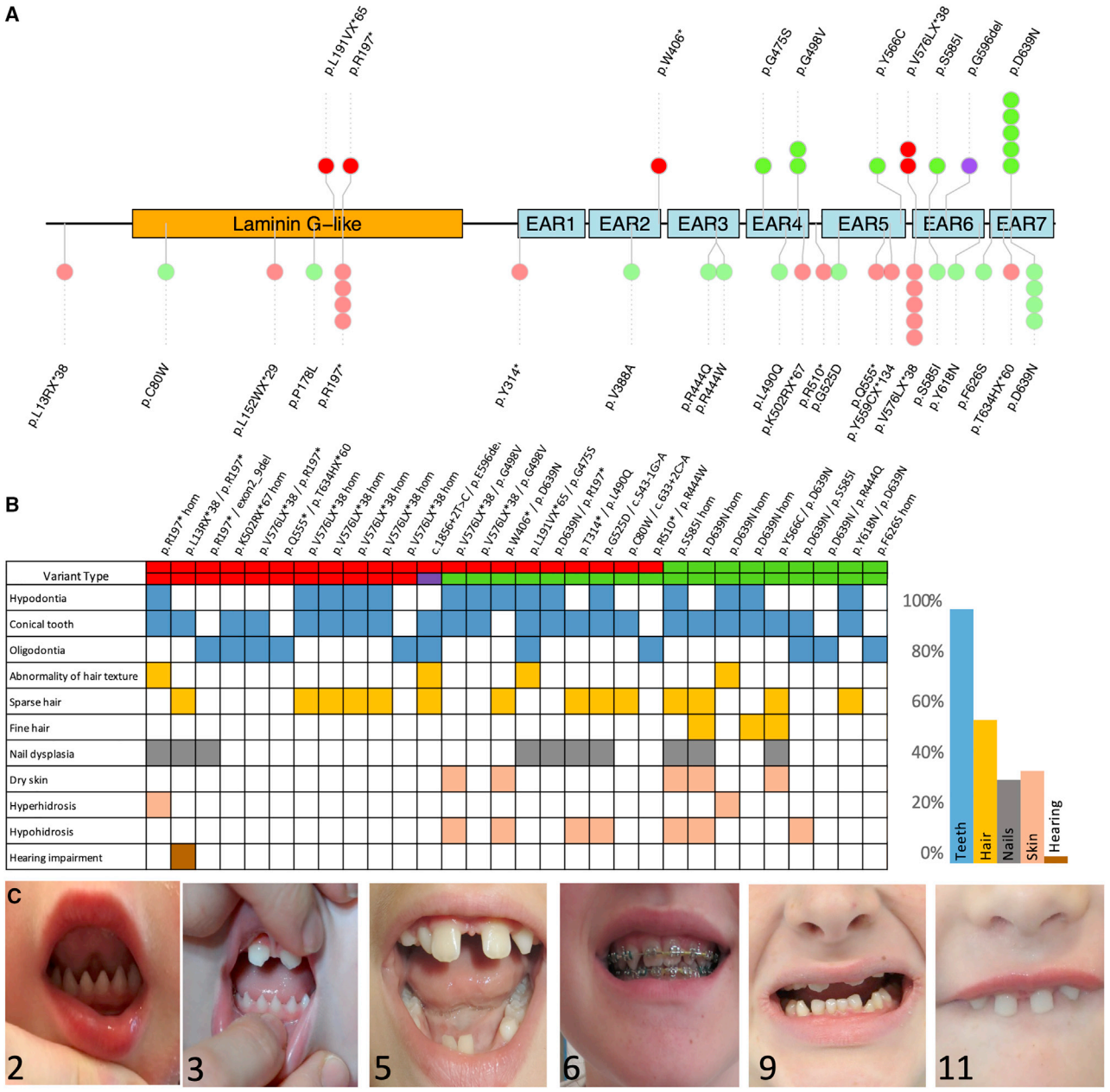

Figure 1. Genotype and phenotype of TSPEAR-related autosomal recessive ectodermal dysplasia
(A) Gene diagram showing all protein-coding variants identified in this study (top) and previously reported (bottom), mapped onto the known protein domains.
(B) Phenotype heatmap among newly and previously described cases showing a predominant dental phenotype.
(C) Images of individuals 2, 3, 5, 6, 9, and 11. Note conical-shaped and widely spaced teeth.

Connolly et al.[55] After the staining, teeth were dissected for visualization using a stereo microscope.

Results

Dental anomalies are the predominant phenotype of ARED14

We performed panel-agnostic analysis[14–17] of trio whole-genome sequencing (WGS) data from a 6-year-old girl with a clinical diagnosis of ED in whom initial panel-led analysis had not yielded a genetic diagnosis. We detected pathogenic compound heterozygous *TSPEAR* variants in her confirming a diagnosis of ARED14 (individual 1 in Table S1). At the time of this finding, only 5 individuals from three families with ARED14 were known in the medical literature.[12] We therefore set out to discover additional patients with biallelic pathogenic or likely pathogenic *TSPEAR* variants and identified 10 additional individuals from eight families (six individuals from five families in the 100KGP, three individuals from two families via clinical multi-gene panel testing for ED, and one individual via DECIPHER) (Figure 1A).

Next, we collected information about the clinical features of these individuals using a standardized proforma. While this work was ongoing, additional individuals with biallelic *TSPEAR* variants were reported.[13,56,57] We combined our clinical data with those from published cases. Combined analysis of phenotypes of 30 individuals revealed dental anomalies in all individuals, with conical-shaped teeth (77%), hypodontia (the absence of between one and five teeth disregarding the third molar, 50%), and oligodontia (the absence of six or more teeth disregarding the third molar, 37%) being the commonest features (Table S1; Figures 1B and 1C). For newly described cases, primary dentition was affected in all and secondary dentition was affected in 64%—affection of secondary dentition in the remaining 36% was unknown, as they had yet to erupt and panoramic radiographs were unavailable. Sparse hair was reported in 47% of individuals, predominantly affecting anterior scalp and temples. Hypohidrosis was reported in 23% and hyperhidrosis in 10% of individuals. Febrile convulsions that are seen in some other EDs were not reported in any individuals. Notably, five individuals from two families have been previously reported with biallelic *TSPEAR* variants and non-syndromic sensorineural hearing loss.[58,59] However, none of the affected individuals in our cohort were reported with sensorineural hearing impairment.

Collectively, these data expand the variant spectrum of ARED14 and show that conical-shaped teeth and hypodontia are the commonest features of this condition, while hair and hypohidrosis phenotypes are relatively mild or infrequent in the majority of affected individuals. At present, there is insufficient evidence to link *TSPEAR* variants as a cause of AR hearing loss.

Destabilization of the β-propeller is the likely mechanism for most missense *TSPEAR* variants

Including individuals presented here, 27 distinct disease-causing *TSPEAR* variants are now known (Figure 1A). Of these, 11 variants (41%) are predicted loss-of-function (pLOF), and 16 (59%) are missense or in-frame deletions. The effects of missense and in-frame *TSPEAR* variants are poorly understood. We first used SpliceAI to check if any had likely splice-altering effects at the nucleotide level (Table S1). This showed a possible splice donor loss (SpliceAI delta 0.42) for the c.1754G>T (p.S585I) variant, suggesting that this allele could lead to LOF through mis-splicing. The SpliceAI scores of the other 15 missense or in-frame insertion or deletion (indel) variants were all below 0.1. 14/16 missense *TSPEAR* ARED14 variants are in EAR domains, while C80W and P178L are located in the laminin-G-like domain (Figure 1A). Variant burden analysis in gnomAD also showed a propensity for predicted deleterious missense *TSPEAR* variants to affect the EAR domains (Figure S3).

Next, to examine the structural basis of deleteriousness of *TSPEAR* variants in more detail, we performed a BLAST search using the amino acid sequence of TSPEAR, which did not reveal any homologous structures in the Protein Data Bank. We next examined the AlphaFold-predicted structure, which revealed that the EAR domains of TSPEAR are predicted to assume a β-propeller formation and can function as part of a ligand-binding domain (Figure 2A).[60] Within the laminin-G-like domain, p.C80W results in the loss of a disulfide bond with residue 26, and p.P178L is likely to introduce a steric clash, thereby disrupting protein folding. The internal walls of the β-propeller are enriched for highly conserved negatively charged residues. Within the β-propeller, p.R444Q, p.R444W, p.G475S, p.L490N, p.G525D, and p.S585I are likely to introduce steric clashes due to the introduction of bulky side chains (Figure 2B). p.E596del, also within the β-propeller, may interfere with a polar contact between E596 and K302 (Figure 2B). Although D597 may rescue this interaction, another charge interaction with R619 is lost, likely destabilizing the β-propeller. p.D639N is predicted to result in the loss of a polar contact between D639 and I330 on an adjacent blade of the β-propeller. These data suggest that destabilization of the β-propeller is the likely mechanism for most missense *TSPEAR* variants.

Multiple founder *TSPEAR* variants exist in the non-Finnish European population and other ethnicities

We noted that a c.1915G>A (p.D639N) missense variant was homozygous in three individuals and was in *trans* with another variant in six individuals among the patients described so far (Figure 1A). We also identified two further homozygotes in the 100KGP dataset for which we were unable to ascertain full phenotype information, and therefore these individuals were not included in the clinical analysis. Furthermore, in gnomADv.2.1.1 also, there is a one individual who is homozygous for p.D639N. Notably, all five individuals in the 100KGP with a homozygous p.D639N variant were of White European origin, and none of their parents were reported to be consanguineous. Furthermore, the p.D639N variant in all five individuals was located within 0.26–1.33 Mb ROHs, including an identical 258 kb (chr21(GRCh38): 44,332,294–44,590,585) stretch (Figure 3A; Table S3). Within this region, we identified 12 SNVs representing core haplotype SNVs. Next, we identified 449 individuals in the 100KGP aggregated gVCF (total = 59,464) who were heterozygous for the p.D639N variant. By extracting the genotypes for these 12 SNVs from the VCF files of all p.D639N carriers, we found the core haplotype SNVs to be in strong linkage disequilibrium in the carrier population (n = 449) (Figure 3B).

Three other variants were recurrent in our cohort: p.R197* (n = 8), p.V576Lfs*38 (n = 7), and p.S585I (n = 2). The p.R197* and p.S585I variants are seen in gnomAD at low frequency and are most commonly found in the non-Finnish European population (Table S4). The p.V576Lfs*38 variant is absent from gnomAD. We identified heterozygous carriers for p.R197* (n = 66), p.S585I (n = 27), and p.V576Lfs*38 (n = 5) in the 100KGP aggregated gVCF. Haplotype analysis showed that all 100KGP

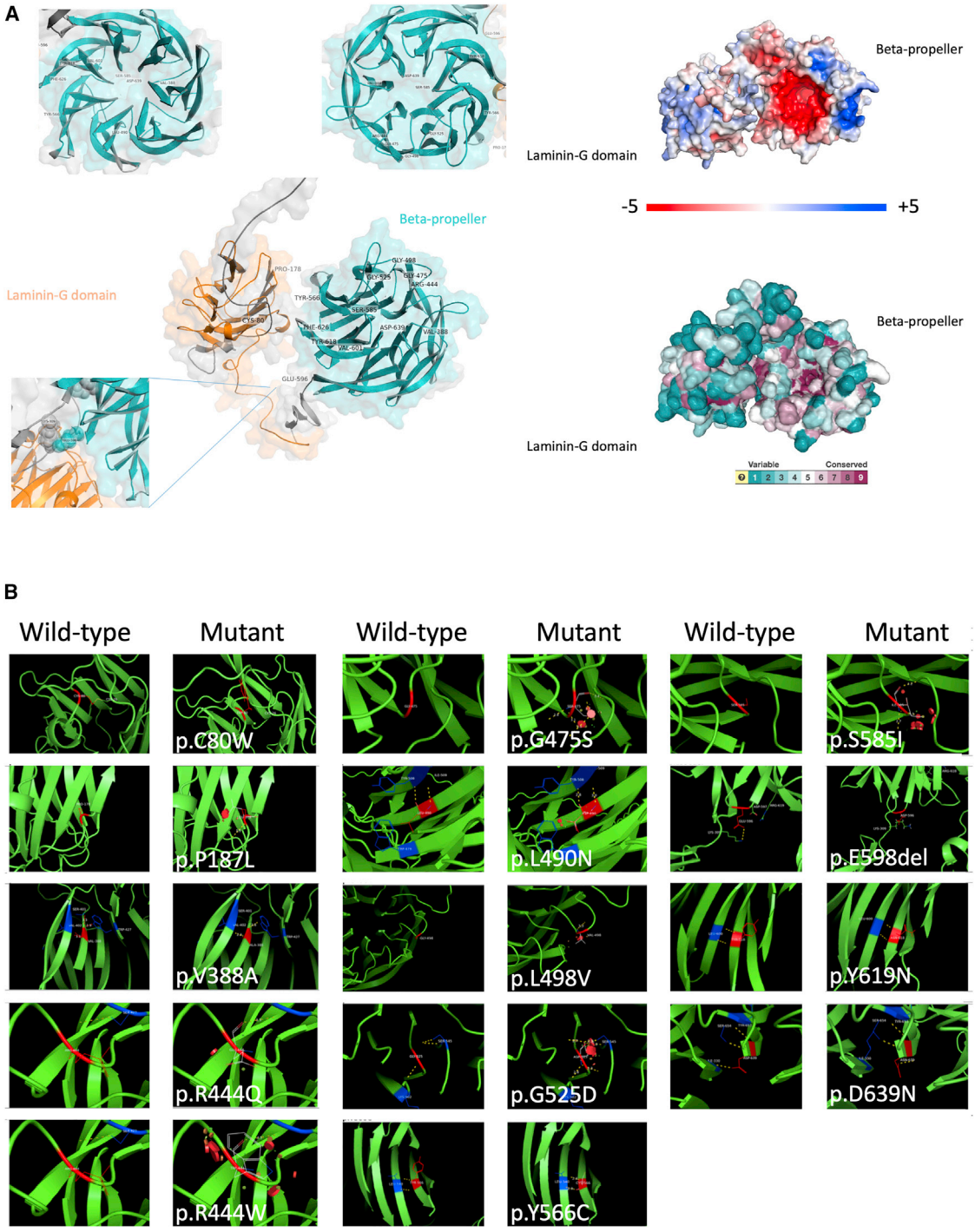

Figure 2. ***In silico*** **protein modeling of pathogenic/likely pathogenic non-truncating TSPEAR variants**
(A) Left shows the AlphaFold-predicted structure of TSPEAR with all missense/in-frame indels plotted; note the propensity for these to affect the β-propeller. Right shows the surface of the predicted TSPEAR structure showing overall charge. Note the pocket of negative charge (red) within the inner surface of the β-propeller. Conservation scores are shown below mapped to the predicted structure surface showing conservation of the residues (purple) within the inner surface of the β-propeller.
(B) Molecular models for all missense/in-frame indels in this study produced by Modeller 9.24. Steric clashes are shown by red discs. Yellow dotted lines indicated interacting residues.

carriers for p.R197*, p.S585I, and p.V576Lfs*38 shared minimal haplotypes of 24, 63, and 51 kb, respectively. Interestingly, all previous reports of the p.V576fs*38 variant have been in individuals of Middle Eastern origin.[12,13,58] However, we detected the p.V576Lfs*38 variant in family 6 with reported White European ancestry. We therefore performed pairwise comparisons of allele sharing between five carriers of the p.V576LX*38 variant, four of White European and one of Asian origin, in the 100KGP dataset. We detected shared haplotype

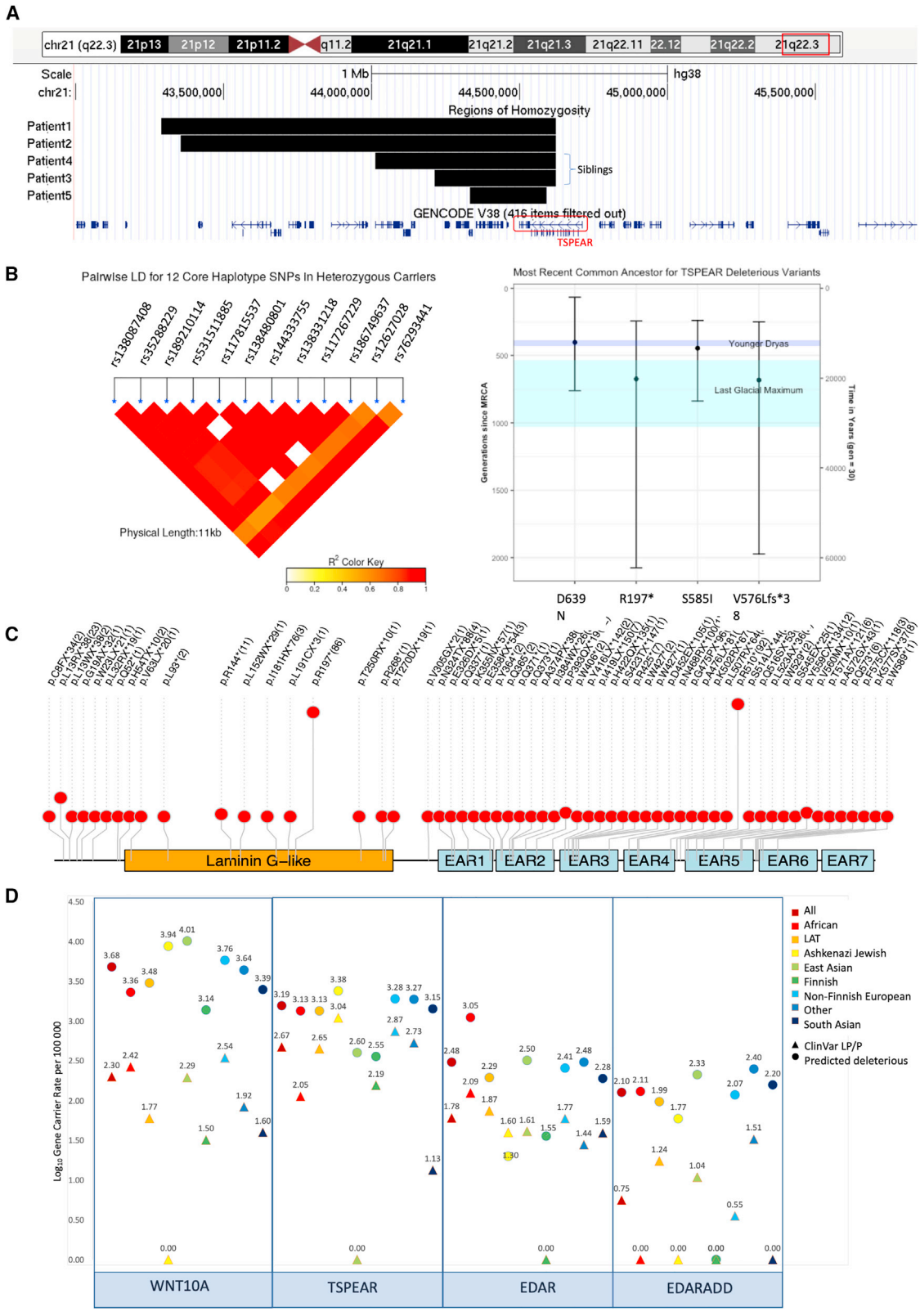

Figure 3. Population and evolutionary genetics of TSPEAR variants

(A) Screenshot from the UCSC browser (GRCh38) showing minimal haplotypes for all founder variants in this study. Regions of homozygosity (ROHs) for the five p.D639N homozygous individuals in the 100KGP are also shown on the lower track.

(B) A linkage disequilibrium heatmap of the 12 core haplotype SNVs for heterozygous p.D639N variant carriers in 100KGP (n = 449).

(C) Bar plot showing the estimated ages of founder variants seen in out TSPEAR cohort. Estimates shown were calculated using the gamma method (recombination clock) after estimating haplotypes using core SNVs for p.D639N and inconsistent homozygous

(legend continued on next page)

between the White European individuals and an Asian individual in the 100KGP (Figure S3). Together, these data establish p.R197*, p.S585I, and p.D639N as founder variants in the non-Finnish European population and the p.V576Lfs*38 variant as a shared founder variant for multiple ethnicities.

Origins of the non-Finnish European founder *TSPEAR* variants converge on the end of the last ice age

Presence of multiple founder variants suggested possible heterozygote advantage. To investigate this possibility, we first set out to calculate the age of the origin of the commonest founder p.D639N *TSPEAR* variant. We identified two discordant mutually exclusive SNVs that likely reflect derived mutations (rs765050949 and rs541272871) within the ROH. Within the 100KGP aggregated gVCF, these SNVs occurred only in individuals who were also heterozygous for the p.D639N variant. We, therefore, concluded that these SNVs likely represent two separate branches of an ancient haplotype. To allow enough time for these derived mutations to have arisen, we estimate that the MRCA for all samples is 286.5 (mean, range: 215.3–387.6) generations. Using the regions of homozygosity as input for the recombination-based gamma method,[27] the MRCA for the homozygotes was estimated to be 130.8 generations ago (confidence interval [CI]: 64.5–269.4). GEVA, using a combined recombination and mutation clock, estimates this variant to have arisen 481.1 generations ago (CI: 165.8–789.5), based on a single concordant pair (i.e., two carriers: NA12489 [Utah Residents (CEPH) with Northern and Western European ancestry (CEU)] and HG02756 [African (AFR)] in the 1000 Genomes Project [1KGP]).[28] To resolve the discrepancy between our results and GEVA, we extracted genotypes from the 1KGP p.D639N carriers for the core SNVs identified above. NA12489 carried 9 of the core SNVs, whereas the AFR sample, HG02756, carried none of the core SNVs. The intervening genomic distance between the core SNVs immediately adjacent to p.D639N (rs117815537 and rs144333755) is ~75 kb, and hence we hypothesized that these two individuals shared a haplotype within this region. Allele-sharing analysis between these two individuals indicated a ~60 kb identity-by-descent (IBD1) segment overlapping the p.D639N variant, indicating a shared haplotype (Figure S5). This suggests that the homozygotes in this cohort may actually share more recent ancestry than the MRCA for the entire sample. For this reason, we delineated haplotypes from the 449 carriers in the 100KGP using genotypes from the 12 core SNVs, which identified 11 distinct haplotypes (Figure S6). Indeed, 23 heterozygous individuals in the 100KGP dataset did not carry any of the core SNVs and hence must share haplotypes of less than ~75 kb. The MRCA based on these 11 haplotypes is 402.1 generations (CI: 67.0–760.7), which is more in line with the estimate from GEVA. Collectively, these data suggest that the MRCA for carriers of this variant lived approximately 400–480 generations ago. We then calculated the MRCA for the other recurrent variants: p.R197*, p.V576Lfs*38, and p.S585I. As these variants are absent from the 1KGP, there are no pre-calculated allele age estimates available from GEVA. Using the haplotypes identified previously, we estimated the MRCA for p.R197*, p.S585I, and p.V576Lfs*38 to be 673.7 (CI: 243.7–2,075.3), 445.3 (CI: 239.2–838.4), and 681.7 (CI: 249.6–1,972.7), respectively. Taking a generation time of 30 years, the mean MRCAs for the four founder variants fall between 12,000 and 20,211 ya, which coincides with the end of the last ice age, indicating a possible correlation between the origin of *TSPEAR* variants and a period of climatic change (Figure 3C).

Next, we searched for *TSPEAR* variants in BAM files for 1,670 published ancient human genomes (ranging from 45,000 BCE to 300 CE) and five published archaic hominin genomes (ranging from 50,000 to 120,000 ya). We identified two reads supportive of the *TSPEAR* p.R197* variant in a heterozygous state in one sample from Ireland (Primrose 9, 3,500–3,360 BCE, DP: 9). In parallel, we also searched for any variants in other common ARED genes—*EDAR*, *EDARADD*, and *WNT10A* in the ancient cohort. We identified a sample from the UK (3,656–3,521 BCE) heterozygous for a novel *WNT10A* variant, G250E, which is predicted to be damaged by multiple *in silico* tools (Table S5; Data S2). We identified the same *EDARADD* missense variant, p.L122Q, in two archaic hominins (Vinidija33.19, 50,000–65,000 BCE, and Denisova11, 84,000–100,000 BCE; Figures S7 and S8). This variant is absent from gnomAD and occurs in a region of high missense constraint (we were unable to identify any protein-changing variation affecting EDARADD residues 120–123 in gnomADv.2.1.1, gnomADv.3, 1KGP, 100KGP, and RD-CONNECT). However, three missense variants in this region, p.D120Y,[8] p.L122R,[61] and p.D123N,[62] have been described in families with autosomal dominant anhidrotic ED. All of these variants lead to near complete loss of downstream nuclear factor κB (NF-κB) activation by a dominant negative mechanism.[10] Furthermore, we also identified a known pathogenic variant in *EDAR* in Denisova11 (p.L397H, 6/13 reads), which has recently been associated with an autosomal dominant form of ED in a single Indian family.[63]

Collectively, these results indicate that several *TSPEAR* variants may have arisen in the non-Finnish European

genotypes for p.R197*, p.S585I, and p.V576Lfs*38. The mean for all variants falls within the last ice age, with both p.R197* and p.V576Lfs*38 falling within the Last Glacial Maximum.
(D) Lolliplot showing pLoF variants in gnomAD for TSPEAR with allele counts in brackets.
(E) Dotplot showing gene-carrier rate for both ClinVar likely pathogenic (LP) and pathogenic (P) alelles (triangles) and all predicted deleterious alleles (circles) in gnomADv.2.1.1. We include p.D639N as a ClinVar LP/P allele given the data we present in this study upgrade this variant to LP based on ACMG criteria (PS4, PM3 and PP3).

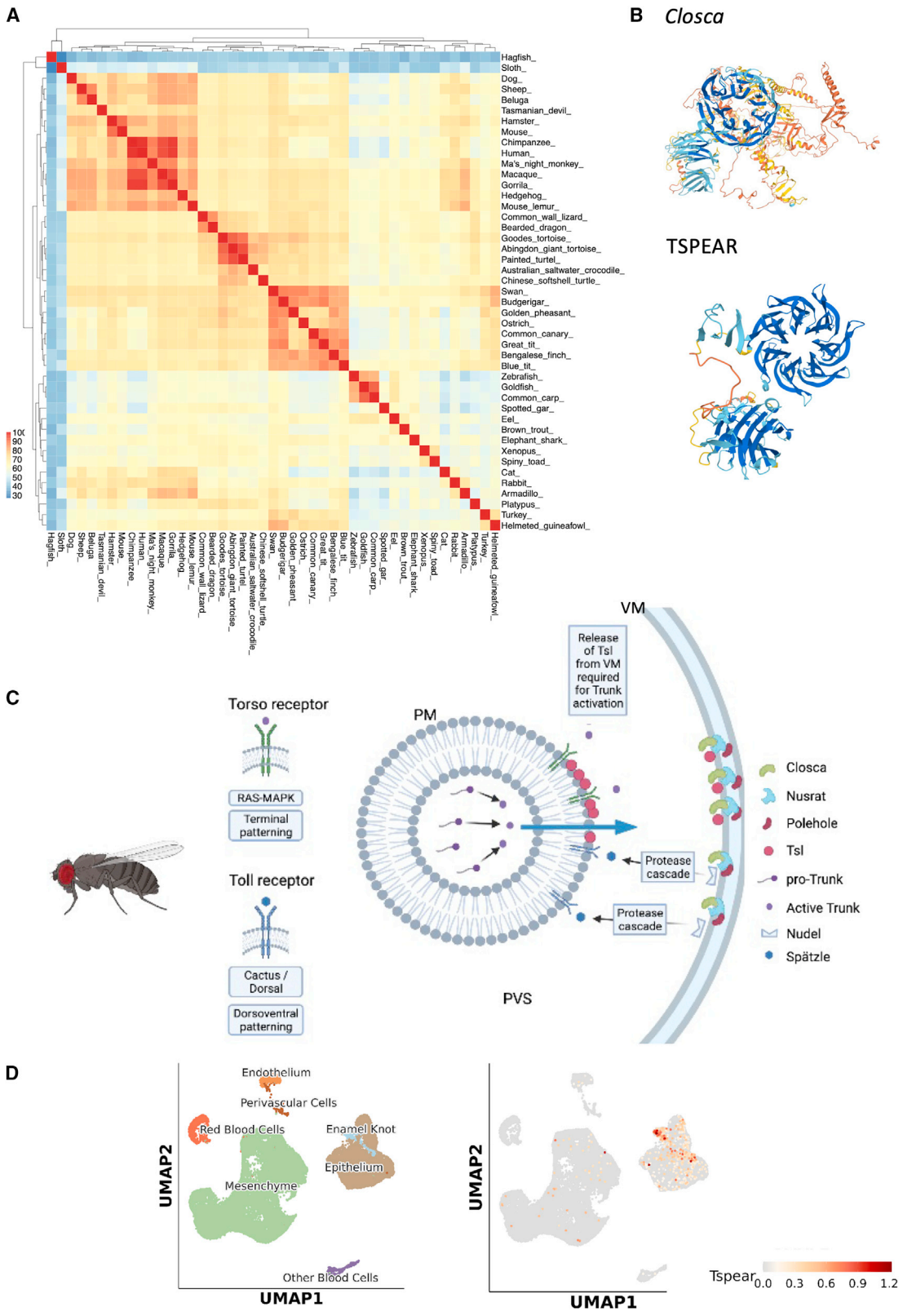

Figure 4. Prediction of TSPEAR function

(A) Heatmap of pairwise sequence identities using cDNA sequences for 45 diverse species. Note the broad conservation in vertebrates and distinct triangles for mammals, birds, and fish.

(B) AlphaFold-predicted structures for both Drosophila *Closca* and TSPEAR; note a predicted β-propeller in both.

(C) Schematic of the known function of *Closca* in both Torso and Toll receptor activation and terminal and dorsoventral patterning, respectively. *Closca*, along with Nusrat and Polehole, sequester Torso-like (tsl) and also Nudel at the vitelline membrane (VM), allowing

(legend continued on next page)

populations around the end of the last ice age and that evidence of pathogenic alleles in ED genes possibly extends as far back as 100,000 years.

TSPEAR-associated ED is likely to be one of the common forms of ARED

The allele frequency of the p.D639N variant in 100KGP aggregated gVCF was 0.0038 (1 in 263 individuals). The frequency of this variant in non-Finnish Europeans in gnomADv.2.1.1 is similar at 0.0035. Furthermore, we noted several pLOF *TSPEAR* variants in gnomAD, suggesting a relatively high carrier frequency in the general population (Figure 3D). We, therefore, calculated the carrier frequencies for ARED14 deleterious alleles across major populations for which data are available in gnomAD. This showed the GCR for ARED14 ranges to be between 0.004 and 0.024 (1 in 287 in the Finnish population to 1 in 43 in the Ashkenazi Jewish population) (Figure 3E). Next, we employed the same criteria to compare carrier frequencies of ARED14 with the three commonest AREDs caused by variants in *EDAR*, *EDARADD*, or *WNT10A*. This analysis showed the GCR for ARED14 to be among the highest in this group. TSPEAR-associated ED is, therefore, likely to be one of the common forms of ARED.

TSPEAR is a functionally conserved gene with a role in ECM- dependent signaling

The understanding of the mechanism of ARED14 has to date been limited because the function of TSPEAR is largely unknown and because the cellular consequences of pathogenic *TSPEAR* variants are poorly understood. To investigate the function of TSPEAR, we performed a BLAST search for the cDNA sequence of human *TSPEAR* to identify orthologous genes in a diverse range of species. Pairwise cDNA sequence identity analysis revealed high conservation in vertebrates with clustering across primates, birds, and fish (Figure 4A). Analysis of protein sequence also reveals conservation in vertebrates with or without teeth, suggesting TSPEAR is likely to be important for the development or function of structures other than tooth (Figures S9 and S10). Next, we generated a *TSPEAR* gene tree based on the Gene Orthology/Paralogy prediction method pipeline in Ensembl, which suggested that *TSPEAR* has a *Drosophila melanogaster* ortholog, *Closca* (Figure S11). *Closca* and human TSPEAR share 26.3% sequence identity at the amino acid level with an E-value of $1.6e^{-6}$. Like TSPEAR, *Closca* contains three putative EAR domains, although its crystalline structure is unknown. The predicted structure of *Closca* in AlphaFold shows a β-propeller (residues 58–419) and an abundance of β strands toward the C-terminal (residues 1534–1835), which is reminiscent of the TSPEAR-predicted structure (Figure 4B). *Closca* is secreted by the *Drosophila* oocyte and incorporated into the vitelline membrane, the inner lining of the eggshell, where it is part of a multi-functional complex required for anchorage and accumulation of morphogens at the poles of the embryo allowing for appropriate spatiotemporal distribution of such molecules (Figure 4C).[64] Eggs of homozygous *Closca* null mutant *Drosophila* females collapse after deposition.[64] The eggshell is an example of a specialized ECM whereby the vitelline membrane supplies positional information to the developing oocyte from the ovary to allow normal dorsoventral and terminal patterning. Importantly, the laminin-G-like domain is related by sequence to the N-terminal domain of Thrombospondin-1, which interacts with heparin sulfate proteoglycans (HSPGs) in the ECM. The laminin-G-like domain of TSPEAR also contains a putative heparin-binding site with an area of prominent positive charge on the external surface of the domain (Figure S12). Spatial and temporal sequestration of growth factors such as ectodermal FGFs, BMPs, and their ECM-dependent signaling via pathways like Wnt and NF-κB is essential to the normal development of ectodermal placodes.[65] Hence, we concluded that *TSPEAR* may have a role in ectodermal development via coordinating ECM-dependent signaling.

Expression of *tspear* is restricted to specific cells in ectodermal tissues

In tooth morphogenesis, the enamel knot coordinates ECM-dependent signaling in the patterning of tooth cusps.[66] We therefore hypothesized that TSPEAR may be functionally related to enamel knot. To investigate this possibility, we analyzed scRNA-seq data from mouse embryonic day 14 (E14) cap-stage tooth cells.[67] Consistent with our hypothesis, this showed expression of *Tspear* to be predominantly in the cluster representing enamel knot (Figure 4D). We also examined scRNA-seq data from murine full-thickness skin in both anagen (5 weeks) and telogen (9 weeks).[68] This showed TSPEAR expression to be restricted to the cluster representing anagen hair follicle keratinocytes in adult mouse skin (http://kasperlab.org/mouseskin; Figure S13). These data showed that *TSPEAR* expression is predominantly in cells of that coordinate ECM-dependent signaling of growth factors in development of ectodermal tissues.

Zebrafish *tspeara*$^{-/-}$;*tspearb*$^{-/-}$ knockout model shows ectodermal anomalies

Having gained insight into the likely role of TSPEAR, we set out to test the effects of its loss in an *in vivo* zebrafish model. Human TSPEAR has two paralogs in zebrafish,

for their precise spatial and temporal accumulation and release, which transfers spatial information to the developing embryo. PM, plasma membrane; PVS, perivitelline space.
(D) scRNA-seq data in the murine developing molar showing restricted expression of *Tspear* to the enamel knot, an important signaling center.

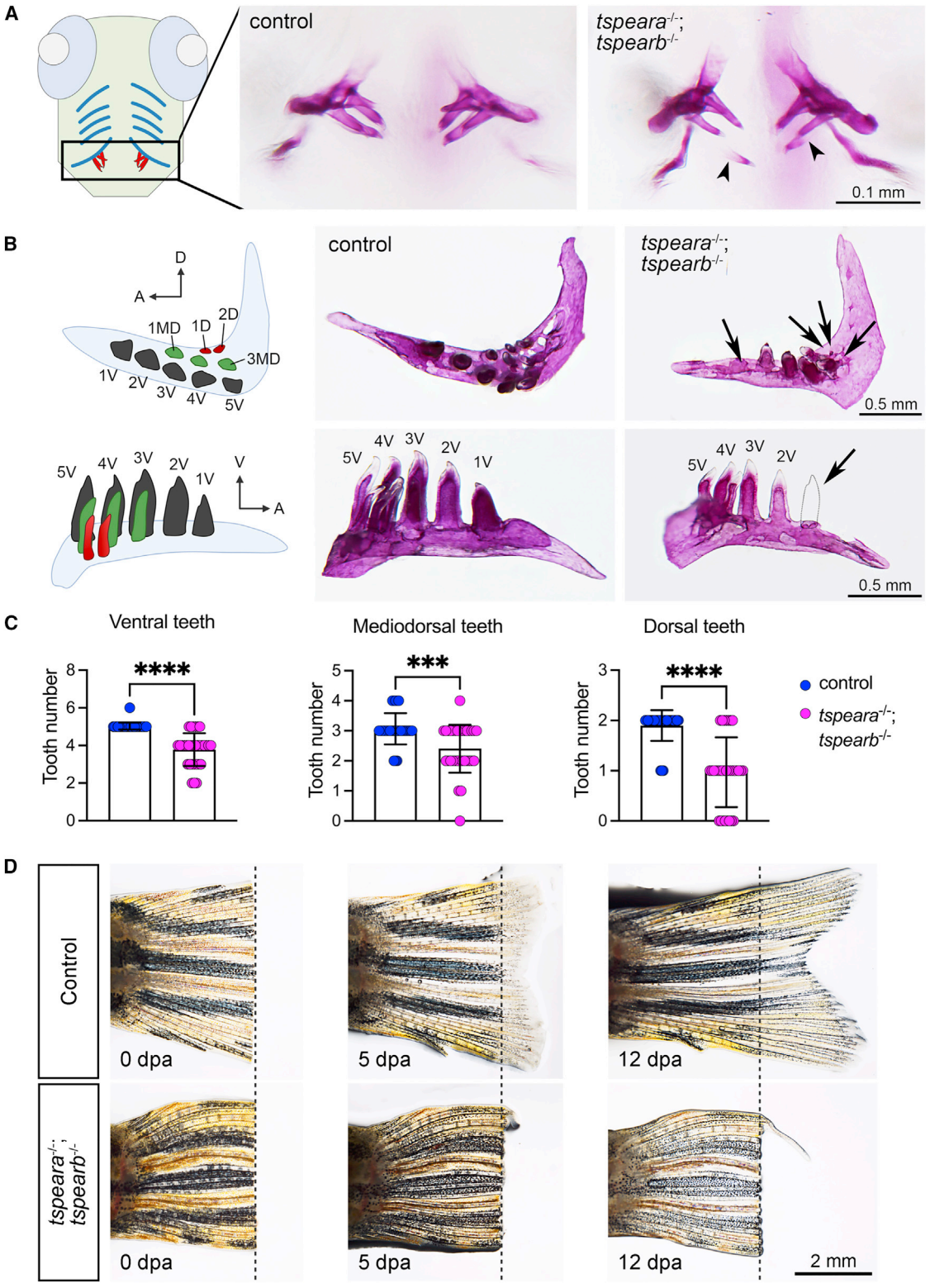

Figure 5. Zebrafish tspear double-knockout model

(A) A schematic figure showing ventral view of larval zebrafish with ventral cartilages (blue) and teeth (red). Alizarin red S staining of pharyngeal teeth shows that a double homozygous mutant (*tspeara*$^{-/-}$;*tspearb*$^{-/-}$) displayed thinner and aberrant mineralized teeth (arrowhead) at 12 days post-fertilization (dpf).

(B) Schematic figure for pharyngeal dentition of adult zebrafish shown in ventral view (top panel) and lateral view (bottom panel). Arrow indicates missing tooth. A, anterior; D, dorsal; MD, mediodorsal; V, ventral.

(legend continued on next page)

tspeara and tspearb, with 73.5%, and 72% similarity, respectively. To understand the function of tspear paralogs in zebrafish, we generated knockouts using CRISPR-Cas9-mediated targeted mutagenesis followed by morphological phenotyping in homozygous animals by breeding two heterozygous animals for each gene. We did not observe any obvious morphological phenotypes by 5 dpf and further raised the animals until adulthood. Adults were genotyped to identify homozygous animals, which are morphologically indistinguishable from their siblings. We hypothesized that since zebrafish mutant might show genetic compensation through transcriptional adaptation in single mutants,[69,70] a tspear LOF phenotype could be assessed in double-mutant fish; therefore, we generated double mutants by breeding *tspeara*$^{-/-}$ and *tspearb*$^{-/-}$ homozygous mutants together, and the resulting *tspeara*$^{+/-}$;*tspearb*$^{+/-}$ animals were further bred to generate double homozygous mutants (*tspeara*$^{-/-}$;*tspearb*$^{-/-}$). Double mutants for both paralogs did not show any significant morphological phenotypes at the larval stage and were viable until adulthood. The adult *tspeara*$^{-/-}$;*tspearb*$^{-/-}$ mutants showed normal morphology except abnormal fin development (Figure S14). In zebrafish, the first tooth develops by 2 dpf, becomes functional by 4 dpf, and is attached to the fifth ceratobranchial cartilage (pharyngeal jaws) in three different rows: ventral, mediodorsal, and dorsal (Figures 5A and 5B, schematics). We analyzed tooth development at larval and adult stages by staining with alizarin red S, which stains mineralized bone structures. Zebrafish *tspeara*$^{-/-}$;*tspearb*$^{-/-}$ mutant analysis revealed thinner and aberrant mineralized teeth at the larval stage by 12 dpf (Figure 5A). We further analyzed teeth in adults, and the mutants consistently showed thinner teeth than age-matched WT controls (Figure 5B). The tooth quantification further showed that many mutant animals did not develop at least one tooth in each row (missing six out of 20 teeth) (Figure 5C), which is similar to the hypodontia phenotype in humans. Furthermore, teeth were firmly attached to the pharyngeal bone in double mutants, but some regions were unstained or poorly stained with alizarin red S stain, indicating poor or missing mineralization (Figure 5B). Interestingly, in addition to tooth anomalies, we observed that caudal, dorsal, anal, and pelvic fins did not show primary bifurcations (Figure S14), which led us to investigate skeleton and bone development in mutants. We stained mutants with alizarin red S stain to visualize the whole skeleton; while mutants did not show any curvature in the spine, the individual vertebrae in the abdominal spine were wider and bifurcated at the base (Figure S14) compared with control animals, and all fins in mutants lacked branching, making their appearance abnormal.

Next, we performed RNA-seq using 5 dpf *tspeara*$^{-/-}$;*tspearb*$^{-/-}$ mutant larvae and validated selected differential expressed genes associated with tooth and bone development by qRT-PCR[71–75] (Figure S15; Data S1). Several genes including *fgf1b*, *sox8b*, *enam*, *scpp5*, *scpp7*, *mustn1a*, and *kcnk5a* were found to be significantly downregulated, and *dlx2b* and *cdkn1a* were upregulated (Figure S16). These DEGs are known to be involved in tooth and bone development; however, they have a relatively minor role in neonatal skeletogenesis and tooth formation. In contrast, expression of *dlx2a*, *fgf1a*, *fgf3*, *fgf4*, *lhx6a*, *lhx8a*, and *pitx2* did not change in mutants. These genes are essential for tooth and bone development, but mutations in these genes cause other severe gross morphological defects.[71,76–78] The RNA-seq analysis from adult *tspear* mutant teeth (Data S1) and qRT-PCR result confirmed significant reduction in the expression of enamel-related genes including *ambn*, *odam*, *enam*, *mmp20a*, *mmp20b*, *fgf1b*, and *scpp5* (Figure S17; Data S1), suggesting that enamel matrix formation, organization, and mineralization genes were affected in mutants. Taken together, the gene expression analysis suggests *tspear* to have a role in regulating key pathways of tooth and bone development in zebrafish.

Together, these data demonstrate that *tspear* is not required for normal viability. The zebrafish double-knockout model recapitulates the predominant clinical feature of ARED14 and presents a potential model to investigate its role in disease pathogenesis.

TSPEAR may be required for the spatial and temporal localization of WNT ligands

TSPEAR likely acts as a sequestrant for growth factors in the ECM essential for ectodermal development (Figure 4C). We set out to utilize the zebrafish double-knockout model to identify TSPEAR-interacting growth factors. The dental phenotype of odontoonychodermal dysplasia, an ARED caused by *WNT10A* variants, closely resembles that of ARED14. WNT10A is a secreted signaling molecule, and like TSPEAR, it demonstrates restricted expression to the enamel knot (Figure S18). *Wnt10a* knockdown disrupts tooth development in zebrafish.[78] We therefore identified WNT10A as a possible TSPEAR-interacting growth factor relevant to ARED14. WNT signaling is known to orchestrate fin regeneration by defining organizing centers,[79] and ECM molecules such as laminin-associated proteins play an important role during fin regeneration.[80–83] We, therefore, investigated the capacity for regeneration of amputated caudal fins in *tspeara*$^{-/-}$;*tspearb*$^{-/-}$ zebrafish.

(C) Quantification of the number of teeth in the ventral, mediodorsal, and dorsal rows. Each dot represents one animal. Control = 30 animals, mutant = 32 animals. Error bars: mean ± SD. Difference was tested using two-tailed unpaired t test with Mann-Whitney test. ***$p < 0.001$, ****$p < 0.0001$.

(D) Adult tail fins were amputated from wild-type control (n = 4) and double homozygous mutants (n = 4). Live fins were imaged at 0, 5, and 12 days post-amputation (dpa). Dashed line indicates the position of amputation.

We observed that mutant fish had a strikingly reduced capacity for fin regeneration and did not regenerate even 12 days post-amputation (Figure 5D) compared with the control animals. This observation supports a likely interaction between TSPEAR and WNT ligands during ectodermal development.

Discussion

We establish that ARED14 is one of the commonest ARED and that it predominantly results in a dental phenotype (Figures 1 and 3). However, the predominant phenotype being dental may mean that cases without any other ectodermal features remain undiagnosed; hence, TSPEAR should be considered in cases of non-syndromic tooth agenesis. This study substantially expands the variant spectrum of *TSPEAR*-related ARED14 and identifies several founder *TSPEAR* variants in multiple ancestry groups. Interestingly, the MRCAs for the NFE founder variants are estimated to have lived during the last ice age, with p.R197* falling within the Last Glacial Maximum (LGM) and both p.D639N and p.S585I falling in the Younger Dryas, another episode of cooling that marked the end of the last ice age. Previous studies on the gain-of-function *EDAR* p.V370A variant demonstrated its association with incisor shoveling, thick hair, and an increased number of eccrine glands.[84] It has been postulated that an increased number of eccrine glands could have been beneficial and led to positive selection for this allele in the Beringia refugium, an area that remained free from ice in the LGM and served as a refuge for humans migrating from Asia to North America. Enhanced thermoregulation is hypothesized to have been beneficial in hunter-gatherer societies in this region.[85] We propose that LOF or hypomorphic *TSPEAR* alleles may lead to fewer sweat glands, providing selective advantage for ancient humans during the ice age who were exposed to cold weather extremes. This theory is supported by our identification of likely deleterious alleles in *EDARADD* and *EDAR* in two archaic hominins sequenced to high (<10×) or medium (3–4×) coverage (the Vindija 33.19 Neanderthal and Denisova 11, an F1 Denisova/Neanderthal hybrid). This finding is particularly intriguing given the unique dental morphology described for *Homo neanderthalensis*, including taurodontism,[86] which is a frequent observation in EDs.[87]

The perturbed *tspeara*$^{-/-}$;*tspearb*$^{-/-}$ mutant zebrafish transcriptome explains ectodermal phenotypes of ARED14. For example, *Enam*$^{-/-}$ knockout mice show no true enamel formation with smaller teeth,[88] while expression of *enam* was also downregulated in the teeth of our zebrafish mutants. The actinopterygians-specific *scpp5* encodes enameloid matrix protein and has a similar expression pattern as *ambn* and *enam* in the inner dental epithelial and in odontoblasts[89]; moreover, *scpp5* is also enriched in bone and dorsal fin of zebrafish.[90] Reduced *scpp5* expression could also be contributing to the defects of teeth and bone in mutants. The *Tbx1/Pitx2/Cdkn1a* cascade is known to play a crucial role in regulating cellular apoptosis during tooth morphogenesis.[91,92] Upregulated *Pitx2* expression transcriptionally activates the *Cdkn1a* in the dental epithelium, which results in smaller teeth in *Tbx1*$^{-/-}$ mice. In agreement with previous study, our transcriptomic data showed that *cdkn1a* expression is upregulated in double mutants and possibly explains the smaller teeth phenotypes in *tspear* mutants of zebrafish larvae and adults.

Our results also suggest a putative role of TSPEAR in the ECM dependent on ectodermal development signaling. The laminin-G-like domain in other proteins has been shown to participate in heparin binding, which is interesting given the preponderance of HSPGs in the ECM. Recently, it has been shown that laminin-G-containing proteins are able to bind growth factors directly, in competition with heparin, using this heparin-binding domain.[93] HSPGs play an important role in the trafficking and sequestration of growth factors including FGF and Wnt signaling pathways.[94] In fact, one of the receptors for downstream signaling molecules of *Closca* (TSPEAR's *drosophila* ortholog) is Torso, which shares 52.9% sequence identify with human FGFR1. Notably, LOF *FGFR1* variants cause Kallman syndrome (OMIM: 147950), in which dental anomalies such as hypodontia and oligodontia have been described.[95–97] *FGFR1* polymorphisms have also been independently associated with tooth agenesis without Kallman syndrome.[98] FGF signaling is important at the initiation stage of tooth development in mammals and subsequently controls the invagination of the dental epithelium into the underlying mesenchyme. Our hypothesis-driven experiments show that WNT10A could be one of the key signaling molecules regulated by TSPEAR. Interestingly, WNT10A is upregulated after amputation,[99] while treatment with inhibitors of WNT ligands, such as, IWP-2, actually prevents fin regeneration.[100] Following amputation, the blastema, which is reminiscent of ectodermal placodes, is formed and acts as a signaling and proliferation center for fin regeneration. WNT signaling is restricted to the distal non-proliferative zone of the blastema but regulates epidermal patterning and blastemal cell proliferation through its downstream effects of FGF and BMP.[79] Moreover, *tspeara* shows a skewed distribution in the regenerating fin, with higher expression in the distal region (2× fold change, $p = 2.36 \times 10^{-10}$, false discovery rate [FDR] 3.19×10^{-9}).[101] Finally, WNT10A has been postulated to be a critical regulator of adult epithelial regeneration, and hence the observation of a regenerative phenotype in *tspeara*$^{-/-}$;*tspearb*$^{-/-}$ knockout animals supports their interaction.[102]

Our work sheds light on a previously uncharacterized protein and confirms the causal relationship between LOF of TSPEAR and ARED14. We also demonstrate that the clinical features can be variable; however, ARED14 is likely to be one of the most common forms of ARED.

Data availability

Researchers can apply to access data from the 100KGP at www.genomicsengland.co.uk/join-a-gecip-domain; upon approval, access is via a secure research environment. Otherwise, the published article includes all datasets generated or analyzed during this study.

Consortia

The members of the Genomics England Research Consortium are J.C. Ambrose, P. Arumugam, R. Bevers, M. Bleda, F. Boardman-Pretty, C.R. Boustred, H. Brittain, M.A. Brown, M.J. Caulfield, G.C. Chan, A. Giess, J.N. Griffin, A. Hamblin, S. Henderson, T.J.P. Hubbard, R. Jackson, L.J. Jones, D. Kasperaviciute, M. Kayikci, A. Kousathanas, L. Lahnstein, A. Lakey, S.E.A. Leigh, I.U.S. Leong, F.J. Lopez, F. Maleady-Crowe, M. McEntagart, F. Minneci, J. Mitchell, L. Moutsianas, M. Mueller, N. Murugaesu, A.C. Need, P. O'Donovan, C. A. Odhams, C. Patch, D. Perez-Gil, M.B. Pereira, J. Pullinger, T. Rahim, A. Rendon, T. Rogers, K. Savage, K. Sawant, R.H. Scott, A. Siddiq, A. Sieghart, S.C. Smith, A. Sosinsky, A. Stuckey, M. Tanguy, A.L. Taylor Tavares, E.R.A. Thomas, S.R. Thompson, A. Tucci, M.J. Welland, E. Williams, K. Witkowska, S.M. Wood, and M. Zarowiecki.

The members of the Solve-RD consortium are Olaf Riess, Tobias B. Haack, Holm Graessner, Birte Zurek, Kornelia Ellwanger, Stephan Ossowski, German Demidov, Marc Sturm, Julia M. Schulze-Hentrich, Rebecca Schüle, Christoph Kessler, Melanie Wayand, Matthis Synofzik, Carlo Wilke, Andreas Traschütz, Ludger Schöls, Holger Hengel, Peter Heutink, Han Brunner, Hans Scheffer, Nicoline Hoogerbrugge, Alexander Hoischen, Peter A.C. 't Hoen, Lisenka E.L.M. Vissers, Christian Gilissen, Wouter Steyaert, Karolis Sablauskas, Richarda M. de Voer, Erik-Jan Kamsteeg, Bart van de Warrenburg, Nienke van Os, Iris te Paske, Erik Janssen, Elke de Boer, Marloes Steehouwer, Burcu Yaldiz, Tjitske Kleefstra, Anthony J. Brookes, Colin Veal, Spencer Gibson, Marc Wadsley, Mehdi Mehtarizadeh, Umar Riaz, Greg Warren, Farid Yavari Dizjikan, Thomas Shorter, Ana Töpf, Volker Straub, Chiara Marini Bettolo, Sabine Specht, Jill Clayton-Smith, Siddharth Banka, Elizabeth Alexander, Adam Jackson, Laurence Faivre, Christel Thauvin, Antonio Vitobello, Anne-Sophie Denommé-Pichon, Yannis Duffourd, Emilie Tisserant, Ange-Line Bruel, Christine Peyron, Aurore Pélissier, Sergi Beltran, Ivo Glynne Gut, Steven Laurie, Davide Piscia, Leslie Matalonga, Anastasios Papakonstantinou, Gemma Bullich, Alberto Corvo, Carles Garcia, Marcos Fernandez-Callejo, Carles Hernández, Daniel Picó, Ida Paramonov, Hanns Lochmüller, Gulcin Gumus, Virginie Bros-Facer, Ana Rath, Marc Hanauer, Annie Olry, David Lagorce, Svitlana Havrylenko, Katia Izem, Fanny Rigor, Giovanni Stevanin, Alexandra Durr, Claire-Sophie Davoine, Léna Guillot-Noel, Anna Heinzmann, Giulia Coarelli, Gisèle Bonne, Teresinha Evangelista, Valérie Allamand, Isabelle Nelson, Rabah Ben Yaou, Corinne Metay, Bruno Eymard, Enzo Cohen, Antonio Atalaia, Tanya Stojkovic, Milan Macek, Jr., Marek Turnovec, Dana Thomasová, Radka Pourová Kremliková, Vera Franková, Markéta Havlovicová, Vlastimil Kremlik, Helen Parkinson, Thomas Keane, Dylan Spalding, Alexander Senf, Peter Robinson, Daniel Danis, Glenn Robert, Alessia Costa, Christine Patch, Mike Hanna, Henry Houlden, Mary Reilly, Jana Vandrovcova, Francesco Muntoni, Irina Zaharieva, Anna Sarkozy, Vincent Timmerman, Jonathan Baets, Liedewei Van de Vondel, Danique Beijer, Peter de Jonghe, Vincenzo Nigro, Sandro Banfi, Annalaura Torella, Francesco Musacchia, Giulio Piluso, Alessandra Ferlini, Rita Selvatici, Rachele Rossi, Marcella Neri, Stefan Aretz, Isabel Spier, Anna Katharina Sommer, Sophia Peters, Carla Oliveira, Jose Garcia Pelaez, Ana Rita Matos, Celina São José, Marta Ferreira, Irene Gullo, Susana Fernandes, Luzia Garrido, Pedro Ferreira, Fátima Carneiro, Morris A. Swertz, Lennart Johansson, Joeri K. van der Velde, Gerben van der Vries, Pieter B. Neerincx, Dieuwke Roelofs-Prins, Sebastian Köhler, Alison Metcalfe, Alain Verloes, Séverine Drunat, Caroline Rooryck, Aurelien Trimouille, Raffaele Castello, Manuela Morleo, Michele Pinelli, Alessandra Varavallo, Manuel Posada De la Paz, Eva Bermejo Sánchez, Estrella López Martín, Beatriz Martínez Delgado, F. Javier Alonso García de la Rosa, Andrea Ciolfi, Bruno Dallapiccola, Simone Pizzi, Francesca Clementina Radio, Marco Tartaglia, Alessandra Renieri, Elisa Benetti, Peter Balicza, Maria Judit Molnar, Ales Maver, Borut Peterlin, Alexander Münchau, Katja Lohmann, Rebecca Herzog, Martje Pauly, Alfons Macaya, Anna Marcé-Grau, Andres Nascimiento Osorio, Daniel Natera de Benito, Hanns Lochmüller, Rachel Thompson, Kiran Polavarapu, David Beeson, Judith Cossins, Pedro M. Rodriguez Cruz, Peter Hackman, Mridul Johari, Marco Savarese, Bjarne Udd, Rita Horvath, Gabriel Capella, Laura Valle, Elke Holinski-Feder, Andreas Laner, Verena Steinke-Lange, Evelin Schröck, and Andreas Rump.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2023.100186.

Acknowledgments

This research was made possible through access to the data and findings generated by the 100KGP. The 100KGP is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The 100KGP is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK, and the Medical Research Council have also funded research infrastructure. The 100KGP uses data provided by patients and collected by the National Health Service as part of their care and support. This study makes use of data generated by the DECIPHER community. A full list of centers who contributed to the generation of the data is available from https://deciphergenomics.org/about/stats and via email from contact@deciphergenomics.org. Funding for the DECIPHER project was provided by Wellcome. A.J. and S.B. acknowledge

the support of Solve-RD. The Solve-RD project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 779257.

Declaration of interests

The authors declare no competing interests.

Received: October 4, 2022
Accepted: February 27, 2023

Web resources

The AlphaFold Structural Prediction Database, www.alphafold.ebi.ac.uk

scRNA-seq data for the murine molar, http://kasperlab.org/mouseskin

References

1. Best, A., and Kamilar, J.M. (2018). The evolution of eccrine sweat glands in human and nonhuman primates. J. Hum. Evol. *117*, 33–43.
2. Oeschger, E.S., Kanavakis, G., Halazonetis, D.J., and Gkantidis, N. (2020). Number of teeth is associated with facial size in humans. Sci. Rep. *10*, 1820.
3. Sujon, M.K., Alam, M.K., and Rahman, S.A. (2016). Prevalence of third molar agenesis: associated dental anomalies in non-syndromic 5923 patients. PLoS One *11*, e0162070.
4. Itin, P.H., and Fistarol, S.K. (2004). Ectodermal dysplasias. Am. J. Med. Genet. C Semin. Med. Genet. *131C*, 45–51.
5. Murdock, S., Lee, J.Y., Guckes, A., and Wright, J.T. (1939). A costs analysis of dental treatment for ectodermal dysplasia. J. Am. Dent. Assoc. *136*, 1273–1276.
6. Pinheiro, M., and Freire-Maia, N. (1994). A clinical classification and a causal review. Am. J. Med. Genet. *53*, 153–162.
7. Martin, A.R., Williams, E., Foulger, R.E., Leigh, S., Daugherty, L.C., Niblock, O., Leong, I.U.S., Smith, K.R., Gerasimenko, O., Haraldsdottir, E., et al. (2019). PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. Nat. Genet. *51*, 1560–1565.
8. Cluzeau, C., Hadj-Rabia, S., Jambou, M., Mansour, S., Guigue, P., Masmoudi, S., Bal, E., Chassaing, N., Vincent, M.C., Viot, G., et al. (2011). Only four genes (EDA1, EDAR, EDARADD, and WNT10A) account for 90% of hypohidrotic/anhidrotic ectodermal dysplasia cases. Hum. Mutat. *32*, 70–72.
9. Okita, T., Asano, N., Yasuno, S., and Shimomura, Y. (2019). Functional studies for a dominant mutation in the EDAR gene responsible for hypohidrotic ectodermal dysplasia. J. Dermatol. *46*, 710–715.
10. Asano, N., Yasuno, S., Hayashi, R., and Shimomura, Y. (2021). Characterization of EDARADD gene mutations responsible for hypohidrotic ectodermal dysplasia. J. Dermatol. *48*, 1533–1541. https://doi.org/10.1111/1346-8138.16044.
11. Plaisancié, J., Bailleul-Forestier, I., Gaston, V., Vaysse, F., Lacombe, D., Holder-Espinasse, M., Abramowicz, M., Coubes, C., Plessis, G., Faivre, L., et al. (2013). Mutations in WNT10A are frequently involved in oligodontia associated with minor signs of ectodermal dysplasia. Am. J. Med. Genet. *161A*, 671–678.
12. Peled, A., Sarig, O., Samuelov, L., Bertolini, M., Ziv, L., Weissglas-Volkov, D., Eskin-Schwartz, M., Adase, C.A., Malchin, N., Bochner, R., et al. (2016). Mutations in TSPEAR, encoding a regulator of notch signaling, affect tooth and hair follicle morphogenesis. PLoS Genet. *12*, e1006369.
13. Bowles, B., Ferrer, A., Nishimura, C.J., Pinto E Vairo, F., Rey, T., Leheup, B., Sullivan, J., Schoch, K., Stong, N., Agolini, E., et al. (2021). TSPEAR variants are primarily associated with ectodermal dysplasia and tooth agenesis but not hearing loss: a novel cohort study. Am. J. Med. Genet. *185*, 2417–2433.
14. Vaz, F.M., McDermott, J.H., Alders, M., Wortmann, S.B., Kölker, S., Pras-Raves, M.L., Vervaart, M.A.T., van Lenthe, H., Luyf, A.C.M., Elfrink, H.L., et al. (2019). Mutations in PCYT2 disrupt etherlipid biosynthesis and cause a complex hereditary spastic paraplegia. Brain *142*, 3382–3397.
15. Zweier, M., Begemann, A., McWalter, K., Cho, M.T., Abela, L., Banka, S., Behring, B., Berger, A., Brown, C.W., Carneiro, M., et al. (2019). Spatially clustering de novo variants in CYFIP2, encoding the cytoplasmic FMRP interacting protein 2, cause intellectual disability and seizures. Eur. J. Hum. Genet. *27*, 747–759.
16. O'Donnell-Luria, A.H., Pais, L.S., Faundes, V., Wood, J.C., Sveden, A., Luria, V., Abou Jamra, R., Accogli, A., Amburgey, K., Anderlid, B.M., et al. (2019). Heterozygous variants in KMT2E cause a spectrum of neurodevelopmental disorders and epilepsy. Am. J. Hum. Genet. *104*, 1210–1222.
17. Jackson, A., Banka, S., Stewart, H., Genomics England Research Consortium, Robinson, H., Lovell, S., and Clayton-Smith, J. (2021). Recurrent KCNT2 missense variants affecting p.Arg190 result in a recognizable phenotype. Am. J. Med. Genet. *185*, 3083–3091. https://doi.org/10.1002/ajmg.a.62370.
18. Turnbull, C., Scott, R.H., Thomas, E., Jones, L., Murugaesu, N., Pretty, F.B., Halai, D., Baple, E., Craig, C., Hamblin, A., et al. (2018). The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. BMJ *361*, k1687. https://doi.org/10.1136/bmj.k1687.
19. Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., and Carter, N.P. (2009). DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. Am. J. Hum. Genet. *84*, 524–533.
20. Zurek, B., Ellwanger, K., Vissers, L.E.L.M., Schüle, R., Synofzik, M., Töpf, A., de Voer, R.M., Laurie, S., Matalonga, L., Gilissen, C., et al. (2021). Solve-RD: systematic pan-European data sharing and collaborative analysis to solve rare diseases. Eur. J. Hum. Genet. *29*, 1325–1331.
21. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). 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. *17*, 405–424.
22. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., et al. (2021). Highly accurate protein structure prediction for the human proteome. Nature *596*, 590–596. https://doi.org/10.1038/s41586-021-03828-1.
23. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016). ConSurf 2016: an

improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. *44*, W344–W350.
24. Eswar, N., and Benjamin, W. (2006). Comparative protein structure modeling using Modeller. Curr. Protoc. Bioinforma. *54*, 5.6.1–5.6.37. Chapter 5, Unit-5.6. https://doi.org/10.1002/cpbi.3.
25. Shin, J.-H., Blay, S., Graham, J., and McNeney, B. (2006). **LDheatmap** : an *R* function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. J. Stat. Softw. *16*.
26. Mathieson, I., and McVean, G. (2014). Demography and the age of rare variants. PLoS Genet. *10*, e1004528.
27. Gandolfo, L.C., Bahlo, M., and Speed, T.P. (2014). Dating rare mutations from small samples with dense marker data. Genetics *197*, 1315–1327.
28. Albers, P.K., and McVean, G. (2020). Dating genomic variants and shared ancestry in population-scale sequencing data. PLoS Biol. *18*, e3000586.
29. Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S.R.F., WGS500 Consortium, Wilkie, A.O.M., McVean, G., and Lunter, G. (2014). Integrating mapping-assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat. Genet. *46*, 912–918.
30. Acuna-Hidalgo, R., Veltman, J.A., and Hoischen, A. (2016). New insights into the generation and role of de novo mutations in health and disease. Genome Biol. *17*, 241.
31. Moorjani, P., Sankararaman, S., Fu, Q., Przeworski, M., Patterson, N., and Reich, D. (2016). A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. Proc. Natl. Acad. Sci. USA *113*, 5652–5657.
32. Wang, R.J., Al-Saffar, S.I., Rogers, J., and Hahn, M.W. (2021). Human generation times across the past 250,000 years. Preprint at bioRxiv. https://doi.org/10.1101/2021.09.07.459333.
33. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of ice age europe. Nature *534*, 200–205.
34. Lamnidis, T.C., Majander, K., Jeong, C., Salmela, E., Wessman, A., Moiseyev, V., Khartanovich, V., Balanovsky, O., Ongyerth, M., Weihmann, A., et al. (2018). Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. Nat. Commun. *9*, 5018.
35. Wang, C.-C., Reinhold, S., Kalmykov, A., Wissgott, A., Brandt, G., Jeong, C., Cheronet, O., Ferry, M., Harney, E., Keating, D., et al. (2019). Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. Nat. Commun. *10*, 590.
36. Brace, S., Diekmann, Y., Booth, T.J., van Dorp, L., Faltyskova, Z., Rohland, N., Mallick, S., Olalde, I., Ferry, M., Michel, M., et al. (2019). Ancient genomes indicate population replacement in Early Neolithic Britain. Nat. Ecol. Evol. *3*, 765–771.
37. Olalde, I., Brace, S., Allentoft, M.E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., et al. (2018). The Beaker phenomenon and the genomic transformation of northwest Europe. Nature *555*, 190–196.
38. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature *513*, 409–413.
39. Patterson, N., Isakov, M., Booth, T., Büster, L., Fischer, C.E., Olalde, I., Ringbauer, H., Akbari, A., Cheronet, O., Bleasdale, M., et al. (2022). Large-scale migration into britain during the Middle to late bronze age. Nature *601*, 588–594.
40. Sánchez-Quinto, F., Malmström, H., Fraser, M., Girdland-Flink, L., Svensson, E.M., Simões, L.G., George, R., Hollfelder, N., Burenhult, G., Noble, G., et al. (2019). Megalithic tombs in western and northern Neolithic Europe were linked to a kindred society. Proc. Natl. Acad. Sci. USA *116*, 9469–9474.
41. Schiffels, S., Haak, W., Paajanen, P., Llamas, B., Popescu, E., Loe, L., Clarke, R., Lyons, A., Mortimer, R., Sayer, D., et al. (2016). Iron age and anglo-saxon genomes from east England reveal British migration history. Nat. Commun. *7*, 10408.
42. Saupe, T., Montinaro, F., Scaggion, C., Carrara, N., Kivisild, T., D'Atanasio, E., Hui, R., Solnik, A., Lebrasseur, O., Larson, G., et al. (2021). Ancient genomes reveal structural shifts after the arrival of Steppe-related ancestry in the Italian Peninsula. Curr. Biol. *31*, 2576–2591.e12.
43. Maisano Delser, P., Jones, E.R., Hovhannisyan, A., Cassidy, L., Pinhasi, R., and Manica, A. (2021). A curated dataset of modern and ancient high-coverage shotgun human genomes. Sci. Data *8*, 202.
44. Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. Nature *528*, 499–503.
45. Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L.F., Aximu-Petri, A., Prüfer, K., de Filippo, C., et al. (2014). The genome sequence of a 45,000-year-old modern human from western Siberia. Nature *514*, 445–449.
46. Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P.H., de Filippo, C., et al. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. Nature *505*, 43–49.
47. Mafessoni, F., Grote, S., de Filippo, C., Slon, V., Kolobova, K.A., Viola, B., Markin, S.V., Chintalapati, M., Peyrégne, S., Skov, L., et al. (2020). A high-coverage neandertal genome from Chagyrskaya cave. Proc. Natl. Acad. Sci. USA *117*, 15132–15136.
48. Prüfer, K., de Filippo, C., Grote, S., Mafessoni, F., Korlević, P., Hajdinjak, M., Vernot, B., Skov, L., Hsieh, P., Peyrégne, S., et al. (2017). A high-coverage neandertal genome from Vindija cave in Croatia. Science *358*, 655–658.
49. Meyer, M., Kircher, M., Gansauge, M.T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., de Filippo, C., et al. (2012). A high-coverage genome sequence from an archaic Denisovan individual. Science *338*, 222–226.
50. Slon, V., Mafessoni, F., Vernot, B., de Filippo, C., Grote, S., Viola, B., Hajdinjak, M., Peyrégne, S., Nagel, S., Brown, S., et al. (2018). The genome of the offspring of a Neanderthal mother and a Denisovan father. Nature *561*, 113–116.
51. Wang, J., Raskin, L., Samuels, D.C., Shyr, Y., and Guo, Y. (2015). Genome measures used for quality control are dependent on gene function and ancestry. Bioinformatics *31*, 318–323.
52. Guo, M.H., and Gregg, A.R. (2019). Estimating yields of prenatal carrier screening and implications for design of expanded carrier screening panels. Genet. Med. *21*, 1940–1947.

53. Ou, J., and Zhu, L.J. (2019). A Bioconductor package for interactive and integrative visualization of multi-omics data. Nat. Methods *16*, 453–454.
54. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539.
55. Connolly, M.H., and Yelick, P.C. (2010). High-throughput methods for visualizing the teleost skeleton: capturing autofluorescence of alizarin red. J. Appl. Ichthyol. *26*, 274–277.
56. Du, R., Dinckan, N., Song, X., Coban-Akdemir, Z., Jhangiani, S.N., Guven, Y., Aktoren, O., Kayserili, H., Petty, L.E., Muzny, D.M., et al. (2018). Identification of likely pathogenic and known variants in TSPEAR, LAMB3, BCOR, and WNT10A in four Turkish families with tooth agenesis. Hum. Genet. *137*, 689–703.
57. Song, J.-S., Bae, M., and Kim, J.-W. (2020). Novel TSPEAR mutations in nonsyndromic oligodontia. Oral Dis. *26*, 847–849. https://doi.org/10.1111/odi.13316.
58. Delmaghani, S., Aghaie, A., Michalski, N., Bonnet, C., Weil, D., and Petit, C. (2012). Defect in the gene encoding the EAR/EPTP domain-containing protein TSPEAR causes DFNB98 profound deafness. Hum. Mol. Genet. *21*, 3835–3844.
59. Sloan-Heggen, C.M., Bierer, A.O., Shearer, A.E., Kolbe, D.L., Nishimura, C.J., Frees, K.L., Ephraim, S.S., Shibata, S.B., Booth, K.T., Campbell, C.A., et al. (2016). Comprehensive genetic testing in the clinical evaluation of 1119 patients with hearing loss. Hum. Genet. *135*, 441–450.
60. Yamagata, A., Miyazaki, Y., Yokoi, N., Shigematsu, H., Sato, Y., Goto-Ito, S., Maeda, A., Goto, T., Sanbo, M., Hirabayashi, M., et al. (2018). Structural basis of epilepsy-related ligand–receptor complex LGI1–ADAM22. Nat. Commun. *9*, 1546.
61. Bal, E., Baala, L., Cluzeau, C., El Kerch, F., Ouldim, K., Hadj-Rabia, S., Bodemer, C., Munnich, A., Courtois, G., Sefiani, A., and Smahi, A. (2007). Autosomal dominant anhidrotic ectodermal dysplasias at the EDARADD locus. Hum. Mutat. *28*, 703–709.
62. Wohlfart, S., Söder, S., Smahi, A., and Schneider, H. (2016). A novel missense mutation in the gene EDARADD associated with an unusual phenotype of hypohidrotic ectodermal dysplasia. Am. J. Med. Genet. *170A*, 249–253.
63. Chaudhary, A.k., Mohapatra, R., Nagarajaram, H.A., Ranganath, P., Dalal, A., Dutta, A., Danda, S., Girisha, K.M., and Bashyam, M.D. (2017). The novel EDAR p.L397H missense mutation causes autosomal dominant hypohidrotic ectodermal dysplasia. J. Eur. Acad. Dermatol. Venereol. *31*, e17–e20.
64. Ventura, G., Furriols, M., Martín, N., Barbosa, V., and Casanova, J. (2010). closca, a new gene required for both Torso RTK activation and vitelline membrane integrity. Germline proteins contribute to Drosophila eggshell composition. Dev. Biol. *344*, 224–232.
65. Graham, A., and Shimeld, S.M. (2013). The origin and evolution of the ectodermal placodes. J. Anat. *222*, 32–40.
66. Jung, S.-Y., Green, D.W., Jung, H.-S., and Kim, E.-J. (2018). Cell cycle of the enamel knot during tooth morphogenesis. Histochem. Cell Biol. *149*, 655–659.
67. Hallikas, O., Das Roy, R., Christensen, M.M., Renvoisé, E., Sulic, A.M., and Jernvall, J. (2021). System-level analyses of keystone genes required for mammalian tooth development. J. Exp. Zool. B Mol. Dev. Evol. *336*, 7–17.
68. Joost, S., Annusver, K., Jacob, T., Sun, X., Dalessandri, T., Sivan, U., Sequeira, I., Sandberg, R., and Kasper, M. (2020). The molecular anatomy of mouse skin during hair growth and rest. Cell Stem Cell *26*, 441–457.e7.
69. El-Brolosy, M.A., and Stainier, D.Y.R. (2017). Genetic compensation: a phenomenon in search of mechanisms. PLoS Genet. *13*, e1006780.
70. Salanga, C.M., and Salanga, M.C. (2021). Genotype to phenotype: CRISPR gene editing reveals genetic compensation as a mechanism for phenotypic disjunction of morphants and mutants. Int. J. Mol. Sci. *22*, 3472.
71. Jackman, W.R., Draper, B.W., and Stock, D.W. (2004). Fgf signaling is required for zebrafish tooth development. Dev. Biol. *274*, 139–157.
72. Lombardo, F., Komatsu, D., and Hadjiargyrou, M. (2004). Molecular cloning and characterization of Mustang, a novel nuclear protein expressed during skeletal development and regeneration. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. *18*, 52–61.
73. Ornitz, D.M., and Marie, P.J. (2002). FGF signaling pathways in endochondral and intramembranous bone development and human genetic disease. Genes Dev. *16*, 1446–1465.
74. Kawasaki, K., Kawasaki, M., Watanabe, M., Idrus, E., Nagai, T., Oommen, S., Maeda, T., Hagiwara, N., Que, J., Sharpe, P.T., and Ohazama, A. (2015). Expression of Sox genes in tooth development. Int. J. Dev. Biol. *59*, 471–478.
75. Sock, E., Schmidt, K., Hermanns-Borgmeyer, I., Bösl, M.R., and Wegner, M. (2001). Idiopathic weight reduction in mice deficient in the high-mobility-group transcription factor Sox8. Mol. Cell Biol. *21*, 6951–6959.
76. Sperber, S.M., Saxena, V., Hatch, G., and Ekker, M. (2008). Zebrafish dlx2a contributes to hindbrain neural crest survival, is necessary for differentiation of sensory ganglia and functions with dlx1a in maturation of the arch cartilage elements. Dev. Biol. *314*, 59–70.
77. Hendee, K.E., Sorokina, E.A., Muheisen, S.S., Reis, L.M., Tyler, R.C., Markovic, V., Cuturilo, G., Link, B.A., and Semina, E.V. (2018). PITX2 deficiency and associated human disease: insights from the zebrafish model. Hum. Mol. Genet. *27*, 1675–1695.
78. Yuan, Q., Zhao, M., Tandon, B., Maili, L., Liu, X., Zhang, A., Baugh, E.H., Tran, T., Silva, R.M., Hecht, J.T., et al. (2017). Role of WNT10A in failure of tooth development in humans and zebrafish. Mol. Genet. Genomic Med. *5*, 730–741.
79. Wehner, D., Cizelsky, W., Vasudevaro, M.D., Ozhan, G., Haase, C., Kagermeier-Schenk, B., Röder, A., Dorsky, R.I., Moro, E., Argenton, F., et al. (2014). Wnt/β-catenin signaling defines organizing centers that orchestrate growth and differentiation of the regenerating zebrafish caudal fin. Cell Rep. *6*, 467–481.
80. Chen, C.-H., Merriman, A.F., Savage, J., Willer, J., Wahlig, T., Katsanis, N., Yin, V.P., and Poss, K.D. (2015). Transient laminin beta 1a induction defines the wound epidermis during zebrafish fin regeneration. PLoS Genet. *11*, e1005437.
81. Webb, A.E., Sanderford, J., Frank, D., Talbot, W.S., Driever, W., and Kimelman, D. (2007). Laminin alpha5 is essential for the formation of the zebrafish fins. Dev. Biol. *311*, 369–382.
82. Carney, T.J., Feitosa, N.M., Sonntag, C., Slanchev, K., Kluger, J., Kiyozumi, D., Gebauer, J.M., Coffin Talbot, J., Kimmel, C.B., Sekiguchi, K., et al. (2010). Genetic analysis of fin development in zebrafish identifies furin and hemicentin1 as

potential novel fraser syndrome disease genes. PLoS Genet. *6*, e1000907.
83. Domogatskaya, A., Rodin, S., and Tryggvason, K. (2012). Functional diversity of laminins. Annu. Rev. Cell Dev. Biol. *28*, 523–553.
84. Kamberov, Y.G., Wang, S., Tan, J., Gerbault, P., Wark, A., Tan, L., Yang, Y., Li, S., Tang, K., Chen, H., et al. (2013). Modeling recent human evolution in mice by expression of a selected EDAR variant. Cell *152*, 691–702.
85. (2018). Environmental selection during the last ice age on the mother-to-infant transmission of vitamin D and fatty acids through breast milk. Proc. Natl. Acad. Sci. USA *115*, E4426–E4432. https://doi.org/10.1073/pnas.1711788115.
86. Bailey, S.E. (2002). A closer look at Neanderthal postcanine dental morphology: the mandibular dentition. Anat. Rec. *269*, 148–156.
87. Gros, C.-I., Clauss, F., Obry, F., Manière, M.C., and Schmittbuhl, M. (2010). Quantification of taurodontism: interests in the early diagnosis of hypohidrotic ectodermal dysplasia. Oral Dis. *16*, 292–298.
88. Hu, J.C.-C., Hu, Y., Smith, C.E., McKee, M.D., Wright, J.T., Yamakoshi, Y., Papagerakis, P., Hunter, G.K., Feng, J.Q., Yamakoshi, F., and Simmer, J.P. (2008). Enamel defects and ameloblast-specific expression in Enam knock-out/lacz knock-in mice. J. Biol. Chem. *283*, 10858–10871.
89. Kawasaki, K., Keating, J.N., Nakatomi, M., Welten, M., Mikami, M., Sasagawa, I., Puttick, M.N., Donoghue, P.C.J., and Ishiyama, M. (2021). Coevolution of enamel, ganoin, enameloid, and their matrix SCPP genes in osteichthyans. iScience *24*, 102023.
90. Liu, Z., Liu, S., Yao, J., Bao, L., Zhang, J., Li, Y., Jiang, C., Sun, L., Wang, R., Zhang, Y., et al. (2016). The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. Nat. Commun. *7*, 11757.
91. Cao, H., Florez, S., Amen, M., Huynh, T., Skobe, Z., Baldini, A., and Amendt, B.A. (2010). Tbx1 regulates progenitor cell proliferation in the dental epithelium by modulating Pitx2 activation of p21. Dev. Biol. *347*, 289–300.
92. Jernvall, J., Aberg, T., Kettunen, P., Keränen, S., and Thesleff, I. (1998). The life history of an embryonic signaling center: BMP-4 induces p21 and is associated with apoptosis in the mouse tooth enamel knot. Development *125*, 161–169.
93. Ishihara, J., Ishihara, A., Fukunaga, K., Sasaki, K., White, M.J.V., Briquez, P.S., and Hubbell, J.A. (2018). Laminin heparin-binding peptides bind to several growth factors and enhance diabetic wound healing. Nat. Commun. *9*, 2163.
94. Margosio, B., Marchetti, D., Vergani, V., Giavazzi, R., Rusnati, M., Presta, M., and Taraboletti, G. (2003). Thrombospondin 1 as a scavenger for matrix-associated fibroblast growth factor 2. Blood *102*, 4399–4406.
95. Sato, N., Katsumata, N., Kagami, M., Hasegawa, T., Hori, N., Kawakita, S., Minowada, S., Shimotsuka, A., Shishiba, Y., Yokozawa, M., et al. (2004). Clinical assessment and mutation analysis of Kallmann syndrome 1 (KAL1) and fibroblast growth factor receptor 1 (FGFR1, or KAL2) in five families and 18 sporadic patients. J. Clin. Endocrinol. Metab. *89*, 1079–1088.
96. Bailleul-Forestier, I., Gros, C., Zenaty, D., Bennaceur, S., Leger, J., and de Roux, N. (2010). Dental agenesis in Kallmann syndrome individuals with FGFR1 mutations. Int. J. Paediatr. Dent. *20*, 305–312.
97. Dodé, C., Levilliers, J., Dupont, J.M., De Paepe, A., Le Dû, N., Soussi-Yanicostas, N., Coimbra, R.S., Delmaghani, S., Compain-Nouaille, S., Baverel, F., et al. (2003). Loss-of-function mutations in FGFR1 cause autosomal dominant Kallmann syndrome. Nat. Genet. *33*, 463–465.
98. Vieira, A.R., Modesto, A., Meira, R., Barbosa, A.R.S., Lidral, A.C., and Murray, J.C. (2007). Interferon regulatory factor 6 (IRF6) and fibroblast growth factor receptor 1 (FGFR1) contribute to human tooth agenesis. Am. J. Med. Genet. *143A*, 538–545.
99. Stoick-Cooper, C.L., Weidinger, G., Riehle, K.J., Hubbert, C., Major, M.B., Fausto, N., and Moon, R.T. (2007). Distinct Wnt signaling pathways have opposing roles in appendage regeneration. Development *134*, 479–489.
100. Stewart, S., Gomez, A.W., Armstrong, B.E., Henner, A., and Stankunas, K. (2014). Sequential and opposing activities of wnt and BMP coordinate zebrafish bone regeneration. Cell Rep. *6*, 482–498.
101. Rabinowitz, J.S., Robitaille, A.M., Wang, Y., Ray, C.A., Thummel, R., Gu, H., Djukovic, D., Raftery, D., Berndt, J.D., and Moon, R.T. (2017). Transcriptomic, proteomic, and metabolomic landscape of positional memory in the caudal fin of zebrafish. Proc. Natl. Acad. Sci. USA *114*, E717–E726.
102. Xu, M., Horrell, J., Snitow, M., Cui, J., Gochnauer, H., Syrett, C.M., Kallish, S., Seykora, J.T., Liu, F., Gaillard, D., et al. (2017). WNT10A mutation causes ectodermal dysplasia by impairing progenitor cell proliferation and KLF4-mediated differentiation. Nat. Commun. *8*, 15397.