J Genet Resour2022;8(2): 165-177 *Homepage: http://sc.journals.umz.ac.ir/*

RESEARCH ARTICLE *DOI: 10.22080/jgr.2022.23318.1304*

J|G|R

In silico Analysis of Possible Novel RNA Interactions and Deleterious Single Nucleotide Polymorphisms Related to *MSX2*, *SHH*, *SMAD7* and *TFAP2* Genes Involved in Odontogenesis

Sajedeh Naghiyan Fesharaki and Sajjad Sisakhtnezhad*

Department of Biology, Faculty of Science, Razi University, Kermanshah, Iran

ARTICLE INFO

Article history:
Received 12 December 2021
Accepted 10 February 2022
Available online 22 February 2022

Keywords:
Bioinformatics analysis
lncRNA
miRNA
SNP
Tooth developmental genes

Corresponding authors:
✉ S. Sisakhtnezhad
s.sisakhtnezhad@razi.ac.ir

p-ISSN 2423-4257
e-ISSN 2588-2589

ABSTRACT

Identification of gene expression profiles, RNA interactions, gene regulation patterns, and single nucleotide polymorphisms (SNPs) is important for determining the molecular mechanisms underlying the normal odontogenesis and the pathology of oral and dental disorders. Therefore, this *in silico* study aimed to identify novel proteins, RNA interactions, and deleterious SNPs related to four major genes (*MSX2*, *SHH*, *SMAD7*, *TFAP2*) involved in the odontogenesis process. After pathway enrichment and gene ontology analysis, the protein-protein, microRNA (miRNA)-mRNA, and miRNA-long noncoding RNA (lncRNA) interactions and networks were determined for the selected genes using integrated bioinformatics analyses. Moreover, the potential deleterious SNPs in the selected genes were identified and finally, their validation and implications on the structure of proteins were investigated by specific bioinformatics tools. The results of this study introduced UBE2I, RNF111, MYBL2, and VEGFA as novel factors that may involve in odontogenesis. It was also found that the *MSX2*, *SHH*, and *TFAP2A* are targeted by hsa-miR-6775-5p, hsa-miR-149-3p, and hsa-miR-432-5p, respectively. Moreover, the hsa-miR-134-5p regulated the *SHH* and *TFAP2A* gene expression. LINC02035 and C3orf35 were also introduced as important lncRNAs that may involve in competitive endogenous RNA interaction with the *SHH* for binding to the hsa-miR-149-3p. Moreover, LINC00319, interacting with the hsa-miR-6775-5p, indirectly regulated the *MSX2* expression. We also identified various SNPs in the investigated genes that changed the normal structure and thus the function of their related proteins. This study, for the first time, introduces different new proteins, miRNAs, lncRNAs, and SNPs that may be important for normal odontogenesis and the pathology of oral and dental disorders.

Please cite this paper as: Naghiyan Fesharaki S, Sisakhtnezhad S. 2022. *In silico* analysis of possible novel RNA interactions and deleterious single nucleotide polymorphisms related to *MSX2*, *SHH*, *SMAD7* and *TFAP2* genes involved in odontogenesis. *J Genet Resour* 8(2): 165-177. doi: 10.22080/jgr.2022.23318.1304.

Introduction

Identification of the cellular and molecular basis of tooth formation (odontogenesis) is very important for understanding the natural development of teeth as well as the pathology of oral and dental disorders and thus their treatment. Different signaling pathways and genes regulate the odontogenesis process. For example, Fibroblast growth factor (FGF), Sonic hedgehog (SHH), Wingless-related integration site (WNT), Ectodysplasin (EDA), and Bone morphogenetic protein (BMP) signaling pathways are some of the most important signaling molecules that regulate the developmental process of teeth by the induction of the odontogenic progenitor cells (Oshima & Tsuji, 2015). In this regard, exclusive epithelial activation of the WNT signaling at the dental

lamina stage initiates multiple tooth buds and results in supernumerary teeth (Balic, 2019). Mutation in the *Wnt* genes (*e.g.,* WNT10A) leads to hypodontia (missing teeth) (Thesleff, 2018). BMP4 expression is confined to tooth germ in the last bud and cap stages. Msh homeobox 1 (*MSX1*) and *MSX2* genes are expressed as soon as dental lamina is initiated and their expression continues until the end of tooth formation, but in different areas (Babajko *et al.,* 2014). MSX1 can favorably control BMP4 expression and is essential for the initiation of tooth formation and morphogenesis. Also, MSX2 as a direct downstream target of SMAD-mediated BMP signaling is important for molar root development (Berdal *et al.,* 2009; Yuan & Chai, 2019). In addition, Mothers against decapentaplegic homolog 7 (SMAD7) is expressed at a high level in the dental epithelium but a moderate to weak level in the dental mesenchyme in growing mouse molars. Functionally, the SMAD7 is critical for the regulation of tooth size (Liu *et al.,* 2019). In addition to the WNT and BMP signaling pathways, the SHH signaling pathway is also crucial in the odontogenesis process. The SHH signaling is important for the epithelial tissue's localized proliferation and invaginating into the underlying mesenchyme to create a tooth bud and initiate tooth development. Moreover, SHH has a significant role in the regulation of enamel formation and the development of tooth roots (Hosoya *et al.,* 2020; Li *et al.,* 2015). It has also been found that SHH signaling is downstream of WNT signaling and both WNT and SHH signaling are mediated by epithelial Islet expression (Li *et al.,* 2017; Hermans *et al.,* 2021). In addition to the above-mentioned genes, the activator protein-2 (AP-2) transcription factors, including transcription factors AP-2 alpha and beta (TFAP2A and TFAP2B), are essential components in the regulation of tooth development and shape. Loss of the epithelial domain of the TFAP2A and TFAP2B changes the quantity and spatial arrangement of the incisors, resulting in duplicate lower incisors, according to tissue-specific deletions. However, the deletion of these two genes in the mesenchymal domain does not affect tooth formation (Woodruff *et al.,* 2021).

Single nucleotide polymorphisms (SNPs) are the most common type of genetic variation among people. Nowadays, SNPs may help predict an individual's response to certain drugs, susceptibility to environmental factors such as toxins, and risk of developing particular diseases. In this regard, nonsynonymous SNPs (nsSNPs), which cause an amino acid change in the corresponding protein product, are of particular relevance since they account for roughly half of all the known human hereditary genetic defects (Akhoundi *et al.,* 2016; Pandey *et al.,* 2020).

Identification of the SNPs related to the hub genes involved in the odontogenesis process is important to know the pathology of oral and dental diseases. Moreover, non-coding RNAs (ncRNAs) are a large segment of the cell transcriptome that are not translated into proteins. However, they have been verified to play critical roles in diverse physiological and pathological processes (Sun & Chen, 2020). MicroRNAs (miRNAs) and long non-coding RNAs (lncRNAs) are two classes of non-coding RNAs that play important roles in fine-tuning gene expression during physiological and pathological conditions. Identifying miRNAs and lncRNAs regulating the genes involved in tooth development as well as their cross-talking is important in understanding the cellular and molecular basis of odontogenesis under physiological and pathological conditions (Wang *et al.,* 2021a; Wang *et al.,* 2021b; Yao *et al.,* 2022). Collectively, according to the literature review of previous studies, *MSX2*, *SHH*, *SMAD7*, and *TFAP2A* are important genes in the odontogenesis process. Despite the identification of some molecular factors involved in tooth development, there may still be other factors influencing this process. Moreover, given the importance of miRNAs and lncRNAs for regulating the genes involved in tooth development, this study aimed to identify possible novel factors, miRNAs and lncRNAs, and also their cross-talk with the *MSX2*, *SHH*, *SMAD7,* and *TFAP2A* genes by *in silico* data analysis. We also attempted to identify the potential deleterious SNPs in the coding and untranslated regions of the *MSX2*, *SHH*, *SMAD7,* and *TFAP2A* genes that influence their protein structures and functions. In addition, the effect of

the detected SNPs in 3′-UTR was evaluated on the interaction of miRNAs with mRNAs of the target genes.

Materials and Methods

Gene selection, pathway enrichment, and gene ontology analysis

According to the literature review of previous studies, *MSX2*, *SHH*, *SMAD7*, and *TFAP2A* were selected as important genes involved in the odontogenesis process. Pathway enrichment for the *MSX2*, *SHH*, *SMAD7*, and *TFAP2A* genes was performed by Reactome (https://reactome.org/ PathwayBrowser/; Griss *et al.,* 2020) and KEGG (https://www.genome.jp /kegg/pathway.html; Kanehisa *et al.,* 2017) online databases. Moreover, gene ontology analysis was carried out by enrichr database (https://maayanlab.cloud/Enrichr/; Xie *et al.,* 2021) to identify the biological processes, molecular functions, and cellular components that are associated with the selected genes.

Protein-protein and RNAs interaction

The direct and indirect protein-protein interactions (PPIs) of the MSX2, SHH, SMAD7, and TFAP2A with each other and other proteins were evaluated by the STRING (Ver 11.5) database (https://string-db.org/cgi/input?sessionId=b IOnO6OxkLkQ&input_page_show_search=off), based on the experimental repositories, genomic context, co-expression, and public text collections (Jensen *et al.,* 2009). The obtained PPIs with high confidence interactions (CIs) $\geq$ 0.7 and P-values less than 0.05 ($P < 0.05$) were selected. In addition, miRNA-mRNA and lncRNA-miRNA interactions were determined by miRWalk (http://mirwalk.umm.uni-heidelberg.de/; Sticht *et al.,* 2018) and ENCORI/Starbase v2 (https://bio.tools/starbase; Dashti *et al.,* 2020) databases, respectively. The complex interactions between the RNAs were visualized by Cytoscape Software (Otasek *et al.,* 2019).

SNPs analysis

In this study, the possible SNPs in the human *MSX2*, *SHH*, *SMAD7,* and *TFAP2A* genes were obtained from the SNP database of NCBI (dbSNP: https://www.ncbi.nlm.nih.gov/snp/?term=; Bhagwat, 2010).

Functional context of the missense mutations

The functional analysis of non-synonymous nsSNPs was performed by sorting intolerant from tolerant (SIFT) (https://sift.bii.a-star.edu.sg/; Kumar *et al.,* 2009) and protein variation effect analyzer (PROVEAN) (http://provean.jcvi.org/index.php; Venselaar *et al.,* 2010) online bioinformatics tools. SIFT can tell if an amino acid alteration in a protein will be harmful or not. When the value is less than or equal to 0.05, the amino acid substitution is expected to be harmful, and when the score is more than 0.05, it is projected to be tolerable. PROVEAN can predict any sort of protein sequence change, including amino acid substitutions, in-frame insertions, deletions, etc.

Biophysical validation of nsSNPs

HOPE web server (https://www3.cmbi.umcn.nl/hope/) was performed to identify the effects of mutations on the proteins' 3D structures and their functions. HOPE uses WHAT IF for structural calculations, YASARA and HSSP for conservation scores, DAS-server for sequence-based predictions, and UniProtKB for sequence annotations.

The input option is a FASTA format sequence of the whole protein and the mutation of interest via the web interface; the output is based on the structural differences between the mutant and wild-type residues (Venselaar *et al.,* 2010). HOPE produces a report, completed with results, figures, and animations. It is built from small pieces of text combined into a complete story. The report shows whether a structure was known, a model was built or predictions were made. This is followed by the effect of the mutation, illustrated by figures (in case a structure is available) and animations.

The data is combined with the known properties of the wild-type and mutant residues and their influence on the hydrogen bonds and/or disturb the correct folding of proteins.

miRNASNP-v3 database for the SNPs in the miRNAs binding sites

In the present study, miRNASNP–v3 database (http://bioinfo.life.hust.edu.cn/miRNASNP/) was

used to identify the top polymorphisms on the 3′-untranslated region (3′-UTR) of the *MSX2*, *SHH*, *SMAD7*, and *TFAP2A* genes. The output indicated novel SNPs leading to higher (gain) or lower (loss) binding affinity to specific microRNAs (Liu *et al.,* 2021).

Results

Gene ontology and functional analysis

Pathway enrichment analysis of the KEGG database revealed that the SHH protein is involved in the hedgehog signaling pathway, axon guidance, and pathways in cancer. KEGG also introduced SMAD7 as a protein involved in Hippo and Transforming growth factor-beta (TGF-β) signaling pathways. Moreover, Reactome pathway analysis showed that MSX2 protein is contributed to the generic transcription pathway by inhibiting Runt-related transcription factor 2 (RUNX2) (Fig. 1). Gene ontology analysis by Enrichr demonstrated that the *TFAP2A* and *SHH* gene products influence biological processes, including embryonic limb morphogenesis, kidney development, and regulation of cell differentiation (Supplement 1). In addition, the TFAP2A, SHH, and MSX2 proteins regulate the cell differentiation process and are also involved in the negative regulation of transcription by RNA polymerase II. Among the various molecular functions, TFAP2A and MSX2 regulate the following functions: transcription regulatory region nucleic acid binding, transcription cis-regulatory region binding, as well as double-stranded, and sequence-specific DNA binding. Moreover, the SHH protein is a morphogen and can bind to the patched cell surface receptor. Furthermore, the cellular component analysis of the Enrichr database indicated that SHH is located in membrane raft, endoplasmic reticulum lumen, and intracellular organelle lumen, and it is also a sequence-specific DNA binding protein. Moreover, the TFAP2A and MSX2 proteins are located in the nucleus and intracellular membrane-bounded organelle

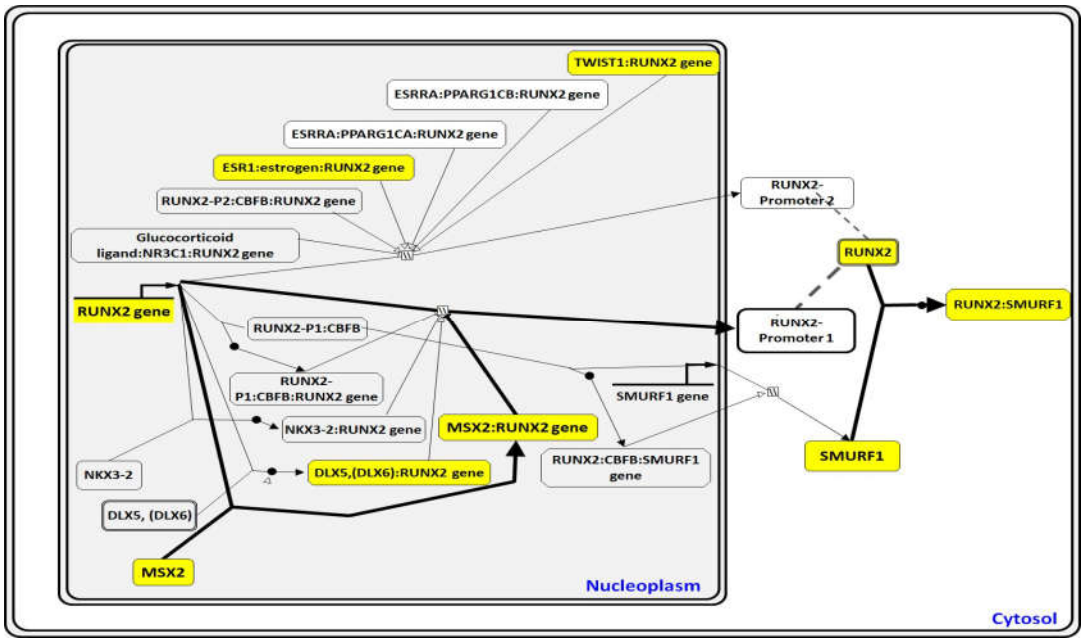

Fig. 1. Pathway enrichment analysis by Reactome revealed that the expression of *RUNX2* is controlled by the MSX2 in the generic transcription signaling pathway. The homeobox transcription factor MSX2 can bind to Distal-less homeobox 5 (*DLX5*) sites in the promoter of *RUNX2* and inhibit transcription of *RUNX2*. The results also indicated that several transcription factors (including estrogen receptor alpha (ESR1), Estrogen- related receptor alpha (ERRA), Glucocorticoid receptor (NR3C1), Twist family bHLH transcription factor 1 (TWIST1), CBF-beta (CBFB), NKX3-2 (BAPX1), and SMAD specific E3 ubiquitin-protein ligase 1/2 (SMURF1/2)) have been implicated in the regulation of the *RUNX2* gene transcription. Factors we identified for following *in silico* studies are highlighted in yellow color on the figure.

mRNA-miRNA-lncRNA interactions

Evaluating the interactions between the miRNAs and mRNAs of the *MSX2*, *TFAP2A*, *SHH*, and *SMAD7* by miRWalk demonstrated a novel miRNA targetome for these genes (Supplement 2). In this analysis, the binding probability and interaction with the seed region are the criteria for selecting top novel microRNAs. After filtering, the top 20 miRNAs for each gene were selected. The results indicated that hsa-miR-6775-5p, hsa-miR-149-3p, and hsa-miR-432-5p are the miRNAs with the lowest binding energy for *MSX2*, *SHH*, and *TFAP2A* mRNAs, respectively. Therefore, these miRNAs were selected for analyzing the lncRNA-miRNA interaction by ENCORI (Fig. 2). Based on the mRNA-miRNA-lncRNA interaction network (competitive endogenous RNA (ceRNA) interaction) analysis, LINC02035 and C3orf35 have novel competition with the *SHH* gene for binding to the hsa-miR-149-3p. Also, it has been found that there is a similar interaction between the LINC00319 and the mRNA of the *MSX2* for binding to the hsa-miR-6775-5p.

Protein-protein interactions analysis

The PPIs analysis by STRING (CIs ≥ 0.7 and P < 0.05) revealed that there are no direct interactions between the MSX2, SMAD7, SHH, and TFAP2A proteins. Nonetheless, it showed that MSX2 interacts with the RUNX2, CCAAT/enhancer-binding protein alpha (CEBPA), MSX1, TWIST1, and DLX5 (Fig. 3A). Our results also demonstrated that the SMAD7 has protein interactions with SMURF1/2, TGF-beta receptor type-1 (TGF-βR1), and Neural precursor cell expressed developmentally down-regulated 4-like (NEDD4L), and Ring finger protein 111 (RNF111) proteins (Fig. 3B). Moreover, SHH has interactions with Hedge- hog-interacting protein (HHIP), Growth arrest-specific protein 1 (GAS1), Protein patched homolog 1/2 (PTCH1/2), and Cell adhesion molecule-related/down-regulated by oncogenes (CDON) (Fig. 3C). The results of STRING also indicated that Ubiquitin-conjugating enzyme E2 I (UBE2I), Estrogen receptor 1 (ESR1), Potassium channel tetramerization domain containing 15 (KCTD15), Paired-like homeodomain transcription factor 2 (PITX2), Erb-B2 receptor tyrosine kinase 2 (ERBB2), Myb proto-oncogene like 2 (MYBL2), Vascular endothelial growth factor A (VEGFA), Histone acetyltransferase p300 (EP300), and CBP/p300-interacting transactivator 2 (CITED2) are the possible partners for interaction with the TFAP2A protein (Fig. 3D).

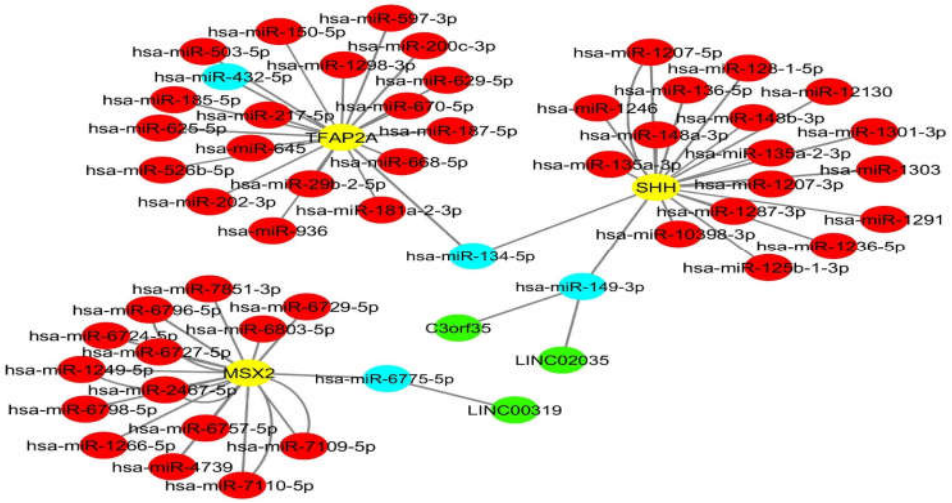

Fig. 2. The competitive endogenous RNA network of *MSX2*, *SHH*, and *TFAP2A*, based on the analysis of the miRNA-mRNA interaction in miRWalk and the miRNA-lncRNA interaction in ENCORI. The red nodes indicate the miRNAs, the green nodes indicate the lncRNAs, and the yellow nodes indicate the protein-coding genes in this study. The width of edges is based on the number of bindings. Visualizing the complex interactions between RNAs was performed by Cytoscape software (V3.7.1).

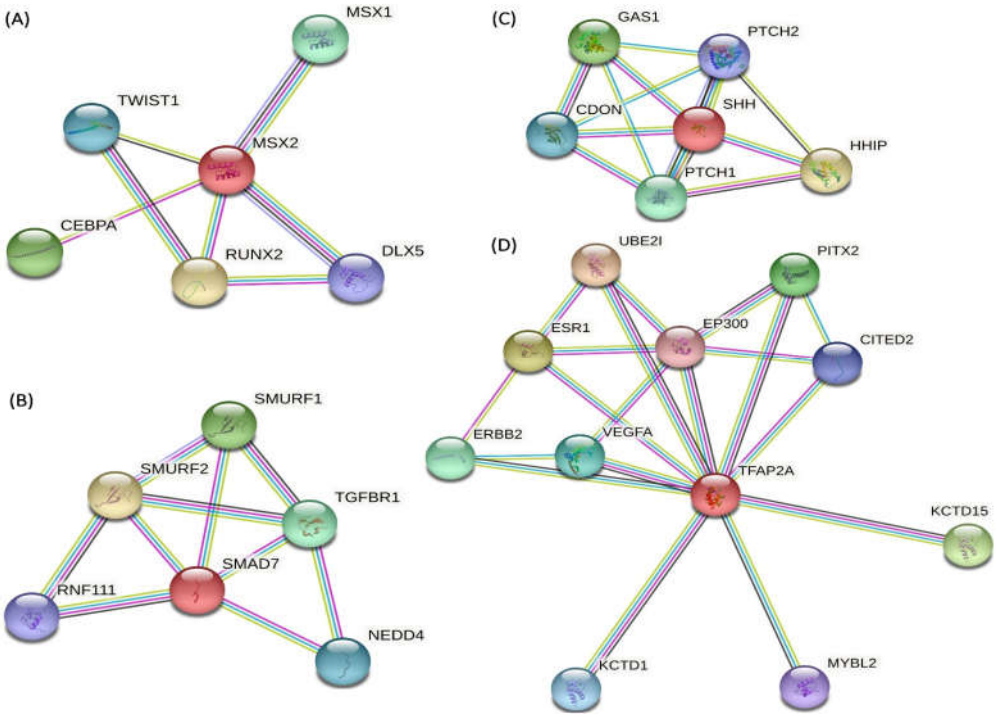

Fig. 3. Protein-protein interaction analysis by STRING database. The PPIs of the MSX2 (a), SMAD7 (b), SHH (c), and TFAP2A (d) with other proteins. By the results of the pathway analysis by Reactome, the results of STRING also show that MSX2 interacts with RUNX2, DLX5, and TWIST1. Furthermore, the SMAD7 and TFAP2A can interact with the SMURF2 and EST1 proteins, respectively. Moreover, the results indicate that there is no direct interaction between the MSX2, SMAD7, SHH, and TFAP2A proteins.

Prediction of the deleterious SNPs in the coding region of genes

In this study, the presence of the possible SNPs in the *MSX2*, *TFAP2A*, *SHH*, and *SMAD7* genes were investigated by the dbSNP database of NCBI. After evaluating the possible SNPs for the investigated genes, the SNPs in the coding region of the genes were selected for subsequent investigations to identify the deleterious polymorphisms. According to the results of the data analysis by SIFT and PROVEAN, two deleterious polymorphisms (rs199732800, rs199856192) were detected from 11 identified nsSNPs in the *MSX2* gene (Supplement 3 and Table 1). Moreover, our results introduced two new deleterious polymorphisms (rs267605193, rs200991750) in the *SMAD7* gene as well as one deleterious SNP (rs267607047) in the *SHH* gene. In this regard, three deleterious SNPs, including rs144275164, rs9350373, and rs143258135, were detected in the *TFAP2A* gene. The deleterious and tolerated SNPs related to the *MSX2*, *TFAP2A*, *SHH,* and *SMAD7* genes are presented in Table S3.

Table 1. List of nsSNPs analysis by SIFT and PROVEAN online bioinformatics tools

Genes	SNP	Ref/Alt	Amino acid change	SIFT score	PROVEAN score
MSX2	rs199732800	G/T	R199I	0	-7.514
	rs199856192	G/A	M263I	0.014	-2.963
SMAD7	rs200991750	T/G	T408P	0.003	-2.653
	rs267605193	C/T	G318D	0.047	-6.046
SHH	rs267607047	G/T	N115K	0	-5.269
TFAP2A	rs9350373	G/C	R356G	0.004	-5.999
	rs144275164	T/G	E311A	0.036	-5.534
	rs143258135	C/A	G20V	0.029	-3.859

Biophysical validation of nsSNPs

After identifying the deleterious SNPs for each gene, the HOPE webserver was used to unveil the 3D structure of the mutated proteins and describe the candidate reactions and physicochemical qualities. Our results for the SNPs of the *MSX2* gene indicated that the rs199732800 converts arginine to isoleucine at position 199 (R199I) (Fig. 4A) and thus makes a mutant residue, which is more hydrophobic than the wild type, and this change can cause the loss of hydrogen bonds and disturb the correct folding. In addition, the smaller size of isoleucine may lead to the loss of the interactions of the MSX2 with other proteins or molecules. Moreover, the rs199856192 that substitutes methionine with isoleucine at position 263 (M263I) of the MSX2 protein, provides a minor residue and therefore can lead to the loss of its interactions with other proteins or molecules. The results of the HOPE webserver for the *SMAD7* gene indicated that rs200991750 converts threonine to proline residue at position 408 (T408P) in the SMAD7 protein (Fig. 4B). The proline residue is more hydrophobic than threonine, and thus this amino acid substitution disturbs the correct folding by reducing the hydrogen bonds. In addition, our results for the SNPs of the *SHH* gene demonstrated that rs267607047 substitutes asparagine with a lysine residue at position 115 (N115K) in the SHH protein (Fig. 4C). This mutation introduces a bigger residue on the surface of the SHH protein and therefore, this SNP can disturb the SHH interaction with other molecules or proteins. Furthermore, biophysical validation of the nsSNPs for the *TFAP2A* gene showed that the rs9350373, which substitutes arginine to glycine residue at position 356 (R356G), introduces a more minor residue in the SNP region of TFAP2A and might lead to the loss of its interactions with other proteins or molecules (Fig. 4D).

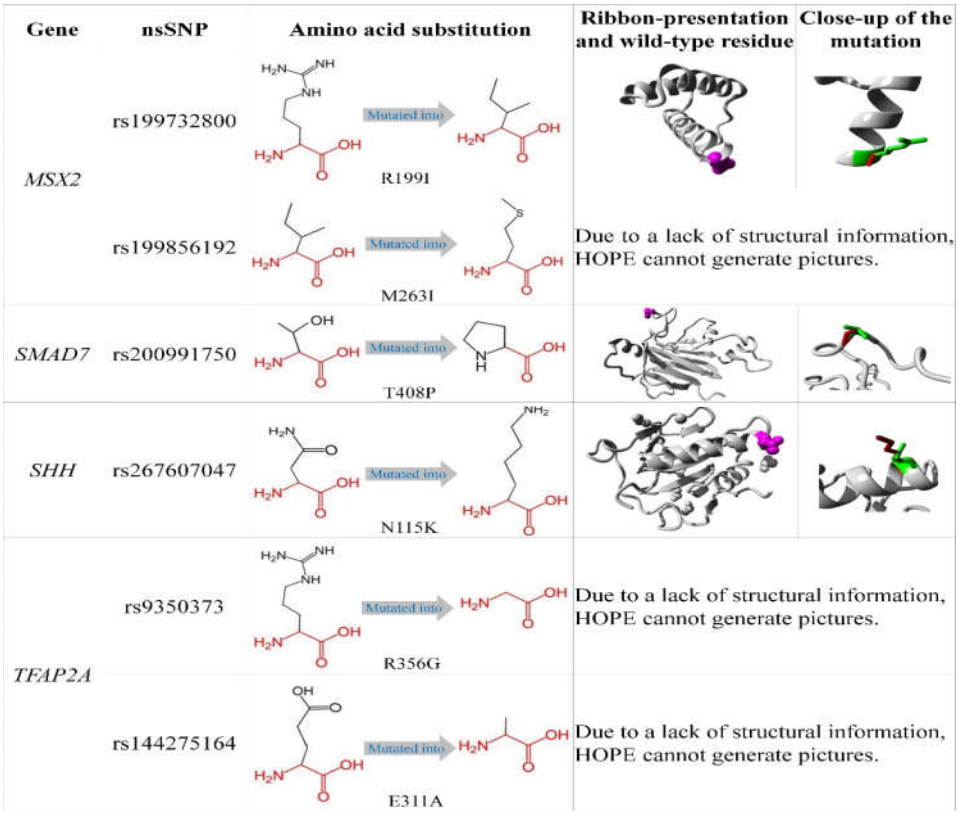

Fig. 4. Biophysical validation of the nsSNPs related to the *MSX2*, *SMAD7*, *SHH,* and *TFAP2A* by HOPE webserver. The HOPE results demonstrated the effects of the identified snSNPs on the amino acid sequence of the MSX2, SMAD7, SHH, and TFAP2A proteins and their 3D structures and functions. See the text and Table 2 for more details.

Also, the mutant residue is more hydrophobic, and thus this SNP can result in the loss of hydrogen bonds or disturb the correct folding. The results of HOPE also revealed that the rs144275164 SNP substitutes glutamic acid with alanine at position 311 (E311A) in the TFAP2A protein. This mutated residue is more hydrophobic than the wild type (glutamic acid) and thus may influence the hydrogen bonds and in this way disturb the correct folding of the TFAP2A protein. The effects of amino acid substitution mediated by the detected nsSNPs are presented in Table 2.

Table 2. The effects of amino acid substitution mediated by the nsSNPs on the properties of the *MSX2*, *SMAD7*, *SHH*, and *TFAP2A* proteins

Protein	AAS*	Effects
MSX2	R199I	•The wild-type residue charge is positive and the mutant residue charge is neutral, therefore mutation of the residue can cause loss of interactions with other molecules or residues. •The mutated residue is located in a highly conserved domain that is important for binding other molecules, therefore mutation of the residue might disturb this function. •The mutant residue is smaller which might lead to loss of interactions. •The mutation introduces a more hydrophobic residue at this position. This can result in loss of hydrogen bonds and/or disturb correct folding.
	M263I	•The wild-type and mutant amino acids differ in size. •The mutant residue is smaller which might lead to loss of interactions.
SMAD7	T408P	•The mutated residue is located near a highly conserved domain that is important for binding other molecules, therefore mutation of the residue might disturb this function. •The mutation introduces a more hydrophobic residue at this position. This can result in loss of hydrogen bonds and/or disturb correct folding.
SHH	N115K	•The wild-type residue charge is neutral and the mutant residue charge is positive. The mutation introduces a charge at this position; this can cause repulsion between the mutant residue and neighboring residues. •The mutant residue is bigger than the wild-type residue. •The residue is located on the surface of the protein. Mutation of this residue can disturb interactions with other molecules or other parts of the protein.
TFAP2A	R356G	•The wild-type residue charge is positive and the mutant residue charge is neutral, therefore mutation of the residue can cause loss of interactions with other molecules or residues. •The mutated residue is located in a domain that is important for the main activity of the protein, therefore mutation of the residue might disturb this function. •The mutant residue is smaller which might lead to loss of interactions. •The mutation introduces a more hydrophobic residue at this position. This can result in loss of hydrogen bonds and/or disturb correct folding.
	E311A	•The wild-type residue charge is negative and the mutant residue charge is neutral, therefore mutation of the residue can cause loss of interactions with other molecules or residues. •The mutated residue is located in a domain that is important for the main activity of the protein, therefore mutation of the residue might disturb this function. •The mutant residue is smaller which might lead to loss of interactions. •The mutation introduces a more hydrophobic residue at this position. This can result in loss of hydrogen bonds and/or disturb correct folding.

*AAS= Amino acid substitution

SNPs analysis in the 3′-UTR region of genes

The SNP analysis by the miRNASNP-v3 webserver revealed that the rs763019404 in the 3′-UTR region of the *MSX2* gene could lead to a higher binding affinity to eight miRNAs (Table 3). Based on the energy of binding, hsa-miR-2276-5p establishes the most robust binding to the mRNA of the *MSX2*, as compared to the other seven gain miRNAs. In addition, the results of the SNP analysis in the 3′-UTR region of the *SHH* gene by miRNASNP-v3 demonstrated that the rs1166368389 SNP causes ten gain miRNAs. According to the energy of binding, hsa-miR-199a-3p and hsa-miR-199b-3p show the most robust binding to the 3′-UTR region of the *SHH* mRNA.

Discussion

Pathway enrichment and gene ontology analysis indicated that the *MSX2*, *SHH*, *SMAD7*, and *TFAP2A* genes are important for different signaling pathways and biological processes. However, there was no information about the role of these genes in odontogenesis in the Reactome, KEGG, and Enrichr databases. This may be due to the lack of data on the role of these genes in tooth development

in these databases. Nonetheless, different studies have demonstrated the central role of the *MSX2*, *SHH*, *SMAD7*, and *TFAP2A* genes in the odontogenesis process (Berdal *et al.,* 2009; Babajko *et al.,* 2014; Li *et al.,* 2015; Liu *et al.,* 2019; Yuan and Chai, 2019; Hosoya *et al.,* 2020; Woodruff *et al.,* 2021).

Table 3. Deleterious nsSNPs and gained miRNAs in the 3′-UTR of *MSX2* and *SHH* genes

Gene	SNP	Ref/Alt	Gain	Loss	miRNA	ΔG binding (kCal/mol)
MSX2	rs763019404	A/G	8	3	hsa-miR-146a-3p	-19.45
					has-miR-301a-3p	-12.27
					hsa-miR-2276-5p	-20.63
					hsa-miR-6892-3p	-13.2
					has-miR-4653-3p	-12.21
					hsa-miR-3921	-15.52
					hsa-miR-4766-5p	-12.76
					hsa-miR-4766-5p	-13.34
SHH	rs1166368389	A/G	10	3	hsa-miR-101-3p	-9.44
					hsa-miR-199a-3p	-18.21
					hsa-miR-199b-3p	-18.21
					hsa-miR-411-5p	-9.4
					hsa-miR-144-3p	-9.31
					hsa-miR-212-3p	-8.9
					hsa-miR-4732-5p	-18.18
					hsa-miR-3129-5p	-15.67
					hsa-miR-6832-5p	-14.39
					hsa-miR-936	-16.03

The PPIs analysis showed that the MSX2, SHH, SMAD7, and TFAP2A interact with different proteins. MSX2 can interact with the RUNX2, CEBPA, MSX1, TWIST1, and DLX5 proteins and previous studies have confirmed the role of these proteins in odontogenesis (Levi *et al.,* 2006; Huang *et al.,* 2013; Meng *et al.,* 2015). The SMAD7 can interact with the SMURF1/2, TGF-βR1, NEDD4L, and RNF111 proteins. The roles of the TGF-βR1, SMURF1/2, and NEDD4L proteins have been shown in tooth development in different studies (Lee *et al.,* 2011). Although our results indicated that the RNF111 may interact with the SMAD7, there is no report regarding its role in odontogenesis. The present study also demonstrated that the SHH interacts with the HHIP, GAS1, PTCH1/2, and CDON proteins (Seppala *et al.,* 2017). All of these proteins are also involved in the tooth development process. In addition, the TFAP2A may interact with the UBE2I, ESR1, KCTD15, PITX2, ERBB2, MYBL2, VEGFA, EP300, and CITED2. Although the importance of ESR1, KCTD15, PITX2, ERBB2, P300, and CITED2 is confirmed in tooth development (Jiménez-Farfán *et al.,* 2005; Chen *et al.,* 2012; Heffer *et al.,* 2017; Yu *et al.,* 2020; Cunha *et al.,* 2021), there is no report regarding UBE2I, MYBL2, and VEGFA. Therefore, the findings of this study for the first time suggest that UBE2I, RNF111, MYBL2, and VEGFA may involve in the odontogenesis process. However, experimental studies are required to confirm the exact role of these proteins in tooth development.

In this study, we also identified that miRNAs and lncRNAs may regulate the expression level of the *MSX2*, *SHH*, *SMAD7,* and *TFAP2A* genes. The integrated bioinformatics analysis introduced hsa-miR-6775-5p, hsa-miR-149-3p, and hsa-miR-432-5p that target *MSX2*, *SHH*, and *TFAP2A* mRNAs, respectively. Although there is no report on the role of hsa-miR-6775-5p and hsa-miR-149-3p in tooth development, Huang *et al.* found that hsa-miR-432 may be a potential miRNA that targets the dentin sialo-phospho-protein (DSPP) 3′-UTR (Huang *et al.,* 2011). Our results also indicated that hsa-miR-134-5p regulates both *SHH* and *TFAP2A* mRNAs expression, which are important for the development of tooth roots and spatial arrangement of the tooth germs (Li *et al.,* 2017; Woodruff *et al.,* 2021). Given the importance of the *SHH* and *TFAP2A* genes in odontogenesis, we suggest that abnormality in the expression or interaction of the miR-134-5p with the mRNA may disturb the

development of tooth roots and the spatial arrangement of teeth. Although previous studies revealed the role of miR-134-5p in parathyroid tumor development (Wang *et al.,* 2021c), Alzheimer's disease (Baby *et al.,* 2020), and vascular dementia (Kobayashi *et al.,* 2017), this is the first study that reports the possible role of hsa-miR-134-5p in tooth development.

Our investigation revealed that hsa-miR-149-3p could regulate the expression of the *SHH* gene. In this regard, it has also been found that LINC02035 and C3orf35 lncRNAs prevent the inhibitory function of hsa-miR-149-3p on the *SHH* gene expression. An emerging role of lncRNAs is that they compete for binding to miRNAs, acting as a sponge to regulate the gene activity (Zhang *et al.,* 2019). Therefore, we suggest that these two identified lncRNAs may indirectly regulate the expression of the *SHH* gene and thus the development of tooth by binding to hsa-miR-149-3p. Despite studying the possible role of hsa-miR-149-3p in some pathological conditions such as renal cell carcinoma (Xiao *et al.,* 2020), myocardial infarction (Nong *et al.,* 2021), and colorectal cancer (Liao *et al.,* 2019), its function has not been evaluated in the normal development of tooth as well as oral and dental disorder. Thus, our *in silico* study provides a good basis for the experimental investigation of the possible role of miR-149-3p in the physiological and pathological process of odontogenesis.

In the present study, we discovered that hsa-miR-6775-5p regulates the expression of the *MSX2* gene. The MSX2 protein is crucial for the development of molar roots (Berdal *et al.,* 2009; Yuan and Chai, 2019). Therefore, hsa-miR-6775-5p, through regulating the *MSX2* gene expression, may affect the development of molar roots. In addition, abnormal changes in binding affinity and expression of this miRNA may lead to shorter molar roots. Our results also elucidate that the function of hsa-miR-6775-5p may be regulated through interactions with LINC00319. Therefore, we propose that down-regulation of this lncRNA and suppression of its interaction with the hsa-miR-6775-5p may induce the formation of shorter molar roots. In general, this study is the first report on the importance of has-miR-6775-5p for tooth development, and thus further experimental studies are required to confirm it.

Based on the SNP analysis, rs199732800 (R199I) and rs199856192 (M263I) may change the structure of the MSX2 protein in a harmful way leading to the abnormal development of molar roots. The rs200991750 (T408P) and rs267605193 (G318D) SNPs are the two potential deleterious polymorphisms in the *SMAD7* gene. These two SNPs may significantly disturb tooth size and interrupt tooth development. The rs267607047 (N115K) polymorphism in the *SHH* gene is a novel and significant SNP that may disturb the SHH signaling and interrupt the dental epithelium growth and the development of tooth roots. In addition, rs9350373 (R356G), rs144275164 (E311A), and rs143258135 (G20V) in the *TFAP2A* gene are the potential deleterious SNPs that probably change the spatial arrangement of the incisors. Functionally, our results indicated that some of these SNPs could change the typical structure of related proteins in different ways – including changing the surface, hydrophobicity, and the number of hydrogen bonds – that can suppress the regular function of the related proteins. In general, all of the SNPs identified in this study are novel based on our literature review, and there has been no previous experimental research on these polymorphisms.

In conclusion, this *in silico* study introduces a set of possible novel SNPs and protein-protein, mRNA-miRNA, lncRNA-miRNA, and SNP-miRNA interactions that may influence the expression of the *MSX2*, *SMAD7*, *SHH*, and *TFAP2A* genes, which are crucial for the odontogenesis process. Indeed, our study provides a basis for conducting experimental research to identify novel molecular factors and cross-talking between them that affect tooth development during physiological and pathological conditions.

Conflicts of interest

The authors declared no conflicts of interest.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

Akhoundi F, Parvaneh N, Modjtaba EB. 2016. In silico analysis of deleterious single nucleotide polymorphisms in human

BUB1 mitotic checkpoint serine/threonine kinase B gene. *Meta Gene* 9: 142-150.

Babajko S, de La Dure-Molla M, Jedeon K, Berdal A. 2014. MSX2 in ameloblast cell fate and activity. *Front Physiol* 5: 510. doi: 10.3389/fphys.2014.00510.

Baby N, Alagappan N, Dheen ST, Sajikumar S. 2020. MicroRNA-134-5p inhibition rescues long-term plasticity and synaptic tagging/capture in an Aβ (1-42)-induced model of Alzheimer's disease. *Aging Cell* 19(1): e13046.

Balic A. 2019. Concise review: cellular and molecular mechanisms regulation of tooth initiation. *Stem Cells* 37(1): 26-32.

Berdal A, Molla M, Hotton D, Aïoub M, Lézot F, Néfussi JR, Goubin G. 2009. Differential impact of MSX1 and MSX2 homeogenes on mouse maxillofacial skeleton. *Cells Tissues Organs* 189(1-4): 126-132.

Bhagwat M. 2010. Searching NCBI's dbSNP database. *Curr Protoc Bioinform* Chapter 1: Unit 1.19.

Chen Z, Gao B, Zhou X. 2012. Expression patterns of histone acetyltransferases p300 and CBP during murine tooth development. *In Vitro Cell Dev Biol Anim* 48(1): 61-68.

Cunha AS, Vertuan Dos Santos L, Schaffer Pugsley Baratto S, Abbasoglu Z, Gerber JT, Paza A, Küchler EC. 2021. Human permanent tooth sizes are associated with genes encoding oestrogen receptors. *J Orthod* 48(1): 24-32.

Dashti S, Taheri M, Ghafouri-Fard S. 2020. An in-silico method leads to recognition of hub genes and crucial pathways in survival of patients with breast cancer. *Sci Rep* 10(1): 1-13.

Griss J, Viteri G, Sidiropoulos K, Nguyen V, Fabregat A, Hermjakob H. 2020. ReactomeGSA - efficient multi-omics comparative pathway analysis. *Mol Cell Proteomics* 19(12): 2115-2125.

Heffer A, Marquart GD, Aquilina-Beck A, Saleem N, Burgess HA, Dawid IB. 2017. Generation and characterization of Kctd15 mutations in zebrafish. *PLoS One* 12(12): e0189162.

Hermans F, Hemeryck L, Lambrichts I, Bronckaers A, Vankelecom H. 2021. Intertwined signaling pathways governing tooth development: a give-and-take between canonical Wnt and Shh. *Front Cell Dev Biol* 9: 758203. doi: 10.3389/fcell.2021.758203.

Hosoya A, Shalehin N, Takebe H, Shimo T, Irie K. 2020. Sonic Hedgehog signaling and tooth development. *Int J Mol Sci* 21(5): 1578. doi: 10.3390/ijms21051587

Huang B, Takahashi K, Sakata-Goto T, Kiso H, Togo Y, Saito K, Bessho K. 2013. Phenotypes of CCAAT/enhancer-binding protein beta deficiency: hyperdontia and elongated coronoid process. *Oral Dis* 19(2): 144-150.

Huang X, Xu S, Gao J, Liu F, Yue J, Chen T, Wu B. 2011. miRNA expression profiling identifies DSPP regulators in cultured dental pulp cells. *Int J Mol Med* 28(4): 659-667.

Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, von Mering C. 2009. STRING 8-a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* 37(Database issue): D412-D416.

Jiménez-Farfán D, Guevara J, Zenteno E, Malagón H, Hernández-Guerrero JC. 2005. EGF-R and erbB-2 in murine tooth development after ethanol exposure. *Birth Defects Res A Clin Mol Teratol* 73(2): 65-71.

Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 45(D1): D353-D361.

Kobayashi T, Tomofuji T, Machida T, Yoneda T, Ekuni D, Azuma T, Morita M. 2017. Expression of salivary miR-203a-3p was related with oral health-related quality of life in healthy volunteers. *Int J Mol Sci* 18(6): 1263. doi: 10.3390/ijms18061263.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 4(7): 1073-1081.

Lee DS, Yoon WJ, Cho ES, Kim HJ, Gronostajski RM, Cho MI, Park JC. 2011. Crosstalk between nuclear factor I-C and transforming growth factor-β1 signaling regulates odontoblast differentiation and homeostasis. *PLoS One* 6(12): e29160.

Levi G, Mantero S, Barbieri O, Cantatore D, Paleari L, Beverdam A, Merlo GR. 2006. Msx1 and Dlx5 act independently in development of craniofacial skeleton, but converge on the regulation of Bmp

signaling in palate formation. *Mech Dev* 123(1): 3-16.

Li J, Feng J, Liu Y, Ho TV, Grimes W, Ho HA. Chai Y. 2015. BMP-SHH signaling network controls epithelial stem cell fate via regulation of its niche in the developing tooth. *Dev Cell* 33(2): 125-135.

Li J, Parada C, Chai Y. 2017. Cellular and molecular mechanisms of tooth root development. *Development* 144(3): 374-384.

Liao C, Huang X, Gong Y, Lin Q. 2019. Discovery of core genes in colorectal cancer by weighted gene co-expression network analysis. *Oncol Lett* 18(3):3137-3149.

Liu CJ, Fu X, Xia M, Zhang Q, Gu Z, Guo AY. 2021. miRNASNP-v3: a comprehensive database for SNPs and disease-related variations in miRNAs and miRNA targets. *Nucleic Acids Res* 49(D1): D1276-D1281.

Liu Z, Chen T, Bai D, Tian W, Chen Y. 2019. Smad7 regulates dental epithelial proliferation during tooth development. *J Dent Res* 98(12):1376-1385.

Meng T, Huang Y, Wang S, Zhang H, Dechow PC, Wang X, Lu Y. 2015. Twist1 is essential for tooth morphogenesis and odontoblast differentiation. *J Biol Chem* 290(49):29593-29602.

Nong Y, Guo Y, Gumpert A, Li Q, Tomlin A, Zhu X, Bolli R. 2021. Single dose of synthetic microRNA-199a or microRNA-149 mimic does not improve cardiac function in a murine model of myocardial infarction. *Mol Cell Biochem* 476(11):4093-4106.

Oshima M, Tsuji T. 2015. Whole tooth regeneration as a future dental treatment. *Adv Exp Med Biol* 881:255-269.

Otasek D, Morris JH, Bouças J, Pico AR, Demchak B. 2019. Cytoscape Automation: empowering workflow-based network analysis. *Genome Biol* 20: 185. https://doi.org/10.1186/s13059-019-1758-4.

Pandey S, Dhusia K, Katara P, Singh S, Gautam B. 2020. An in silico analysis of deleterious single nucleotide polymorphisms and molecular dynamics simulation of disease linked mutations in genes responsible for neurodegenerative disorder. *J Biomol Struct Dyn* 38(14): 4259-4272.

Seppala M, Fraser GJ, Birjandi AA, Xavier GM, Cobourne MT. 2017. Sonic Hedgehog signaling and development of the dentition. *J Dev Biol* 5(2): 6. doi:10.3390/jdb5020006.

Sticht C, De La Torre C, Parveen A, Gretz N. 2018. miRWalk: An online resource for prediction of microRNA binding sites. *PLoS One* 13(10): e0206239.

Sun YM, Chen YQ. 2020. Principles and innovative technologies for decrypting noncoding RNAs: from discovery and functional prediction to clinical application. *J Hematol Oncol* 13(1): 1-27.

Thesleff I. 2018. From understanding tooth development to bioengineering of teeth. *Eur J Oral Sci* 126 : 67-71.

Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. 2010. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinform* 11(1): 1-10.

Wang C, Dong L, Wang Y. Jiang Z, Zhang J, Yang G. 2021a. Bioinformatics analysis identified miR-584-5p and key miRNA-mRNA networks involved in the osteogenic differentiation of human periodontal ligament stem cells. *Front Genet* 12: 750827. doi: 10.3389/fgene.2021.750827.

Wang F, Tao R, Zhao L, Hao XH, Zou Y, Lin Q, Chen S. 2021b. Differential lncRNA/mRNA expression profiling and functional network analyses in bmp2 deletion of mouse dental papilla cells. *Front Genet* 12: 702540. doi: 10.3389/fgene.2021.702540.

Wang J, Wang Q, Zhao T, Liu X, Bai G, Xin Y, Wei B. 2021c. Expression profile of serum-related exosomal miRNAs from parathyroid tumor. *Endocrine* 72(1): 239-248.

Woodruff ED, Gutierrez GC, Van Otterloo E, Williams T, Cohn MJ. 2021. Anomalous incisor morphology indicates tissue-specific roles for Tfap2a and Tfap2b in tooth development. *Dev Biol* 472: 67-74.

Xiao CT, Lai WJ, Zhu WA, Wang H. 2020. MicroRNA derived from circulating exosomes as noninvasive biomarkers for diagnosing renal cell carcinoma. *Onco Targets Ther* 13: 10765. doi: 10.2147/OTT.S271606.

Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, Ma'ayan A. 2021. Gene set knowledge discovery with Enrichr. *Curr Protoc* 1(3): e90.

Yao B, Cheng X, Mei X, Qiu J, Zhang B, Wang J, Xiao M. 2022. Profiling long noncoding RNA alterations during the stromal cell-derived factor-1α-induced odontogenic differentiation of human dental pulp stem cells. *Arch Oral Biol* 137: 105393. doi: 10.1016/j.archoralbio.2022.105393.

Yu W, Sun Z, Sweat Y, Sweat M, Venugopalan SR, Eliason S, Amendt BA. 2020. Pitx2-Sox2-Lef1 interactions specify progenitor oral/dental epithelial cell signaling centers. *Development* 147(11): dev186023. doi: 10.1242/dev.186023.

Yuan Y, Chai Y. 2019. Regulatory mechanisms of jaw bone and tooth development. *Curr Top Dev Biol* 133: 91-118.

Zhang X, Wang W, Zhu W, Dong J, Cheng Y, Yin Z, Shen F. 2019. Mechanisms and functions of long non-coding RNAs at multiple regulatory levels. *Int J Mol Sci* 20(22): 5573. doi: 10.3390/ijms20225573.