



ARTICLE OPEN

Novel *MSX1* variants identified in families with nonsyndromic oligodontiaJinglei Zheng¹, Miao Yu¹, Haochen Liu¹, Tao Cai², Hailan Feng¹, Yang Liu¹ and Dong Han¹

The goal of this study was to identify *MSX1* gene variants in multiple Chinese families with nonsyndromic oligodontia and analyse the functional influence of these variants. Whole-exome sequencing (WES) and Sanger sequencing were performed to identify the causal gene variants in five families with nonsyndromic oligodontia, and a series of bioinformatics databases were used for variant confirmation and functional prediction. Phenotypic characterization of the members of these families was described, and an in vitro analysis was performed for functional evaluation. Five novel *MSX1* heterozygous variants were identified: three missense variants [c.662A>C (p.Q221P), c.670C>T (p.R224C), and c.809C>T (p.S270L)], one nonsense variant [c.364G>T (p.G122*)], and one frameshift variant [c.277delG (p.A93Rfs*67)]. Preliminary in vitro studies demonstrated that the subcellular localization of *MSX1* was abnormal with the p.Q221P, p.R224C, p.G122*, and p.A93Rfs*67 variants compared to the wild type. Three variants (p.Q221P, p.G122*, and p.A93Rfs*67) were classified as pathogenic or likely pathogenic, while p.S270L and p.R224C were of uncertain significance in the current data. Moreover, we summarized and analysed the *MSX1*-related tooth agenesis positions and found that the type and variant locus were not related to the severity of tooth loss. Our results expand the variant spectrum of nonsyndromic oligodontia and provide valuable information for genetic counselling.

International Journal of Oral Science (2021)13:2

; <https://doi.org/10.1038/s41368-020-00106-0>

INTRODUCTION

Congenital tooth agenesis (TA) is one of the most common dental anomalies and may lead to masticatory dysfunction, speech alteration, aesthetic problems, and malocclusion.¹ Based on the existence of other ectodermal symptoms, TA can be classified as syndromic or nonsyndromic. Nonsyndromic TA can be further categorized into hypodontia (absence of one to five permanent teeth, excluding the third molars), oligodontia (absence of six or more permanent teeth, excluding the third molars), and anodontia (complete absence of teeth), based on the number of missing teeth.² The overall incidence of TA is reported to be approximately 2–10%, excluding the third molars,^{3,4} whereas oligodontia is observed in approximately 0.08–0.16% of the population.⁵ TA has been associated with many factors, including genetics, epigenetics, and environmental factors. Among them, genetic defects have been demonstrated to play a major role.⁶

A number of genes have been shown to be associated with nonsyndromic TA, including axin inhibition protein 2 (*AXIN2*), ectodysplasin A (*EDA*), low-density lipoprotein receptor-related protein 6 (*LRP6*), muscle segment homeobox 1 (*MSX1*), paired box 9 (*PAX9*), and wingless-type mouse mammary tumour virus integration site family member 10A (*WNT10A*) and 10B (*WNT10B*). Variants in the above seven genes account for approximately 90% of the variants responsible for nonsyndromic TA.⁷ As the first gene identified in a family with TA,⁸ the *MSX1* gene is a hotspot candidate.

The *MSX1* gene is located on 4p16.2, belongs to the homeobox gene family, and encodes a transcription factor, MSX1.⁹ MSX1 has a highly conserved and important domain called the homeodomain

(HD), which is encoded by exon 2 of the gene.¹⁰ *MSX1* plays crucial roles in regulating epithelial–mesenchymal interactions that are required for the morphogenesis of a variety of embryonic tissues, such as the limb bud, embryonic tail, hair follicle, and tooth bud.¹¹ This molecule also has roles in regulating the development of the nervous system and tumour growth inhibition.¹² *Msx1*^{-/-} mice are characterized by a deficiency of both mandibular and maxillary alveolar bones, a complete cleft palate, and a failure of tooth development, which is arrested at the bud stage.^{13,14} In humans, *MSX1* is associated with nonsyndromic TA (OMIM, #106600—Tooth Agenesis, Selective, 1; STHAG1), nonsyndromic cleft lip with or without cleft palate (CL/P, OMIM #608874, Orofacial Cleft 5, OFC5), Witkop syndrome (OMIM, #189500), and Wolf–Hirschhorn syndrome (WHS, OMIM, #194190).¹² To date, according to the Human Gene Mutation Database (HGMD, <http://www.hgmd.cf.ac.uk/>) and PubMed database, 36 variants have been identified in the *MSX1* gene that contribute to nonsyndromic TA, and most of them are missense or frameshift variants.^{1,6,8,10,15–39}

This study aimed to continue to expand the variant spectrum of *MSX1* associated with nonsyndromic oligodontia, explore the functional impacts of the identified novel variant loci, and analyse the genotype–phenotype correlations.

RESULTS

Clinical findings and variant detection

The TA patterns (excluding the third molars) and pedigrees of five Chinese families are shown in Figs. 1 and 2. Using whole-exome

¹Department of Prosthodontics, Peking University School and Hospital of Stomatology & National Engineering Laboratory for Digital and Material Technology of Stomatology & Beijing Key Laboratory of Digital Stomatology, Beijing, China and ²National Institute of Dental and Craniofacial Research, NIH, Bethesda, MD, USA

Correspondence: Yang Liu (pkussliuyang@bjmu.edu.cn) or Dong Han (donghan@bjmu.edu.cn)

These authors contributed equally: Jinglei Zheng, Miao Yu.

Received: 27 May 2020 Revised: 24 November 2020 Accepted: 26 November 2020

Published online: 08 January 2021

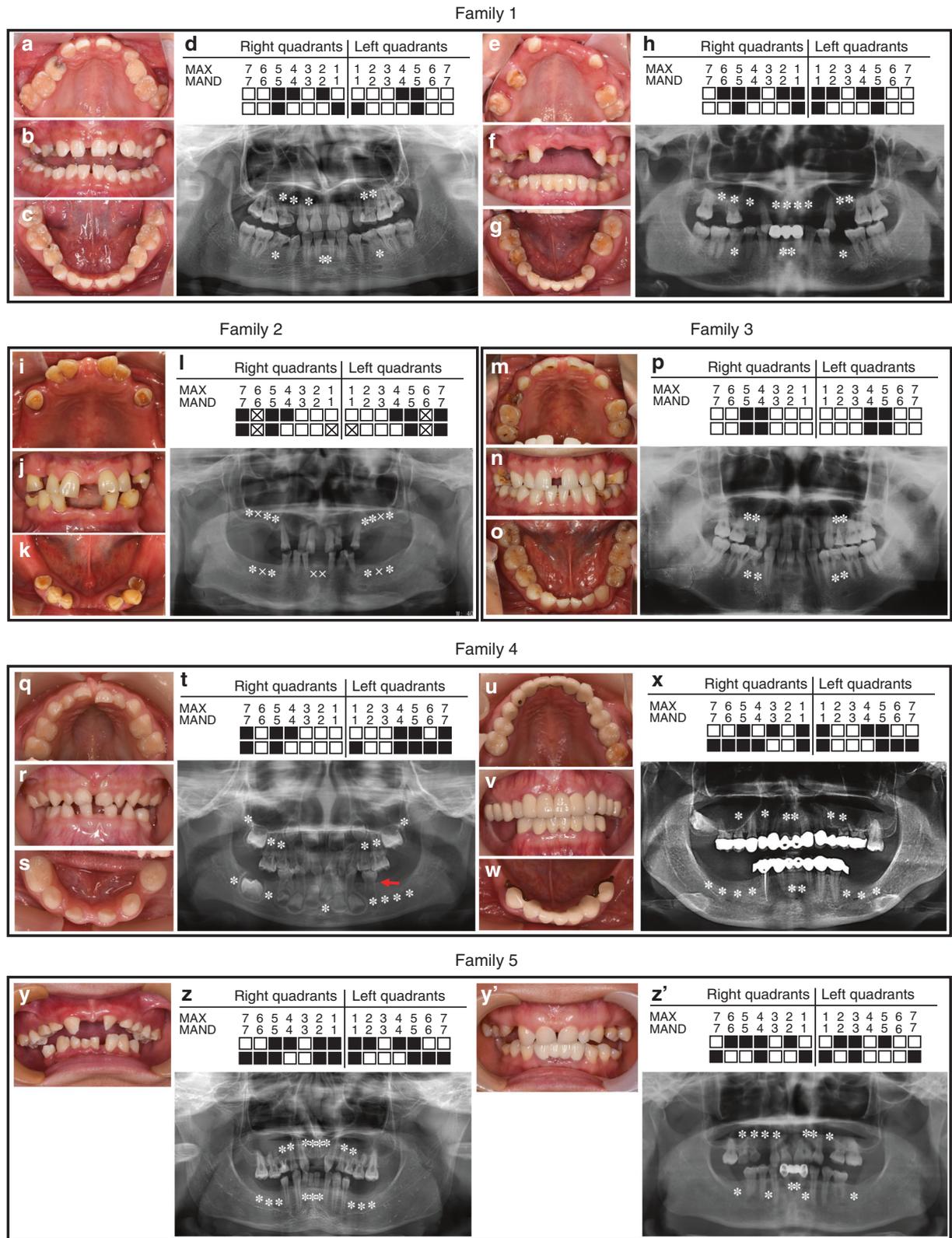


Fig. 1 Dental characteristics of five families with nonsyndromic oligodontia with *MSX1* variants. **a–h** Digital photographs, panoramic radiographs, and schematics of oligodontia in family 1, including proband #1 and her mother. **i–l** Digital photographs, a panoramic radiograph, and a schematic of oligodontia in family 2, including proband #2. **m–p** Digital photographs, a panoramic radiograph, and a schematic of oligodontia in family 3, including proband #3. **q–x** Digital photographs, panoramic radiographs, and schematics of oligodontia in family 4, including proband #4 and his mother. The red arrow indicates taurodontism of the left mandibular first primary molar. **y–z'** Digital photographs, panoramic radiographs, and schematics of oligodontia in family 5, including proband #5 and his mother. Asterisks and solid squares indicate congenitally missing teeth. Crosses and squares with crosses indicate the extracted teeth

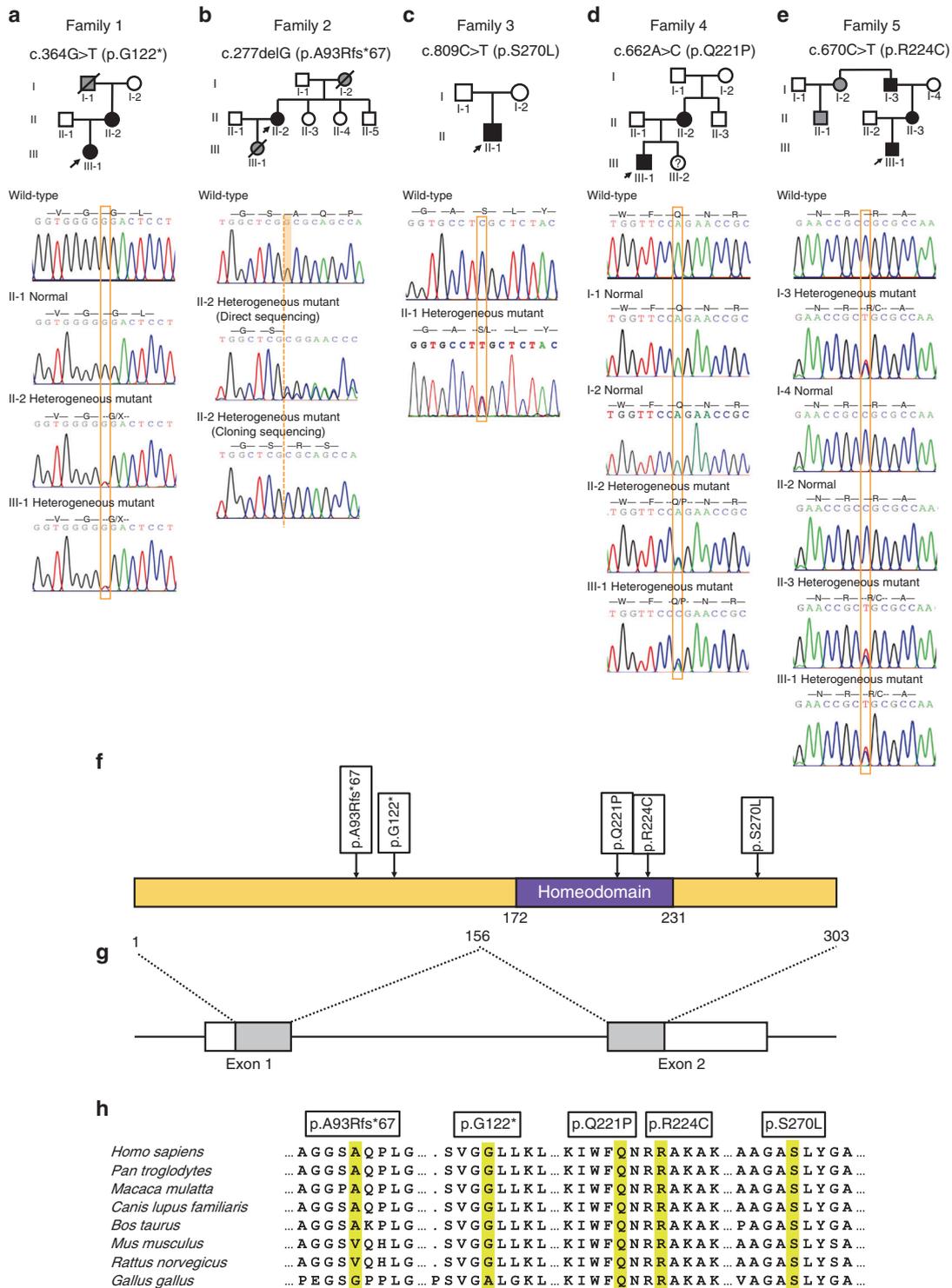


Fig. 2 Sequencing chromatograms of five families and conservation analysis of the *MSX1* variants. **a** Pedigree of Family 1 with nonsyndromic oligodontia. Available DNA sequencing chromatogram showing a heterozygous *MSX1* variant of c.364G>T (p.G122*) in proband #1 (III-1) and her mother (II-2). **b** Pedigree of Family 2 with nonsyndromic oligodontia. Available DNA sequencing chromatogram showing a heterozygous *MSX1* variant of c.277delG (p.A93Rfs*67) in proband #2 (II-2). **c** Pedigree of Family 3 with nonsyndromic oligodontia. Available DNA sequencing chromatogram showing a heterozygous *MSX1* variant of c.809C>T (p.S270L) in proband #3 (II-1). **d** Pedigree of Family 4 with nonsyndromic oligodontia. Available DNA sequencing chromatogram showing a heterozygous *MSX1* variant of c.662A>C (p.Q221P) in proband #4 (III-1) and his mother (II-2). **e** Pedigree of Family 5 with nonsyndromic oligodontia. Available DNA sequencing chromatogram showing a heterozygous *MSX1* variant of c.670C>T (p.R224C) in proband #5 (III-1), his mother (II-3), and his grandfather (I-3). **f** Schematic diagram of the *MSX1* protein structure showing the localization of five novel variants identified in these TA pedigrees. **g** Schematic diagram of the *MSX1* gene. **h** Alignment conservation analysis of the *MSX1* amino acid sequences among different species. Black arrows indicate the proband in each family. Black squares and circles represent TA patients. Grey squares and circles represent individuals with TA, but their DNA samples were not available. Squares and circles with a slash represent individuals who have passed away. A circle with a question mark represents an individual who is too young to be confirmed

Table 1. Harmful prediction of the five novel variants in *MSX1*

Families	Exon	Nucleotide change	Protein change	Variation type	ExAC (MAF)	SIFT	PolyPhen-2	Mutation Taster	ACMG classification (evidence of pathogenicity)
Family 1	1	c.364G>T	p.G122*	Nonsense	—	—	—	Disease causing	Pathogenic (PVS1 + PM2 + PM5 + PP1)
Family 2	1	c.277delG	p.A93Rfs*67	Frameshift	—	—	—	Disease causing	Pathogenic (PVS1 + PM2 + PP3)
Family 3	2	c.809C>T	p.S270L	Missense	—	0.057 (tolerated)	0.213 (benign)	Disease causing	Uncertain significance (PM2 + PP3)
Family 4	2	c.662A>C	p.Q221P	Missense	—	0.032 (damaging)	1.000 (probably damaging)	Disease causing	Likely Pathogenic (PS2 + PM2 + PP1 + PP3)
Family 5	2	c.670C>T	p.R224C	Missense	4.012 × 10 ⁻⁶	0.000 (damaging)	0.998 (probably damaging)	Disease causing	Uncertain significance (PP1 + PP3)

Note: — variant was not found in ExAC
ACMG, American College of Medical Genetics; PM, pathogenic criterion is weighted as moderate; PP, pathogenic as strong; PVS, pathogenic criterion is weighted as very strong

sequencing (WES), we identified five novel heterozygous variants of the *MSX1* gene in these families and confirmed them by Sanger sequencing (Fig. 2a–e). Variants in families 1–4 were not identified in the 1000 Genomes, Exome Aggregation Consortium (ExAC), or gnomAD databases (Table 1).

In family 1, proband #1 was an 18-year-old girl in good health. Clinical and radiographic examinations revealed the absence of nine permanent teeth and the presence of few deciduous teeth. The maxillary central incisors were narrow, and the left maxillary lateral incisor was cone-shaped (Fig. 1a–d). Clinical and panoramic diagnosis revealed that her mother also suffered from oligodontia with the absence of 13 permanent teeth (Fig. 1e–h). The mother declared that the proband’s grandfather also suffered from TA, while the other family members had normal dentition. The entire family had normal hair, skin, and nails, without any other syndromic phenotypes. WES and Sanger sequencing revealed that proband #1 (III-1) and her mother (II-2) carried the same *MSX1* heterozygous nonsense variant, c.364G>T (p.G122*) (Fig. 2a). According to family cosegregation, the variant of the proband was inherited from his mother through an autosomal-dominant mode of inheritance.

In family 2, proband #2 was a 55-year-old woman in good health. According to the dental medical records, all the first molars and mandibular central incisors were lost due to extraction. Clinical and radiographic examinations revealed that she had ten congenitally missing permanent teeth (Fig. 1i–l). She declared that her mother and daughter also suffered from TA, while other family members had normal dentition. WES and Sanger sequencing revealed that proband #2 (II-2) carried an *MSX1* heterozygous frameshift variant, c.277delG (p.A93Rfs*67) (Fig. 2b). Unfortunately, her mother and daughter had passed away and were unavailable for genetic screening.

In family 3, proband #3 was a 28-year-old man in good health. He denied any history of tooth extraction. Clinical and radiographic examinations revealed the absence of eight permanent teeth (Fig. 1m–p). His hair, skin, and nails appeared normal. According to the proband’s description, his parents had normal dentition. Unfortunately, they were unavailable for genetic screening. WES and Sanger sequencing revealed that proband #3 (II-1) carried an *MSX1* heterozygous missense variant, c.809C>T (p.S270L) (Fig. 2c). Based on the current sequencing results, the inheritance pattern and the pathogenicity of this variant could not be clearly defined.

In family 4, proband #4 was a 4-year-old boy in good health. Clinical and radiographic examinations revealed the absence of 13 permanent teeth (Fig. 1q–t) and the presence of taurodontism in his left mandibular first primary molar (Fig. 1t, red arrow).

His mother also suffered from oligodontia, which was restored with maxillary and mandibular fixed bridges (Fig. 1u–w). A panoramic radiograph revealed that she lacked 15 permanent teeth (Fig. 1x). The proband and his mother showed normal development of hair, skin, nails, and facial appearance and denied any history of heat intolerance. The proband had a 1-year-old sister, but dental information and a DNA sample were not available because of her age (III-2). The mother stated that all other family members had normal dentition. WES and Sanger sequencing revealed that proband #4 (III-1) and his mother (II-2) carried an *MSX1* heterozygous missense variant, c.662A>C (p.Q221P), with an autosomal-dominant pattern of inheritance, while the grandparents of the proband (I-1 and I-2) had wild-type sequences (Fig. 2d). Family cosegregation demonstrated that the variant in the proband was inherited from his mother. Furthermore, we confirmed the genetic relationship between the proband’s mother (II-2) and grandparents (I-1 and I-2). Therefore, p.Q221P could be identified as a de novo variant.

In family 5, proband #5 was a 12-year-old boy with good physical health. His facial features, hair, skin, and nails appeared normal. Clinical and radiographic examinations revealed the absence of 17 permanent teeth (Fig. 1y–z). All remaining permanent teeth were smaller than normal, especially the bilateral maxillary first molars, which had a narrow occlusal surface in buccolingual directions and were similar to the deciduous teeth. His mother was missing 13 permanent teeth, which were restored with a fixed bridge on the lower anterior teeth (Fig. 1y’–z’). His mother stated that her father, aunt, and cousin had congenitally missing teeth as well and denied any consanguineous marriages in the family. WES and Sanger sequencing revealed that proband #5 (III-1) carried an *MSX1* heterozygous missense variant, c.670C>T (p.R224C, rs1342784720). We sequenced all the available members in this family (I-3, I-4, II-2, and II-3) and found that the variant was carried by all the affected members (I-3 and II-3) but not by the unaffected members (I-4 and II-2) (Fig. 2e). Thus family cosegregation also revealed an autosomal-dominant inheritance pattern. This variant was found in the ExAC Database (minor allele frequency (MAF): 0.000004012) (Table 1) and Single Nucleotide Polymorphism Database (dbSNP) but had not been reported to be associated with any anomalies.

Conservation and bioinformatics analysis

We conducted conservation and bioinformatics analyses to predict the effects of the five identified novel *MSX1* variants. We found that variants c.277delG (A93Rfs*67) and c.364G>T

(G122*) were located in exon 1 of *MSX1*, and three missense variants, c.662A>C (Q221P), c.670C>T (R224C), and c.809C>T (S270L), were located in exon 2 of *MSX1* (Fig. 2f–g). Among them, Q221P and R224C were located in the highly conserved HD of *MSX1*, which was encoded by exon 2 (Fig. 2f–g). According to Sorting Intolerant from Tolerant (SIFT), polymorphism phenotyping (PolyPhen-2), MutationTaster, and the classification of pathogenic variants with the standards of the 2015 American College of Medical Genetics and Genomics and the Association for Molecular Pathology, p.G122* and p.A93Rfs*67 were predicted to be pathogenic, while p.Q221P was predicted to be likely pathogenic. However, the pathogenicity of p.S270L and p.R224C was of uncertain significance in the current data and requires further investigation (Table 1). The results of amino acid sequence alignment of the *MSX1* protein among multiple species showed that Q221, R224, and S270 were highly evolutionarily conserved but A93 and G122 were not (Fig. 2h).

Preliminary functional analysis

First, the secondary and tertiary protein structures were predicted to analyse the functional effects of *MSX1* variants (Fig. 3a–k). The nonsense variant G122* and frameshift variant A93Rfs led to truncated proteins, resulting in severe changes in the *MSX1* secondary structure (Fig. 3b–c). Missense variants S270L, Q221P, and R224C led to the substitution of one amino acid and caused slight changes in the secondary structure (Fig. 3d–f). Tertiary structure analysis of the Q221P and R224C variants in the HD (Fig. 3g) showed that Q221P resulted in the substitution of a polar residue Gln221 with a nonpolar residue Pro (Fig. 3h–i), and R224C resulted in the substitution of a basic amino acid residue Arg224 with a polar residue Cys, which was much shorter than Arg (Fig. 3j–k).

Second, we performed in vitro functional analysis to evaluate the influence of the five identified novel variants. To test whether these five mutated plasmids could be expressed in vitro, we detected the wild-type and five mutated *MSX1* proteins with green fluorescent protein (GFP) using an anti-GFP antibody. Western blotting and quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis showed that the wild-type and all five mutated proteins were expressed (Fig. 4a and Supplementary Fig. 1). A93Rfs*67 and G122* expressed truncated proteins, as predicted, while Q221P, R224C, and S270L expressed proteins of the same size as the wild type (Fig. 4a).

To evaluate the nuclear localization of the *MSX1* variants, we determined the subcellular localization of the five mutated proteins by immunofluorescence. The results showed that wild-type *MSX1* was located in the nucleus as a transcription factor (Fig. 4c–c’), but the localization of the *MSX1* variants G122*, A93Rfs*67, Q221P, and R224C was altered. GFP-G122* was located in the whole cytoplasm with weak expression in the nucleus (Fig. 4d–d’). GFP-A93Rfs*67 showed a diffuse distribution in the nucleus and dense expression inside the nucleolus (Fig. 4e–e’). GFP-Q221P and GFP-R224C were scattered in the nucleus and densely clumped in the nucleolus (Fig. 4g–g’, h–h’). The results indicated that G122*, A93Rfs*67, Q221P, and R224C severely affected the nuclear localization of *MSX1*, which might contribute to the pathogenic process of oligodontia in our patients. However, the S270L variant seemed to be located in the same place in the cell as the wild type (Fig. 4f–f’), suggesting that the S270L variant might not affect the nuclear localization of *MSX1*.

Statistical analysis of the *MSX1*-related TA pattern

To analyse the specific TA pattern of nonsyndromic TA patients with *MSX1* variants, we compiled the number of congenital missing teeth of 101 patients (Supplementary Table 1) at each of the 7 positions in the 4 quadrants, excluding the third molars and tooth extractions (Table 2). We performed chi-

square tests (or Fisher’s exact test) to determine the statistically significant differences between the different tooth positions (Fig. 5).

We found that the average number of *MSX1*-related missing teeth in 101 patients was 12.26 ± 3.06 . The average incidence rate of missing teeth in the maxillary arch (32.39%) was comparable to that in the mandibular arch (31.05%) ($P > 0.05$). The rate of TA in the right quadrant (31.97%) was also similar to that in the left quadrant (31.47%) ($P > 0.05$). The maxillary second premolars (85.64%), mandibular second premolars (81.19%), and maxillary first premolars (58.42%) were the most frequently missing teeth. In contrast, the teeth with the lowest missing rates were mandibular canines (6.44%), maxillary canines (7.43%), and maxillary first molars (9.41%). Statistically significant differences in the missing tooth rate at different tooth positions are presented in Fig. 5 ($P < 0.05$).

Furthermore, statistical analysis of the relationship between *MSX1* variant domains and missing tooth positions was performed (Supplementary Figs. 2–4) based on the summary of all reported cases with nonsyndromic TA caused by the *MSX1* variant (Supplementary Table 1). Statistical results showed that missing tooth positions appeared to have no significant correlation with the variant domains (exon 1, HD in exon 2, and the rest of the region of exon 2).

DISCUSSION

Nonsyndromic TA caused by *MSX1* variants is inherited in an autosomal-dominant manner.¹ To date, there are 36 *MSX1* variants associated with nonsyndromic TA, and most of them are missense (16/36) and frameshift variants (12/36). We found that, among these variants, 75% of the missense variants are located in the HD, while only 33% of the frameshift variants are located in the HD (Supplementary Table 1). In this study, we identified five novel *MSX1* variants in Chinese families with oligodontia inherited in an autosomal-dominant manner: three missense variants (Q221P, R224C, and S270L), one nonsense variant (G122*), and one frameshift variant (A93Rfs*67). All the missense variants were located in exon 2, while two of them were located in HD, thus supporting the hypothesis that missense variants are more likely to be found in areas of high evolutionary constraint⁴⁰ and thereby suggesting that high conservation could be a marker for amino acids to be a hotspot for missense variants. However, for frameshift variants, the locations were not related to the conservation of amino acids.

Except for S270L, the other four variants showed different subcellular localizations of *MSX1* compared to the wild type. As a highly conserved sequence, the HD is composed of an extended N-terminal arm and three α -helices, which contribute to structural stability, DNA binding specificity, transcriptional repression, and protein interactions.⁴¹ The HD contains two nuclear localization signal (NLS) sequences: NLS1 (₁₆₁RKHKTNRKPR₁₇₀) located upstream of the HD and NLS2 (₂₁₆NRRRAKAKR₂₂₃) located downstream of the HD. NLS1 and NLS2 work cooperatively to mediate the nuclear localization of *MSX1*.⁴² Therefore, the presence or absence of HD variants may result in impaired or lost transcriptional nuclear localization of *MSX1*. Interestingly, only G112* was detected in the cytoplasm and nucleus, thereby suggesting that nuclear localization was impaired. Furthermore, the nucleolar expression of A93Rfs*67, Q221P, and R224C might be due to the opening of chromatin.⁴³ In a previous study, a similar result was found for another *MSX1* missense variant, R196P, which is located in HD.⁴⁴ Notably, although S270L was predicted to cause a change in the *MSX1* secondary structure, the subcellular localization of the S270L variant was not affected. A possible explanation is that S270L is located outside of the HD and does not change the HD structure or nuclear localization. Based on the current results, S270L was identified to have uncertain significance. Whether the variant is pathogenic requires further

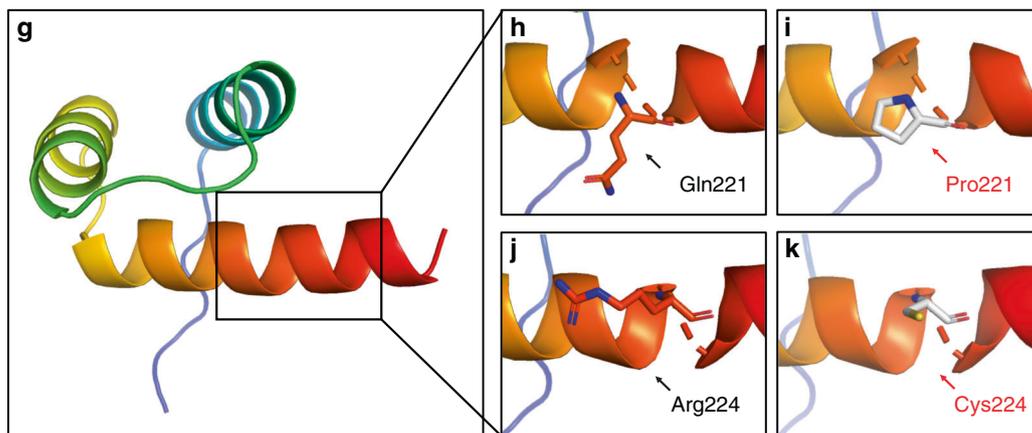
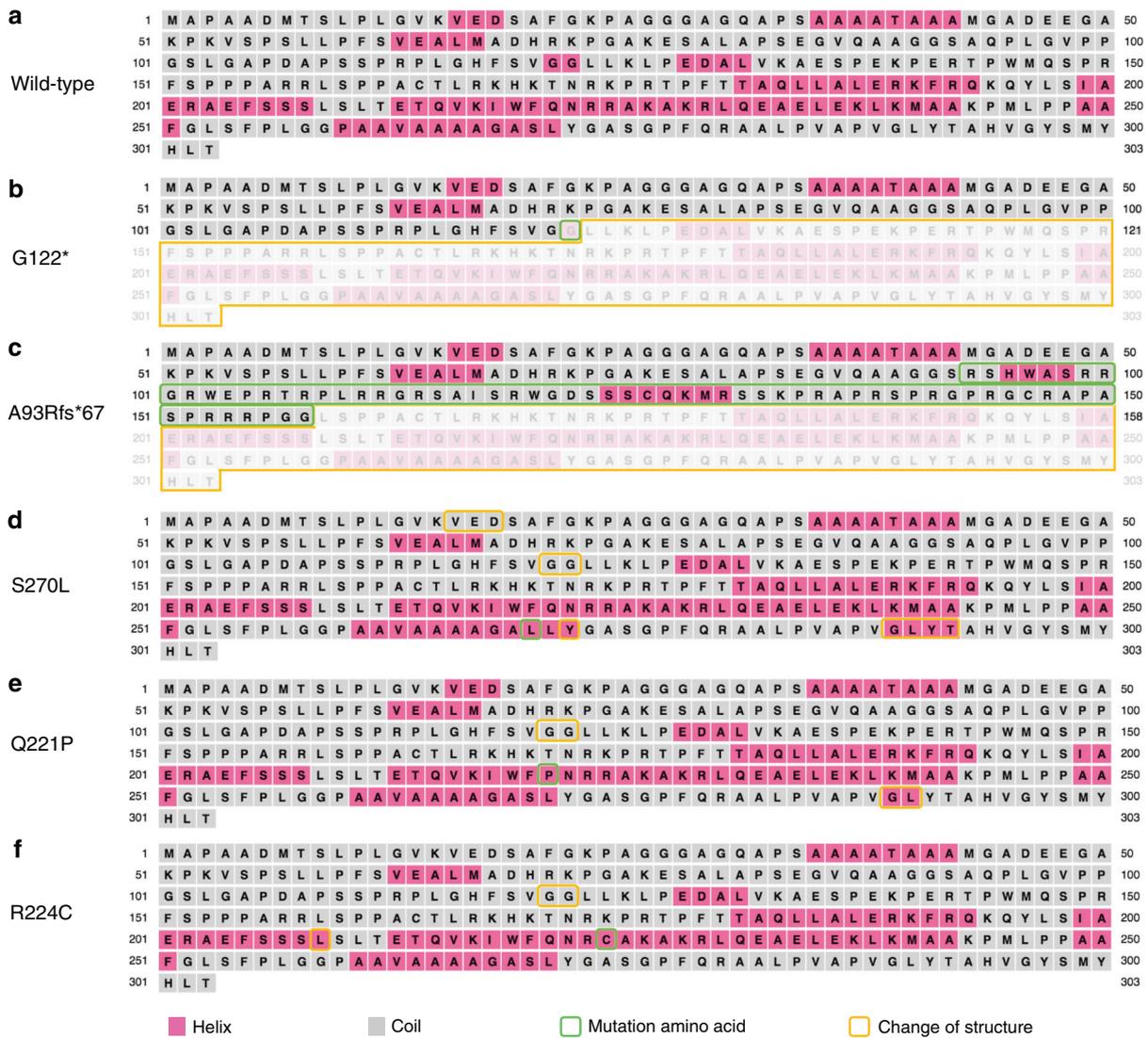


Fig. 3 Secondary and tertiary structure analysis of mutated *MSX1* protein. **a** The predicted secondary structure of the wild-type *MSX1* protein. **b–f** The predicted secondary structure of five mutated *MSX1* proteins. Sites of variants are indicated by green squares. The structural changes in these mutated proteins compared to the wild-type *MSX1* are indicated by orange squares. α -Helices are represented as pink squares, while coils are represented as grey squares. **g** The predicted three-dimensional models of HD in *MSX1*. **h–k** The three-dimensional structural changes of two missense variants. The wild-type (**h**, **j**) and mutated residues (**i**, **k**) are shown in black and red, respectively

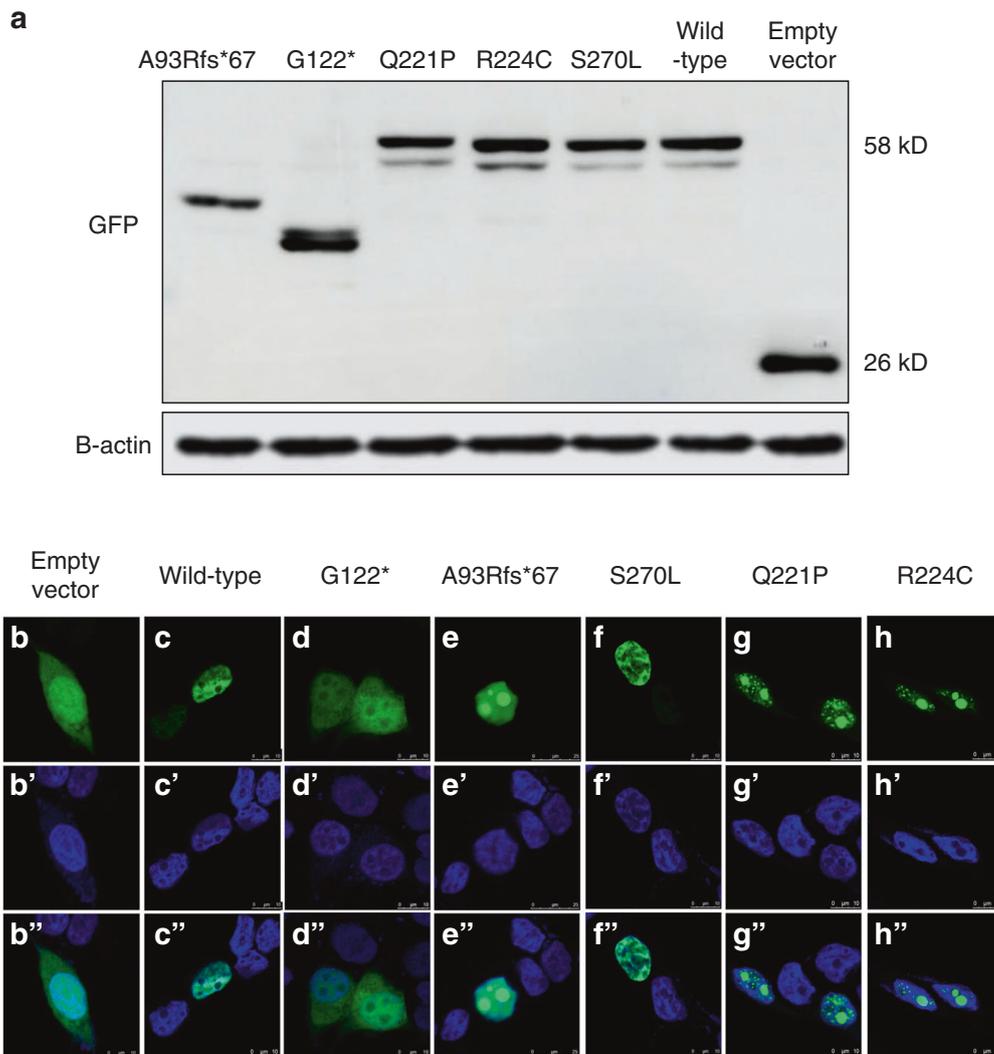


Fig. 4 The expression and subcellular localization of wild-type and mutated *MSX1* proteins. **a** Western blot analysis of total protein from wild-type or mutated *MSX1* cells transfected into 293T cells using anti-GFP and anti- β -actin antibodies. Mutated proteins with normal length and truncated proteins were detected. An empty vector was transfected into cells as a negative control. **b–h** Subcellular localization of wild-type or mutated *MSX1* in vitro. Cells transfected with GFP-G122* show expression in the entire cytoplasm with weak expression in the nucleus. GFP-A93Rfs*67 is detected in the nucleus with dense clumping. GFP-S270L is located in the same location as the wild type. GFP-Q221P and GFP-R224C show densely clumped distributions within the nucleus. The empty vector with GFP was transfected into 293T cells as a negative control. **b'–h'** Nuclei staining by DAPI, 4,6-diamino-2-phenylindole. **b''–h''** Merge of GFP and DAPI. *MSX1* (GFP, green); nuclei (DAPI, blue)

Table 2. The number of missing teeth in 101 patients with *MSX1* mutations

Quadrant	Tooth position							Total missing number
	1	2	3	4	5	6	7	
Max R	13/101	33/101	8/101	61/101	88/101	10/101	24/101	237/707
Max L	11/101	29/101	7/101	57/101	85/101	9/101	23/101	221/707
Mand R	36/101	17/101	5/101	17/101	80/101	26/101	34/101	215/707
Mand L	38/101	13/101	8/101	20/101	84/101	29/101	32/101	224/707
Max	24/202 (11.88%)	62/202 (30.69%)	15/202 (7.43%)	118/202 (58.42%)	173/202 (85.64%)	19/202 (9.41%)	47/202 (23.27%)	458/1414 (32.39%)
Mand	74/202 (36.63%)	30/202 (14.85%)	13/202 (6.44%)	37/202 (18.32%)	164/202 (81.19%)	55/202 (27.23%)	66/202 (32.67%)	439/1414 (31.05%)

Note: (1) The numerators indicate the number of missing teeth and the denominators indicate the number of individuals. (2) There was no statistically significant difference between the left and right arch, therefore the data for equivalent teeth on the left and right were combined at the bottom. (3) The number in the bracket indicates the rate of missing teeth

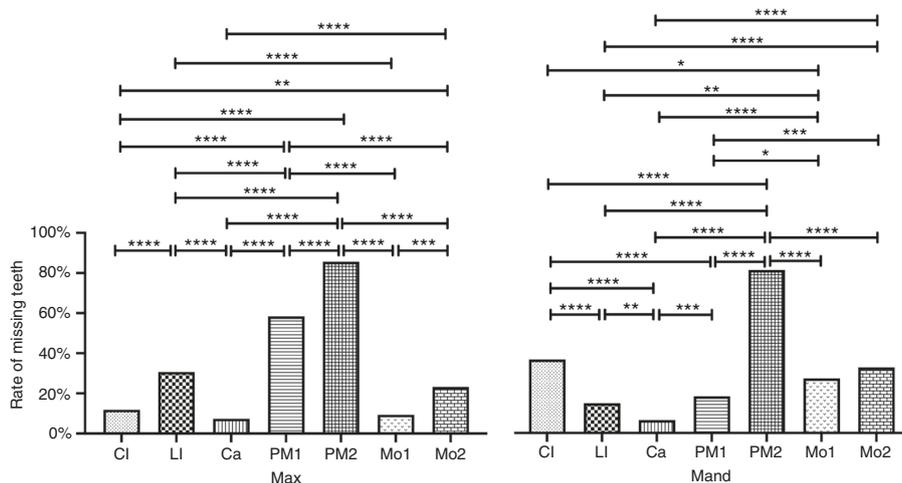


Fig. 5 The pattern of *MSX1*-related tooth agenesis. The rate of missing teeth in 101 patients with *MSX1* variants was compiled at seven tooth positions in the maxillary and mandibular regions. A significant difference is marked by an asterisk (* $P = 0.0332$, ** $P = 0.0021$, *** $P = 0.0002$, and **** $P < 0.0001$). CI, central incisor; LI, lateral incisor; Ca, canine; PM1, first premolar; PM2, second premolar; Mo1, first molar; Mo2, second molar

analysis. Moreover, the G122* variant was mainly located in the cytoplasm, thus indicating that this nonsense variant resulted in a truncated *MSX1* protein without an HD and severely affected the nuclear translocation process of the *MSX1* variant. Although the frameshift variant A93Rfs*67 altered the subsequent sequence after the 93rd amino acid, resulting in a truncated *MSX1* without HD, the A93Rfs*67 variant was detected in the nucleus with dense clumping. This finding might be because the 67 mutated amino acids may have formed a structure similar to that governing *MSX1* nuclear localization in HD. However, further structural and functional experiments are required to elucidate the exact mechanism.

Analysis of the missing tooth positions in 101 patients with *MSX1* variants revealed that the maxillary second premolar was the most vulnerable, followed by the mandibular second premolar and maxillary first premolar, thereby suggesting that *MSX1* is crucial for premolar development. Our findings were consistent with previous studies.^{19,45} Moreover, we found that the maxillary and mandibular canines and the maxillary first molars were least affected. This finding had not been previously reported, suggesting that *MSX1* might be nonessential during the development of these teeth.

In conclusion, the results from our study broaden the variant spectrum of *MSX1*. Information on *MSX1*-related TA patterns offers new evidence for genotypic studies and can guide health care providers in diagnosis, treatment, and genetic counselling.

METHODS

Pedigrees and patients

All the probands and family members of five pedigrees requested treatment advice at the Department of Prosthodontics, Peking University School of Stomatology (Beijing, China) due to their congenital missing teeth. Oral examinations and panoramic dental radiographs were taken to reveal the exact number of missing teeth and other deficiencies. The medical history and physical examinations of the patients confirmed the absence of previous tooth extractions or other developmental abnormalities. Informed consent was obtained from all the participants. This study was approved by the Ethics Committee of Peking University School and Hospital of Stomatology (PKUSSIRB-201736082).

WES, bioinformatics, and conservation analysis

Genomic DNA from these five probands was isolated from peripheral blood lymphocytes using a BioTek DNA Whole-blood Mini Kit (BioTek, Beijing, China) according to the manufacturer's instructions. WES was conducted by Beijing Angen Gene Medicine

Technology (Beijing, China) using DNA from the patients to identify potential pathogenic variants.

A series of bioinformatics databases, such as 1000 Genomes, ExAC, dbSNP, SIFT, PolyPhen-2, and MutationTaster, were used for variant screening and confirmation as well as to predict pathogenicity. Based on the WES results, we filtered all nonsynonymous single-nucleotide variants and insertions/deletions according to the $MAF \leq 0.01$ in the databases mentioned above. Next, we selected all the variants in the genes associated with TA based on reports from previous studies. Then the variants that affect protein function were predicted based on the results obtained from SIFT, PolyPhen-2, and MutationTaster.⁴⁶ The secondary structure of wild-type *MSX1* and five mutated variants was predicted using PsiPred 4.0 (<http://bioinf.cs.ucl.ac.uk/psipred>). An optimum template of the HD (from amino acids 175 to 228) was selected for homology modelling analysis through SWISS-MODEL (<https://swiss-model.expasy.org>). Therefore, the three-dimensional structural changes of Q221P and R224C variants were captured using the PyMOL software (Molecular Graphics System, DeLano Scientific, CA).

For conservation analysis, alignment of the *MSX1* amino acid sequence among multiple species was conducted using ClustalX 2.1. The *MSX1* amino acid sequences of different vertebrate species were obtained from the National Center for Biotechnology Information.

Sanger sequencing and clone sequencing

For verification of the WES results, the related *MSX1* (NM_002448) fragments were sequenced using Sanger sequencing. Genomic DNA from these family members was isolated according to the procedure described above. The primers for two exons of *MSX1*, *MSX1*-1F (5'-CTGGCCTCGCCTTATTAGC-3') and *MSX1*-1R (5'-GCCTGGGTTCTGGCTACTC-3') and *MSX1*-2F (5'-TGGCGGCACTCAATATCTGG-3') and *MSX1*-2R (5'-CAGATCTGTGGTGGTTC-3'), were specifically designed to detect variants. The exonic region of the *MSX1* gene of five families was amplified using PCR with Taq PCR Master Mix (BioTek, Beijing, China). The PCR products were sequenced by Tsingke Biological Technology (Beijing, China). Once the frameshift variant was detected, the PCR product harbouring this variant was cloned into the pClone007 Simple Vector (Tsingke, Beijing, China) to identify the exact status of the variant.

Construction of plasmids

The full-length coding region of the human *MSX1* gene was cloned into the pEGFP-C1 expression vector with enhanced GFP to synthesize the wild-type plasmid pEGFP-C1-*MSX1*. In vitro

site-directed mutagenesis was performed to generate five mutated plasmids: pEGFP-C1-Q221P, pEGFP-C1-R224C, pEGFP-C1-S270 L, pEGFP-C1-G122*, and pEGFP-C1-A93Rfs*67. All plasmids were synthesized by the Beijing Genomic Institute (BGI, Beijing, China), and the entire sequence of the mutated constructs was confirmed by BGI.

Cell culture, transfection, western blot, and qRT-PCR analysis Human embryonic kidney 293T cells were cultured in Dulbecco's modified Eagle's medium (Invitrogen, Grand Island, NY, USA) supplemented with 10% foetal bovine serum and 2 mmol·L⁻¹ L-glutamine in the presence of 5% CO₂. Transient transfection was performed using Lipofectamine 3000 (Invitrogen) according to the manufacturer's instructions. Forty-eight hours after transfection, the cells were collected and lysed with protein lysis buffer (Beyotime Technology, Shanghai, China) containing protease inhibitor cocktail (Solarbio, Beijing, China). Proteins were quantified using the bicinchoninic acid protein concentration assay kit (Invitrogen). Western blot analysis was performed with cell lysates containing 20 µg of total protein. After electrophoresis on a 10% polyacrylamide gel, the proteins were electrophoretically transferred to a polyvinylidene fluoride membrane and incubated with anti-GFP and anti-β-actin mouse polyclonal antibodies (Abcam, Cambridge, UK). The membrane was washed and incubated with peroxidase-conjugated goat anti-mouse secondary antibodies (Abcam).

In addition, total RNA was extracted from the transfected 293T cells using the TRIzol (Carlsbad, CA, USA) method, and 1 µg of the extracted RNA was used to synthesize cDNA with a reverse transcription kit (TaKaRa, Tokyo, Japan) according to the manufacturer's protocol. qRT-PCR was performed on an ABI 7500 system (Applied Biosystems, CA, USA) using 2× Universal SYBR Green Fast qPCR Mix (ABclonal, Wuhan, China). The reaction conditions were 95 °C for 30 s during the initial hold, followed by 40 amplification cycles at 95 °C for 10 s and annealing for 15 s at 60 °C. The relative gene expression was normalized against the gene expression of *GAPDH* and calculated by the 2^{-ΔΔC_t} method. The primers for the human *MSX1* gene were F: 5'-GACATGACT TCTTTGCCACTCG-3' and R: 5'-CAGGAGCGAAGGGGACATTT-3'. All experiments were performed in triplicate.

Subcellular localization assay

After seeding on slides, 293T cells were transiently transfected with pEGFP-C1 expression plasmids containing GFP-wild-type or mutated *MSX1* cDNA fusion proteins. At 48 h post-transfection, the cells were washed three times with phosphate-buffered saline and fixed with 4% paraformaldehyde for 15 min. After three washes with phosphate-buffered saline, the cells were placed in mounting medium with 4',6-diamidino-2-phenylindole (Solarbio, Beijing, China), mounted, and photographed using an LSM 510 Meta confocal microscope (Zeiss, Oberkochen, Germany) with a ×40/1.00 numerical aperture oil objective lens.

Statistical analysis

For verification of the nonsyndromic TA pattern in patients with *MSX1* variants, eight patients with detailed missing tooth site records from this study and previously published studies were included. According to the HGMD and PubMed databases, 36 *MSX1* variants from 31 studies were identified, and 93 patients with detailed documentation of TA sites were assessed (Supplementary Table 1). Therefore, 101 patients were included in the study. Statistical analysis was performed using the GraphPad Prism software (v.8.0). A one-way chi-square test (or Fisher's exact test) was used to compare the differences between tooth positions. Data are presented as the mean ± standard deviation (SD) ($n = 3$), and $P < 0.05$ was considered statistically significant. The P value style was acquired from GraphPad: NS, $P = 0.1234$, * $P = 0.0332$, ** $P = 0.0021$, *** $P = 0.0002$, and **** $P < 0.0001$.

ACKNOWLEDGEMENTS

We thank the participating families. We are also grateful to SAGE for their assistance in editing this article. This study was supported by the National Natural Science Foundation of China (81970902, 81771054, and 81600846).

ADDITIONAL INFORMATION

The online version of this article (<https://doi.org/10.1038/s41368-020-00106-0>) contains supplementary material, which is available to authorized users.

Competing interests: The authors declare no competing interests.

REFERENCES

- Wong, S. W. et al. A novel non-stop mutation in *MSX1* causing autosomal dominant non-syndromic oligodontia. *Mutagenesis* **29**, 319–323 (2014).
- Wang, J. et al. DNA methylation is critical for tooth agenesis: implications for sporadic non-syndromic anodontia and hypodontia. *Sci. Rep.* **6**, 19162 (2016).
- Polder, B. J., Van't Hof, M. A., Van der Linden, F. P. & Kuijpers-Jagtman, A. M. A meta-analysis of the prevalence of dental agenesis of permanent teeth. *Community Dent. Oral Epidemiol.* **32**, 217–226 (2004).
- Ye, X. & Attaie, A. B. Genetic basis of nonsyndromic and syndromic tooth agenesis. *J. Pediatr. Genet.* **5**, 198–208 (2016).
- Rolling, S. & Poulsen, S. Oligodontia in Danish schoolchildren. *Acta Odontol. Scand.* **59**, 111–112 (2001).
- Ma, T. et al. A novel mutation of *MSX1* inherited from maternal mosaicism causes a severely affected child with nonsyndromic oligodontia. *Ann. Hum. Genet.* **84**, 97–101 (2020).
- Yu, M., Wong, S. W., Han, D. & Cai, T. Genetic analysis: Wnt and other pathways in nonsyndromic tooth agenesis. *Oral Dis.* **25**, 646–651 (2019).
- Vastardis, H., Karimbux, N., Guthua, S. W., Seidman, J. G. & Seidman, C. E. A human *MSX1* homeodomain missense mutation causes selective tooth agenesis. *Nat. Genet.* **13**, 417–421 (1996).
- Ivens, A. et al. The human homeobox gene *HOX7* maps to chromosome 4p16.1 and may be implicated in Wolf-Hirschhorn syndrome. *Hum. Genet.* **84**, 473–476 (1990).
- Kimura, M. et al. Novel nonsense mutation in *MSX1* in familial nonsyndromic oligodontia: subcellular localization and role of homeodomain/MH4. *Eur. J. Oral Sci.* **122**, 15–20 (2014).
- Blin-Wakkach, C. et al. Endogenous *Msx1* antisense transcript: in vivo and in vitro evidences, structure, and potential involvement in skeleton development in mammals. *Proc. Natl Acad. Sci. USA* **98**, 7336–7341 (2001).
- Liang, J. et al. *MSX1* mutations and associated disease phenotypes: genotype-phenotype relations. *Eur. J. Hum. Genet.* **24**, 1663–1670 (2016).
- Satokata, I. & Maas, R. *Msx1* deficient mice exhibit cleft palate and abnormalities of craniofacial and tooth development. *Nat. Genet.* **6**, 348–356 (1994).
- Feng, X. Y., Wu, X. S., Wang, J. S., Zhang, C. M. & Wang, S. L. Homeobox protein *MSX-1* inhibits expression of bone morphogenetic protein 2, bone morphogenetic protein 4, and lymphoid enhancer-binding factor 1 via Wnt/β-catenin signaling to prevent differentiation of dental mesenchymal cells during the late bell stage. *Eur. J. Oral Sci.* **126**, 1–12 (2018).
- Lidral, A. C. & Reising, B. C. The role of *MSX1* in human tooth agenesis. *J. Dent. Res.* **81**, 274–278 (2002).
- De Muijnck, S. et al. A novel *MSX1* mutation in hypodontia. *Am. J. Med. Genet. A* **128a**, 401–403 (2004).
- Mostowska, A., Biedziak, B. & Trzeciak, W. H. A novel c.581C>T transition localized in a highly conserved homeobox sequence of *MSX1*: is it responsible for oligodontia? *J. Appl. Genet.* **47**, 159–164 (2006).
- Chishti, M. S., Muhammad, D., Haider, M. & Ahmad, W. A novel missense mutation in *MSX1* underlies autosomal recessive oligodontia with associated dental anomalies in Pakistani families. *J. Hum. Genet.* **51**, 872–878 (2006).
- Kim, J. W., Simmer, J. P., Lin, B. P. & Hu, J. C. Novel *MSX1* frameshift causes autosomal-dominant oligodontia. *J. Dent. Res.* **85**, 267–271 (2006).
- Xuan, K. et al. Identification of a novel missense mutation of *MSX1* gene in Chinese family with autosomal-dominant oligodontia. *Arch. Oral Biol.* **53**, 773–779 (2008).
- Bergendal, B., Klar, J., Stecksén-Blicks, C., Norderyd, J. & Dahl, N. Isolated oligodontia associated with mutations in *EDARADD*, *AXIN2*, *MSX1*, and *PAX9* genes. *Am. J. Med. Genet. A* **155a**, 1616–1622 (2011).
- Mostowska, A., Biedziak, B. & Jagodzinski, P. P. Novel *MSX1* mutation in a family with autosomal-dominant hypodontia of second premolars and third molars. *Arch. Oral Biol.* **57**, 790–795 (2012).
- van den Boogaard, M. J. et al. Mutations in *WNT10A* are present in more than half of isolated hypodontia cases. *J. Med. Genet.* **49**, 327–331 (2012).

24. Liang, J., Zhu, L., Meng, L., Chen, D. & Bian, Z. Novel nonsense mutation in *MSX1* causes tooth agenesis with cleft lip in a Chinese family. *Eur. J. Oral Sci.* **120**, 278–282 (2012).
25. Mu, Y. D. et al. Mutational analysis of *AXIN2*, *MSX1*, and *PAX9* in two Mexican oligodontia families. *Genet. Mol. Res.* **12**, 4446–4458 (2013).
26. Qin, H., Xu, H. Z. & Xuan, K. Clinical and genetic evaluation of a Chinese family with isolated oligodontia. *Arch. Oral Biol.* **58**, 1180–1186 (2013).
27. Arte, S., Parmanen, S., Pirinen, S., Alaluusua, S. & Nieminen, P. Candidate gene analysis of tooth agenesis identifies novel mutations in six genes and suggests significant role for *WNT* and *EDA* signaling and allele combinations. *PLoS ONE* **8**, e73705 (2013).
28. Ceyhan, D., Kirzioglu, Z. & Calapoglu, N. S. Mutations in the *MSX1* gene in Turkish children with non-syndromic tooth agenesis and other dental anomalies. *Indian J. Dent.* **5**, 172–182 (2014).
29. Yamaguchi, S. et al. Characterization of novel *MSX1* mutations identified in Japanese patients with nonsyndromic tooth agenesis. *PLoS ONE* **9**, e102944 (2014).
30. Zhang, X. X., Wong, S. W., Han, D. & Feng, H. L. Simultaneous occurrence of an autosomal dominant inherited *MSX1* mutation and an X-linked recessive inherited *EDA* mutation in one Chinese family with non-syndromic oligodontia. *Chin. J. Dent. Res.* **18**, 229–234 (2015).
31. AlFawaz, S., Plagnol, V., Wong, F. S. & Kelsell, D. P. A novel frameshift *MSX1* mutation in a Saudi family with autosomal dominant premolar and third molar agenesis. *Arch. Oral Biol.* **60**, 982–988 (2015).
32. Tatematsu, T. et al. An aberrant splice acceptor site due to a novel intronic nucleotide substitution in *MSX1* gene is the cause of congenital tooth agenesis in a Japanese family. *PLoS ONE* **10**, e0128227 (2015).
33. Prasad, M. K. et al. A targeted next-generation sequencing assay for the molecular diagnosis of genetic disorders with orodontal involvement. *J. Med. Genet.* **53**, 98–110 (2016).
34. Mitsui, S. N. et al. Novel human mutation and CRISPR/Cas genome-edited mice reveal the importance of C-terminal domain of *MSX1* in tooth and palate development. *Sci. Rep.* **6**, 38398 (2016).
35. Xue, J. et al. A novel *MSX1* intronic mutation associated with autosomal dominant non-syndromic oligodontia in a large Chinese family pedigree. *Clin. Chim. Acta* **461**, 135–140 (2016).
36. Abid, M. F., Simpson, M. A., Petridis, C., Cobourne, M. T. & Sharpe, P. T. Non-syndromic severe hypodontia caused by a novel frameshift insertion mutation in the homeobox of the *MSX1* gene. *Arch. Oral Biol.* **75**, 8–13 (2017).
37. Bonczek, O. et al. Next generation sequencing reveals a novel nonsense mutation in *MSX1* gene related to oligodontia. *PLoS ONE* **13**, e0202989 (2018).
38. Xin, T. et al. A novel mutation of *MSX1* in oligodontia inhibits odontogenesis of dental pulp stem cells via the ERK pathway. *Stem Cell Res. Ther.* **9**, 221 (2018).
39. Yang, L., Liang, J., Yue, H. & Bian, Z. Two novel mutations in *MSX1* causing oligodontia. *PLoS ONE* **15**, e0227287 (2020).
40. Walker, D. R. et al. Evolutionary conservation and somatic mutation hotspot maps of p53: correlation with p53 protein structural and functional features. *Oncogene* **18**, 211–218 (1999).
41. Isaac, V. E., Scivolino, P. & Abate, C. Multiple amino acids determine the DNA binding specificity of the *Msx-1* homeodomain. *Biochemistry* **34**, 7127–7134 (1995).
42. Shibata, A. et al. Identification of nuclear localization signals in the human homeoprotein *MSX1*. *Biochem. Cell Biol.* **96**, 483–489 (2018).
43. Li, Y. J., Fu, X. H., Liu, D. P. & Liang, C. C. Opening the chromatin for transcription. *Int. J. Biochem. Cell Biol.* **36**, 1411–1423 (2004).
44. Wang, Y., Kong, H., Mues, G. & D'Souza, R. *Msx1* mutations: how do they cause tooth agenesis? *J. Dent. Res.* **90**, 311–316 (2011).
45. Nieminen, P. Genetic basis of tooth agenesis. *J. Exp. Zool. B Mol. Dev. Evol.* **312b**, 320–342 (2009).
46. Zheng, J. et al. Tooth defects of EEC and AEC syndrome caused by heterozygous TP63 mutations in three Chinese families and genotype-phenotype correlation analyses of TP63-related disorders. *Mol. Genet. Genomics Med.* **7**, e704 (2019).



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021