Novel Genetic Determinants of Dental Maturation in Children

Olja Grgic[1,2], Vid Prijatelj[1,2], Amel Dudakovic[3], Strahinja Vucic, Brunilda Dhamo[1,2], Katerina Trajanoska[4,5], Claire Monnereau, Masa Zrimsek[6], Kaare M. Gautvik[7], Sjur Reppe[7], Emi Shimizu[8], Simon Haworth[9,10,11], Nicholas J. Timpson[9,11], Vincent W. V. Jaddoe[2], Marjo-Riitta Jarvelin[12,13], David Evans[11,14,15], Andre G. Uitterlinden[2], Edwin M. Ongkosuwito[16], Andre J. van Wijnen[17], Carolina Medina-Gomez[2], Fernando Rivadeneira[1,2*], Eppo B. Wolvius[1,2*]

[1]ErasmusMC, Rotterdam, The Netherland, Department of Oral and Maxillofacial Surgery. [2]ErasmusMC, Rotterdam, The Netherland, The Generation R Study. [3]Mayo Clinic, Rochester, Minnesota, USA, Department of Orthopedic Surgery. [4]McGill University, Montréal, Québec, Canada, Department of Human Genetics. [5]McGill University, Montréal, Québec, Canada, Canada Excellence Research Chair in Genomic Medicine. [6]Medical University of Vienna, Vienna, Austria, Department of Pathology. [7]Oslo University Hospital, Oslo, Norway, Department of Medical Biochemistry. [8]Rutgers School of Dental Medicine, New Jersey, USA, Department of Oral Biology. [9]University of Bristol, Bristol, UK, Bristol Medical School, Department of Population Health Sciences. [10]University of Bristol, Bristol, UK, Bristol Dental School. [11]University of Bristol, Bristol, UK, MRC Integrative Epidemiology Unit, [12]University of Oulu, Finland, Faculty of Medicine, Center for Life Course Health Research. [13]Imperial College, London, UK, Faculty of Medicine, School of Public health. [14]The University of Queensland, Brisbane, Australia, Diamantina Institute. [15]The University of Queensland, Brisbane, Australia, Institute for Molecular Bioscience. [16]Radboud University Nijmegen Medical Center, Dentistry, Section Orthodontics and Craniofacial Biology. [17]University of Vermont, Burlington, Vermont, USA, Department of Biochemistry.

ABSTRACT

Dental occlusion requires harmonious development of teeth, jaws, and other elements of the craniofacial complex, which are regulated by environmental and genetic factors. We performed the first genome-wide association study (GWAS) on dental development (DD) using the Demirjian radiographic method. Radiographic assessments from participants of the Generation R Study (primary study population, N_1= 2,793; mean age 9.8 years) were correlated with ~30 million genetic variants while adjusting for age, sex, and genomic principal components (proxy for population stratification). Variants associated with DD at genome-wide significant (GWS) level ($P<5x10^{-8}$) mapped to 16q12.2 (*IRX5*) (lead variant rs3922616, B=0.16; $P=2.2x10^{-8}$). We used Fisher's combined probability tests weighted by sample size to perform a meta-analysis (N=14,805) combining radiographic DD at mean age 9.8 years from Generation R with data from a previous GWAS (N_2=12,012) on number of teeth (NT) in infants used as proxy of DD at mean age 9.8 years (including the ALSPAC and NFBC1966). This GWAS meta-analysis revealed three novel loci mapping to: 7p15.3 (*IGF2BP3*: $P=3.2x10^{-8}$), 14q13.3 (*PAX9*: $P=1.9x10^{-8}$) and 16q12.2 (*IRX5*: $P=1.2x10^{-9}$), and validated eight previously reported NT loci. A polygenic allele score constructed from these 11 loci was associated with radiographic DD in an independent Generation R set of children (N=703; B=0.05, P=0.004). Furthermore, profiling of the identified genes across an atlas of murine and human stem cells observed expression in the cells involved in the formation of bone and/or dental tissues (>0.3 Frequency Per Kilobase of transcript per Million mapped reads), likely reflecting functional specialization. Our findings provide biological insight into the polygenic architecture of the pediatric dental maturation process.

Abstract word count: 261

Total word count: 3,200

Total number of tables/figures: 5

Number of references: 39

Keywords: odontogenesis, tooth calcification, tooth eruption, GWAS, gene expression, child health

***These authors contributed equally**

INTRODUCTION

Dental development (DD) is a biological process which begins in the eighth week of gestation with primary teeth formation and completes postnatally around the age of 18-25 years (Nanci 2008). Tooth structures' formation, eruption, and emergence are intertwined processes and integral components of human tooth maturation (Nanci 2008). Delayed DD typically results in inadequate dental occlusion, and consequent problems with mastication, pronunciation, and appearance (Dhamo et al. 2018).

DD is a multifactorial process under distinct environmental and genetic influences. Tooth eruption is influenced by multiple factors such as trauma, surgery, nutrition, and medication (Suri et al. 2004; Poureslami et al. 2015). Furthermore, secular acceleration in tooth (root) formation most likely reflects changes in environmental influences over time (lessening of disease, better nutrition, decrease in energy expenditure resulting in high BMI) (Cardoso et al. 2010). Conversely, various gene families have also been associated with DD (FGF, WNT, BMP) mostly in animal models (Frazier-Bowers and Vora 2017). Previous genome-wide association study (GWAS) meta-analysis studying tooth eruption identified 10 loci associated with the age of the first erupted tooth (AFTE) and 11 loci with the number of teeth at one year of age (NT) (Fatemifar et al. 2013). Demirjian and colleagues developed a radiographic method to estimate DD based on dental mineralization, shape, and proportions (Demirijan et al. 1973) which has never been surveyed following a GWAS approach.

We performed the first GWAS on radiographic DD and strengthened our results by combining them with a previously performed meta-analysis studying NT at 12 or 15 months of age (proxies of DD) (Fatemifar et al. 2013). We also constructed a polygenic allele score (PGS) for DD and tested its association with radiographic DD in independent samples. In addition, we

sought supporting evidence for the functional involvement of the discovered candidate gene(s) by investigating the local expression profiles in murine and human cell lines.

METHODS

Study Population

The present study comprised participants from the Generation R Study (primary study population), and two other cohorts – Avon Longitudinal Study of Parents and Children study (ALSPAC) and 1966 Northern Finland Birth Cohort Study (NFBC1966). Details about study populations and ethical approvals are provided in the *Supplementary Material*.

Dental Development Assessment

The *Generation R Study* quantified DD at a mean age of 9.8 (SD=0.34) years (third cohort visit) on panoramic radiographs (exclusively available in the Generation R Study) using the method described by Demirjian (1973) (see *Supplementary Material*).

ALSPAC provided primary tooth eruption phenotypes of children based on questionnaires completed by the mothers and included both AFTE and NT (at 15 months) (Fatemifar et al. 2013).

NFBC1966 measured AFTE and NT in primary teeth based on the judgment of public health professionals during monthly visits of children to welfare centers. AFTE was recorded as the month of visit at which the first tooth was observed. Hence, the first tooth may have erupted at any time between the end of the previous month and the recorded month (i.e., 'interval censoring'). NT was recorded at 12 months (Fatemifar et al. 2013).

Genotyping and Quality Control

Individuals were genotyped using Illumina HumanHap 610 or 660 Quad chips (Illumina Inc., San Diego, USA) in the Generation R Study (Medina-Gomez et al. 2015), Illumina

HumanHap550 Quad genotyping platform (23andMe, Cambridge, UK; Burlington, NC, USA) in ALSPAC, and Illumina HumanCNV370-Duo DNA Analysis BeadChip in NFBC1966. Details about genotyping and quality control are provided in the *Supplementary Material*.

Gene expression profiling

Candidate genes were prioritized based on expression profiles across relevant cells and tissues. These profiles were selected from a very large atlas of RNA-seq data obtained by performing both RNA-sequencing and bioinformatic analysis on >1,000 murine and human RNA samples using the same experimental procedures and in silico pipeline to facilitate broad comparisons among many cell types and tissues (AJVW and AD; unpublished data available on request). Murine RNA-seq data used for the current analysis was obtained for ameloblasts and osteoblasts, while human RNA-seq data was obtained with bone marrow stromal cells (BMSCs), adipose tissue-derived mesenchymal stem cells (AMSCs), dental pulp derived stromal cells (DPSCs), and pluripotent embryonic stem cells (ESCs). Based on the biological characteristics of these cell lines, gene(s) that directly contribute to DD were expected to be endogenously expressed in one or more mesenchymal cell types. Conversely, absence of expression would be a potential disqualification for direct mechanistic contributions to DD regardless of a GWS level of association. Hence, local expression provided supporting data for the prioritization of the best candidate(s) genes per locus. We also considered the evidence for mechanistic involvement of specific genes in bone and tooth formation and mineralization based on the available literature. Details about cell origin and RNA-sequencing analysis are given in the *Supplementary Material.* We evaluated expression for all the genes mapping within +/- 400 KB of the index-associated SNP in the novel loci. Furthermore, for all the loci previously associated with NT (Fatemifar et

al. 2013) and GWS in the joint meta-analysis, the expression of the gene closest to the index SNP was evaluated.

GWAS in the Generation R participants

In the Generation R Study – the primary study population (N_1=2,793), standardized residuals of radiographic DD were calculated using a linear regression model adjusted for age, sex, and 20 genomic principal components (to adjust for population stratification) (Price et al. 2006). Residuals were further used as an outcome in the association analysis with ~ 30 million SNPs using the mach2qtl software package as implemented in GRIMP (Estrada et al. 2009). Genome-wide significance (GWS) was set at $P<5x10^{-8}$.

Meta-Analysis

Single GWS SNP(s) look-up was pursued in an already published GWAS meta-analysis comprising two cohorts (ALSPAC and NFBC1966) (N_2=12,012) (Fatemifar et al. 2013) using proxy phenotypes of DD – NT and AFTE –. In addition, Fisher's combined probability test weighted by sample size and implemented in METAL (Willer et al. 2010) was used in the joint meta-analysis of the discovery cohort in which radiographic DD at mean age 9.8 was assessed and the published results of ALSPAC and NFBC 1966 in which conversely NT (proxy of DD at mean age 9.8 years) was evaluated.

Genetic correlation

Genetic correlation was evaluated between the summary statistics of the NT-GWAS and anthropometric traits, blood lipids, hormones, and glycemic traits using LD-Score regression as implemented in the LD Hub database, making use of the already reported epidemiological association of these traits with dental maturation/eruption (Lal et al. 2008; Partyka et al. 2018; Danze et al. 2021). We used only the NT GWAS summary statistics to avoid combining two

proxies of dental maturation – NT and radiographic DD - in the LD Hub database (Zheng et al. 2016). Heritability estimate for NT was obtained from the LD Hub output.

PGS

Association between PGS (created from loci reported as associated at GWS level with combined dental traits) and radiographic DD (expressed in years) was tested in an independent subset of children of European ancestry in Generation R using a linear regression model. Details about the study population can be found in the *Supplementary Material*.

Gene expression

High throughput next-generation sequencing of RNA and read mapping were performed as discussed previously (Dudakovic et al. 2014; Kalari et al. 2014). Frequency Per Kilobase of transcript per Million mapped reads (FPKM) expression values were standardized by Log2 (x+1) transformation. Log transformation was performed to account for exponential differences in expression values, while one was added to retain zero-valued observations in the dataset. A threshold of detectable expression was set at 0.3 FPKM as described previously (Ramsköld et al. 2009). Differential gene expression was defined as a change greater than 1.5 fold (Van Oudenhove et al. 2016). Fold change in the expression across two cell lines was derived from the ratio of the mean expression values across cell lines. The non-parametric Mann-Whitney test was used to assess the differences in gene expression across cell lines. A significance threshold was set at $P<0.05$ for the analysis of expression in murine cells (2 cell lines) and at $P<0.017$ (0.05/3) for the analysis in human cells after correction for multiple testing (4 cell lines).

RESULTS

The characteristics of the discovery sample are presented in **Appendix Table 1**. The discovery GWAS, showing an adequate control for population stratification (genomic inflation

factor λ: 0.98), identified variants in the 16q12.2 locus associated at GWS level with advanced DD (*IRX5*; lead variant rs3922616-C, B=0.16; P=2.2x10^{-8}; explaining 1% of the variance of radiographic DD) (**Appendix Figure 2**). The SNP rs12444195, found to be in high LD with rs3922616 (r^2=0.87), was used for the cross-trait replication, as the lead variant was not present in the NT meta-analysis set. This variant showed significant evidence of cross-trait replication in the NT and the AFTE meta-analyses (P=2.7x10^{-5} and P=1.5x10^{-5}, respectively). In the genome-wide cross-trait meta-analysis between the radiographic DD discovery and the NT sets, including 2,216,657 SNPs common to all three studies (N=14,805), there was no evidence suggesting population stratification (λ=1.007) (**Figure 1**). Furthermore, we identified variants associated with DD in children in three novel loci, including the one locus (*IRX5*) that was already reported in the discovery set. These new loci are 16q12.2 (*IRX5*; P=1.2x10^{-9}), 7p15.3 (*IGF2BP3*; P=3.2x10^{-8}) and 14q13.3 (*PAX9;* P=1.9x10^{-8}) (**Table 1, Figure 1, Figure 2**). Given that mutations in *PAX9* cause tooth agenesis (Online Mendelian Inheritance in Man dataset; https://www.omim.org/, *167416) (Hamosh et al. 2005), we subsequently adjusted our analysis for the presence of hypodontia in the discovery set where this variable was available. This adjustment did not substantially change the findings of the discovery set or the joint meta-analysis. Furthermore, variants in two loci (1q32.1; *ASCL5* and 14q24.1; *RAD51B*) reached GWS in the combined meta-analysis (**Figure 1**). However, these genes have already been implicated in primary and permanent tooth eruption (Pillas et al. 2010; Geller et al. 2011). Lastly, from the 10 loci reported previously as associated at GWS level with NT (excluding X chromosome), 8 were GWS after the joint meta-analysis (**Appendix Table 2**). However, only 3 of these 8 loci were nominally (P<0.05) associated with DD in the discovery set (*ADK*, *HMGA2*, and *MSRB3*) (**Appendix Table 2**).

NT SNP heritability was estimated at 24.8% (95% CI 24.70 - 24.90). Borderline significant genetic correlation of NT with diverse lipid traits was observed while there was no evidence of genetic correlation with hormones, glycemic and anthropometric traits (**Appendix Table 3**, **Appendix Figure 3**).

There was a significant positive association between the PGS (comprising 11 SNPs from joint effort; 3 novel and 8 replicated) and radiographic DD in the independent European subsample of children (B: 0.05 years, P=0.004) (**Appendix Figure 4**).

We also evaluated the expression of 28 genes from the GWS loci in human-derived cells and 20 genes in mouse-derived cells (8 genes did not have orthologs in the mouse genome) after the exclusion of three recently described genes (*FAM221A*, *NKX2-1-AS1*, and *MIR4503*) that were not yet annotated in the specific transcriptome scaffold used for the queried RNA-seq database.

Heatmap visualization of gene expression across mouse osteoblasts and ameloblasts showed two clusters: one comprising genes with higher expression levels in ameloblasts and the other comprising genes predominantly expressed in osteoblasts (**Figure 3**). However, none of the differences in expression levels of the genes across the two cell lines reached significance ($P \geq 0.1$). Nevertheless, *GPNMB* (7p15.3) showed exclusive expression in murine ameloblasts. More importantly, the DD candidate genes were clearly expressed in either mesenchymal cell type, consistent with some type of endogenous function in either early or late stages of craniofacial and DD.

Heatmap visualization of gene expression across human cell lines also indicated two clusters: one comprising genes with higher expression levels in DPSCs and the other predominantly expressed in ESCs (**Figure 4**). However, after a detailed assessment of gene

expression patterns in the novel loci, the only genes showing a clear difference in expression levels across the four cell lines (expression fold change>1.5, P<0.017) were *GPNMB* (7p15.3; expression in DPSCs close to the one in BMDCs [P=0.69], 2-fold higher than in AMSCs [P=0.008], and close to the detection threshold in ESCs), *IGF2BP3* (7p15.3; expression in ESCs 3-fold higher than in AMSCs [P=0.008] and close to the detection threshold in DPSCs and BMDCs), *PAX9* (exclusively expressed in DPSCs) and *SLC25A21* (expressed in DPSCs and ESCs, with no significant difference in expression between the two cell lines [P=0.1]). Despite the differences in the expression levels between cell types, the fact that mRNAs for each of the genes is expressed at detectable level in distinct cell types suggests that these genes could directly contribute to growth, differentiation or function of mesenchymal cells.

Expression profiles of the genes previously associated with the NT (Fatemifar et al. 2013) and GWS in the joint effort are also presented in both heatmap visualizations. Likewise, expression profiles of all the genes across mouse and human cells are presented in **Appendix Figure 5** and **Appendix Figure 6,** respectively.

DISCUSSION

Radiographic DD constitutes a refined approach to assess the genetic architecture of dental maturation. In the joint meta-analysis, we report 3 novel loci associated with combined dental traits mapping to the vicinity of the *IRX5, IGF2BP3,* and *PAX9* genes, beyond confirming 8 (of 10) loci previously associated with dental eruption. Also, we demonstrated dental maturity PGS as associated with radiographic DD in an independent subsample of children. Moreover, the analysis of gene expression profiles across murine and human cell lines provided additional evidence for the functional involvement of the reported genes in DD.

Earlier dental maturation in girls as compared to boys, likely governed by earlier pubertal onset in girls has been reported (Demirjian and Levesque 1980). In line with this contention, at the mean age of 10 years, in the Generation R Study, girls showed on average more advanced DD as compared to boys (data not shown). Therefore, our GWAS studies (Fatemifar et al. 2013) have implemented sex-adjusted statistical models.

IRX5 may be driving the association with DD at the 16q12.2 locus. The *IRX5* expression profile links it to tooth development and bone metabolism (**Appendix Figures 5** and **6**). Previously, it has been shown that *IRX5* impacts craniofacial development (Bonnard et al. 2012), bone development (Li et al. 2014) and the development of the upper jaw in mice (Jeong et al. 2008). Yet, *CRNDE,* showing a similar expression pattern in human cells as *IRX5*, could also potentially be the gene underlying the reported association. In mice, *Crnde* has been postulated as a regulator of bone metabolism (Mulati et al. 2020). Interestingly, *CRNDE* and *IRX5* share the same promoter which could suggest a functional connection of the genes (Yang et al. 2008). *IRX6* expression was below/close to the detection limit in both mouse and human cell lines and therefore is less likely to be underlying the association with DD in this locus.

In the 14q13.3 locus, *PAX9* is the strongest candidate gene to underlie the reported association with DD. In line with the implication of *PAX9* in disturbances in the number (Sun et al. 2020) and size of teeth (Lee et al. 2012), our profiling indicates an enriched expression of this gene in cell lines of dental origin in both mouse and humans. However, *SLC25A21* can not be fully discarded as a possible gene underlying the reported association. The ablation of *Slc25a21* (14q13.3) in mice results in orofacial abnormalities. This finding is consistent with the exclusive expression of *SLC25A21* in human DPSCs and ESCs. Nevertheless, the effect ascribed to *SLC25A21* could also be explained by a change in the expression of the neighboring *PAX9*

(Maguire et al. 2014). Lastly, *MBIP* shows a broad expression across all cell lines and therefore may represent a ubiquitous gene that could contribute to DD. However, there is a paucity of data on its involvement in the development of mineralizing tissues. The other genes in the locus, exhibit expression values close to the detection threshold and are less likely to play a role in DD regulation.

Information arising from gene expression profiling was not conclusive at the 7p15.3 locus, as multiple genes showed broad expression patterns across many/all cell types. Yet, the *GPNMB* and *IGF2BP3* emerge as attractive candidates deserving further experimental consideration. Exclusive expression of *GPNMB* in murine ameloblasts and expression in human DPSCs (and BMDCs) suggest that this gene may be involved in dental and/or bone phenotypes, consistent with previous results in mouse models (Frara et al. 2016). There is also evidence of *GPNMB* as a negative regulator of miR-508-5p during the odontogenesis of DPSCs (Liu et al. 2019; Zhou et al. 2020). However, the index SNP in this locus maps closer to *IGF2BP3*. This gene shows the largest differential expression in ESCs as compared to the other cell lines. *IGF2BP3* encodes a protein that regulates insulin-like growth factor II (Igf2) translation during late embryonic development in mice and humans (Nielsen et al. 1999). Overexpression of Igf2 is associated with Beckwith–Wiedemann syndrome characterized by abnormal growth of different skull parts (Hamosh et al. 2005).

The association between dental maturation PGS and radiographic DD suggests that genes reported here may have multiple functions in both dental eruption and mineralization, and indicates that these two processes are interdependent components of overall dental maturation.

Generation R is to our knowledge the only large pediatric cohort with dental panoramic radiographs available. This represents a great opportunity to investigate DD but also poses the

limitation of the lack of replication necessary to allow extrapolation of our findings. Hence, in this particular study, we chose proxy phenotypes. Such a decision relied on the successful replication of our discovery findings in the NT meta-analysis and a previously reported negative partial correlation between DD and tooth eruption (ranging from 0.60-0.80 after adjustment for BMI; $P<0.001$) (Nicholas et al. 2018). However, combining proxy traits of DD (radiographic DD and NT at 12 or 15 months) in the current work may have decreased our statistical power to detect associations confined to one of the phenotypes. Another limitation of our study is the low number of RNA samples obtained both from human (N=5) and murine (N=3) cell lines which could result in low statistical power or false-positive findings. Lastly, it is important to know that the association signal in the discovery sample explained only 1% of the variance in radiographic DD. Based on a twin study, it was reported that 2/3 of the variation in the mineralization of root of third molars can be attributed to genetic effects, while 1/3 can be attributed to environmental components (Trakinienė et al. 2019). In other words, environmental factors (such as nutrition, different diseases, and BMI) have an important impact on DD. While beyond the scope of this manuscript, the study of these factors and their possible interaction with genetic markers, should be addressed in future research.

In conclusion, in the first GWAS meta-analysis including radiographic DD, we identified 3 novel loci affecting dental maturity. The association of dental maturity PGS with radiographic DD and cross-trait replication indicated that radiographic DD and tooth eruption are reliable proxies for each other, and may thus serve as mutually corroborating parameters for the general assessment of dental maturation in genetic studies. Taken together, our study provides new biological insights into dental maturation by showing that DD is a complex polygenic process.

Among genes mapping to the identified loci, the *IRX5*, *PAX9*, *GPNMB* and *IGF2BP3* have been shown as functionally involved in the normal development of mineralized tissues.

Authors' Contributions: Olja Grgic contributed to the conception of this work, performed data analysis, interpreted them, and wrote the manuscript. Vid Prijatelj gave his suggestions when performing polygenic score analyses, interpreting them, and critically revising the manuscript. Strahinja Vucic contributed to the conception of this work, statistical analyses, interpretation of the findings, and was involved in manuscript creation. Brunilda Dhamo assessed dental development for the whole Generation R and critically revised the manuscript. Katerina Trajanoska gave her suggestions when performing GWAS meta-analysis and critically revised the manuscript. Claire Monnereau, Masha Zrimsek, Kaare M. Gautvik, Sjur Reppe, Amel Dudakovic, and Emi Shimizu helped with the interpretation of the findings and critically revised the manuscript. Simon Haworth and Nicholas J. Timpson contributed to the conception of this work, were involved in joining the efforts/data, and critically revised the manuscript. Vincent W. V. Jaddoe had a major role in the conception and development of the Generation R Study and critically revised the manuscript. Andre G. Uitterlinden gave his suggestions at the initiation of this project, was involved in genotyping the generation R Study, and critically revised the manuscript. Marjo-Riitta Jarvelin and David Evans contributed to the conception of this work, were involved in joining the efforts/data, and critically revised the manuscript. Andre J. van Wijnen acquired and shared unpublished data for the RNA-seq atlas of murine and human stem cells, interpreted the findings, critically revised the manuscript and provided creative suggestions for the presentation of the work. To generate the RNA-seq atlas, Andre J. van Wijnen and Amel Dudakovic secured funding, IRB and IACUC approvals where appropriate, as well as ensured

patient consenting, rapid sample collection, expert sample processing, RNA-seq read acquisition, bioinformatics, curation of the data sets, and subsequent data selection and interpretation of samples relevant to the current study. Carolina Medina-Gomez and Edwin M. Ongkosuwito jointly supervised every step of this workflow and supported it with their suggestions and comments. Lastly, they critically revised the manuscript. Fernando Rivadeneira and Eppo B. Wolvius substantially and jointly contributed to the conception of this work, interpretation of findings, supervised this work at every stage, and in detail revised this manuscript. All authors gave their final approval and agree to be accountable for all aspects of the work.

Acknowledgments: The Generation R acknowledges the contribution of children and parents, general practitioners, hospitals, midwives, and pharmacies in Rotterdam. The generation and management of GWAS genotype data for the Generation R Study were done at the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, The Netherlands. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera, and Marjolein Peters for their help in creating, managing, and QC of the GWAS database. The general design of Generation R Study is made possible by financial support from the Erasmus Medical Center, Rotterdam, the Erasmus University Rotterdam, the Netherlands Organization for Health Research and Development (ZonMw), the Netherlands Organisation for Scientific Research (NWO), the Ministry of Health, Welfare and Sport and the Ministry of Youth and Families. Additionally, the Netherlands Organization for Health Research and Development supported C.M-G and F.R of this manuscript (ZonMw VIDI 016.136.367 to F.R. and C. M-G.), while the National Institutes of Health provided support for A.J.v.W (R01 AR075803). ALSPAC is extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical

workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and Wellcome (217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. Lastly, N.J.T of this manuscript is a Wellcome Trust Investigator (202802/Z/16/Z), is the PI of the Avon Longitudinal Study of Parents and Children (MRC & WT 217065/Z/19/Z), is supported by the University of Bristol NIHR Biomedical Research Centre (BRC-1215-2001), the MRC Integrative Epidemiology Unit (MC_UU_00011/1) and works within the CRUK Integrative Cancer Epidemiology Programme (C18281/A29019). This publication is the work of the authors and the authors will serve as guarantors for the contents of this paper. There are no conflicts of interest to be stated.

REFERENCES

Bonnard C, Strobl AC, Shboul M, Lee H, Merriman B, Nelson SF, Ababneh OH, Uz E, Guran T, Kayserili H, et al. 2012. Mutations in IRX5 impair craniofacial development and germ cell migration via SDF1. Nat Genet. 44(6):709–713.

Cain CJ, Gaborit N, Lwin W, Barruet E, Ho S, Bonnard C, Hamamy H, Shboul M, Reversade B, Kayserili H, et al. 2016. Loss of Iroquois homeobox transcription factors 3 and 5 in osteoblasts disrupts cranial mineralization. Bone Reports. 5:86–95.

Cardoso HFV, Heuzé Y, Júlio P. 2010. Secular change in the timing of dental root maturation in Portuguese boys and girls. Am J Hum Biol. 22(6):791–800.

Danze A, Jacox LA, Bocklage C, Whitley J, Moss K, Hardigan P, Garcia-Godoy CE, Jackson TH. 2021. Influence of BMI percentile on craniofacial morphology and development in children and adolescents. Eur J Orthod. 43(2):184–192.

Demirijan A, Goldstein H, Tanner JM. 1973. A new system of dental age assessment. Hum Biol.

45(2):211–227.

Demirjian A, Levesque G. 1980. Sexual differences in dental development and prediction of emergence. J Dent Res. 59(7):1110–1122.

Dhamo B, Elezi B, Kragt L, Wolvius EB, Ongkosuwito EM. 2018. Does dental caries affect dental development in children and adolescents? Bosn J Basic Med Sci. 18(2):198–205.

Dudakovic A, Camilleri E, Riester SM, Lewallen EA, Kvasha S, Chen X, Radel DJ, Anderson JM, Nair AA, Evans JM, et al. 2014. High-resolution molecular validation of self-renewal and spontaneous differentiation in clinical-grade adipose-tissue derived human mesenchymal stem cells. J Cell Biochem. 115(10):1816–1828.

Estrada K, Abuseiris A, Grosveld FG, Uitterlinden AG, Knoch TA, Rivadeneira F. 2009. GRIMP: A web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data. Bioinformatics. 25(20):2750–2752.

Fatemifar G, Hoggart CJ, Paternoster L, Kemp JP, Prokopenko I, Horikoshi M, Wright VJ, Tobias JH, Richmond S, Zhurov AI, et al. 2013. Genome-wide association study of primary tooth eruption identifies pleiotropic loci associated with height and craniofacial distances. Hum Mol Genet. 22(18):3807–3817.

Frara N, Abdelmagid SM, Sondag GR, Moussa FM, Yingling VR, Owen TA, Popoff SN, Barbe MF, Safadi FF. 2016. Transgenic expression of osteoactivin/gpnmb enhances bone formation in vivo and osteoprogenitor differentiation ex vivo. J Cell Physiol. 231(1):72–83.

Frazier-Bowers SA, Vora SR. 2017. Genetic disorders of dental development: tales from the bony crypt. Curr Osteoporos Rep. 15(1):9–17.

Geller F, Feenstra B, Zhang H, Shaffer JR, Hansen T, Esserlind AL, Boyd HA, Nohr EA, Timpson NJ, Fatemifar G, et al. 2011. Genome-wide association study identifies four loci

associated with eruption of permanent teeth. PLoS Genet. 7(9):1–9.

Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. 2005. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 33(Database Iss.):D514–D517.

Jeong J, Li X, McEvilly RJ, Rosenfeld MG, Lufkin T, Rubenstein JLR. 2008. Dlx genes pattern mammalian jaw primordium by regulating both lower jaw-specific and upper jaw-specific genetic programs. Development. 135(17):2905-2916.

Kalari KR, Nair AA, Bhavsar JD, O'Brien DR, Davila JI, Bockol MA, Nie J, Tang X, Baheti S, Doughty JB, et al. 2014. MAP-RSeq: Mayo Analysis Pipeline for RNA sequencing. BMC Bioinformatics. 15(1):1–11.

Lal S, Cheng B, Kaplan S, Softness B, Greenberg E, Goland RS, Lalla E, Lamster IB. 2008. Accelerated tooth eruption in children with diabetes mellitus. Pediatrics. 121(5):e1139–e1143.

Lee W-C, Yamaguchi T, Watanabe C, Kawaguchi A, Takeda M, Kim Y-I, Haga S, Tomoyasu Y, Ishida H, Maki K, et al. 2012. Association of common PAX9 variants with permanent tooth size variation in non-syndromic East Asian populations. J Hum Genet. 57(10):654–659.

Li D, Sakuma R, Vakili NA, Mo R, Puviindran V, Deimling S, Zhang X, Hopyan S, Hui C. 2014. Formation of proximal and anterior limb skeleton requires early function of Irx3 and Irx5 and is negatively regulated by Shh signaling. Dev Cell. 29(2):233–240.

Liu F, Wang X, Yang Y, Hu R, Wang W, Wang Y. 2019. The suppressive effects of miR-508-5p on the odontogenic differentiation of human dental pulp stem cells by targeting glycoprotein non-metastatic melanomal protein B. Stem Cell Res Ther. 10(1):35.

Maguire S, Estabel J, Ingham N, Pearson S, Ryder E, Carragher DM, Walker N, Bussell J, Chan W-I, Keane TM, et al. 2014. Targeting of Slc25a21 is associated with orofacial defects and otitis

media due to disrupted expression of a neighbouring gene. PLoS One. 9(3):e91807.
Medina-Gomez C, Felix JF, Estrada K, Peters MJ, Herrera L, Kruithof CJ, Duijts L, Hofman A, van Duijn CM, Uitterlinden AG, et al. 2015. Challenges in conducting genome-wide association studies in highly admixed multi-ethnic populations: the Generation R Study. Eur J Epidemiol. 30(4):317–330.
Mulati M, Kobayashi Y, Takahashi A, Numata H, Saito M, Hiraoka Y, Ochi H, Sato S, Ezura Y, Yuasa M, et al. 2020. The long noncoding RNA Crnde regulates osteoblast proliferation through the Wnt/β-catenin signaling pathway in mice. Bone. 130:115076.
Nanci A. 2008. Ten Cate's oral histology: development, structure, and function. 8th ed. St. Louis (MO): Elsevier, Mosby.
Nicholas CL, Kadavy K, Holton NE, Marshall T, Richter A, Southard T. 2018. Childhood body mass index is associated with early dental development and eruption in a longitudinal sample from the Iowa Facial Growth Study. Am J Orthod Dentofac Orthop. 154(1):72–81.
Nielsen J, Christiansen J, Lykke-Andersen J, Johnsen AH, Wewer UM, Nielsen FC. 1999. A family of insulin-like growth factor II mRNA-binding proteins represses translation in late development. Mol Cell Biol. 19(2):1262–1270.
Van Oudenhove JJ, Grandy RA, Ghule PN, del Rio R, Lian JB, Stein JL, Zaidi SK, Stein GS. 2016. Lineage-specific early differentiation of human embryonic stem cells requires a G2 cell cycle pause. Stem Cells. 34(7):1765–1775.
Partyka M, Chałas R, Dunin-Wilczyńska I, Drohomyretska M, Klatka M. 2018. Influence of growth hormone therapy on selected dental and skeletal system parameters. Ann Agric Environ Med. 25(1):60–65.
Pillas D, Hoggart CJ, Evans DM, O'Reilly PF, Sipilä K, Lähdesmäki R, Millwood IY, Kaakinen

M, Netuveli G, Blane D, et al. 2010. Genome-wide association study reveals multiple loci associated with primary tooth development during infancy. PLoS Genet. 6(2):e1000856.

Poureslami H, Asl Aminabadi N, Sighari Deljavan A, Erfanparast L, Sohrabi A, Jamali Z, Ghertasi Oskouei S, Hazem K, Shirazi S. 2015. Does timing of eruption in first primary tooth correlate with that of first permanent tooth? A 9-years cohort study. J Dent Res Dent Clin Dent Prospects. 9(2):79–85.

Ramsköld D, Wang ET, Burge CB, Sandberg R. 2009. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput Biol. 5(12):e1000598.

Sun K, Yu M, Yeh I, Zhang L, Liu H, Cai T, Feng H, Liu Y, Han D. 2020. Functional study of novel PAX9 variants: The paired domain and non-syndromic oligodontia . Oral Dis. 27(6):1468-1477.

Suri L, Gadari E, Vastardis H. 2004. Delayed tooth eruption: Pathogenesis, diagnosis, and treatment. A literature review. Am J Orthod Dentofac Orthop. 126(4):432–445.

Trakinienė G, Andriuškevičiūtė I, Šalomskienė L, Vasiliauskas A, Trakinis T, Šidlauskas A. 2019. Genetic and environmental influences on third molar root mineralization. Arch Oral Biol. 98:220–225.

Willer CJ, Li Y, Abecasis GR. 2010. METAL: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 26(17):2190–2191.

Yang W, Ng P, Zhao M, Wong TKF, Yiu SM, Lau YL. 2008. Promoter-sharing by different genes in human genome - CPNE1 and RBM12 gene pair as an example. BMC Genomics. 9:456.

Zheng J, Erzurumluoglu M, Elsworth B, Howe L, Haycock P, Hemani G, Tansey K, Laurin C, St. Pourcain B, Warrington N, et al. 2016. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP

heritability and genetic correlation analysis. Bioinformatics. 33(2):272-279.

Zhou D, Gan L, Peng Y, Zhou Y, Zhou X, Wan M, Fan Y, Xu X, Zhou X, Zheng L, et al. 2020.

Epigenetic regulation of dental pulp stem cell fate. Stem Cells Int. 2020.

TABLES

Table 1. Novel loci associated with DD obtained from the joint meta-analysis. SNP – single nucleotide polymorphism, CHR – chromosome, BP – base pair, EA – effect allele, EAF – frequency of the effect allele, N – sample size, B – beta, SE – standard error, GENE – closest gene. SNPs position – given in GRCh37/hg19.

						Generation R N=2,793			ALSPAC+NFBC19 66 N=12,012			Meta N=14,805	
SNP	**CHR**	**BP**	**GENE**	**EA**	**EAF**	**B**	**SE**	**P**	**B**	**SE**	**P**	**Z**	**P**
rs12444195	16	55044333	*IRX5*	A	0.57	-0.16	0.03	$3.0x10^{-8}$	-0.10	0.02	$2.7x10^{-5}$	-6.1	$1.2x10^{-9}$
rs6967145	7	23517468	*IGF2BP3*	C	0.4	-0.07	0.02	$7.0x10^{-3}$	-0.12	0.02	$5.8x10^{-7}$	-5.5	$3.2x10^{-8}$
rs7156439	14	37100738	*PAX9*	T	0.62	-0.10	0.03	$2.3 x10^{-4}$	-0.11	0.02	$4.6x10^{-6}$	-5.6	$1.9x10^{-8}$

FIGURES

Figure 1. Manhattan plot of association statistics for DD in the joint meta-analysis (2,216,657 SNPs tested; N=14,805). Each dot represents one SNP, the x-axis is its chromosomal position-build 37 NCBI, on the y-axis the –log10 (P-value) is reported, horizontal red and yellow lines mark the GWS threshold ($P<5x10^{-8}$) and suggestive threshold ($P<1x10^{-6}$), respectively, SNPs which were GWS in previously published meta-analysis investigating NT are colored in red. Q-Q plot (in the upper left corner) is obtained from the joint meta-analysis; the plot shows the distribution of expected against observed P values, the diagonal red line represents the null distribution, each dot represents one SNP.

Figure 2. Regional plots for three novel loci associated with DD: 16q12.2 (a), 7p15.3 (b) and 14q13.3 (c). The lead variant in the locus is presented as a purple diamond and the flanking variants as circles in different colors according to the level of their linkage disequilibrium (LD) with the lead variant. The x-axis shows all the genes in the region in the window of +/- 500 Kb from the lead variant. Y-axis presents the –log10 (P-value). Regional plot was created using http://locuszoom.org.

Figure 3. **Heat-map representation of expression patterns of the genes mapping to GWS loci reported in the joint meta-analysis across ameloblasts and osteoblasts.** Expression is measured by RNA-seq analysis with values expressed as Frequency Per Kilobase of transcript per Million mapped reads (FPKM). Expression of the genes was available for 3 samples per cell line. Gene expression level is shown in the legend with the darkest blue representing the lowest expression in a certain row (gene min) and the darkest red the highest expression in that row (gene max). Gene clusters are marked with dots (Cluster 1 – black dot, Cluster2 – green dot). The cytogenetic band is presented next to each gene; * genes in the novel loci. Expression from *SLC25A21* (14q13.3), *NKX2-1* (14q13.3), *STK31* (7p15.3), and *IRX6* (16q12.2) was below the detection threshold in both cell lines.
Figure is created in https://software.broadinstitute.org/morpheus/

Figure 4. Heat-map representation of expression patterns of the genes mapping to GWS loci reported in the joint meta-analysis across 4 types of cell lines. BMSCs- bone marrow stromal cells, AMSCs- adipose tissue-derived mesenchymal stromal cells, DPSCs- dental pulp stem cells, ESCs- embryonic stem cells. Expression measured by RNA-seq analysis with values expressed as Frequency Per Kilobase of transcript per Million mapped reads (FPKM). Expression of the genes was available for 5 donors for every cell line. Gene expression level is shown in the legend with the darkest blue representing the lowest expression in a certain row (gene min) and the darkest red the highest expression in that row (gene max). Gene clusters are marked with dots (Cluster 1 – black dot, Cluster2 – green dot). The cytogenetic band is presented next to each gene; * genes in the novel loci. Expression from *IRX6* (16q12.2), *STK31* (7p15.3), *CLK2P* (7p15.3), *KLHL7-AS1* (7p15.3), *RPS2P32* (7p15.3), *NKX2-1* (14q13.3), *NKX2-8* (14q13.3), and *SFTA3* (14q13.3) was close to the detection threshold in all cell lines.
Figure is created in https://software.broadinstitute.org/morpheus/