Study of a Relative Polygenic Risk Score Assay for Common Oral Health Conditions

Steven D.P. Moore [1*], Solon Guzman [2], Priscillia Quan[1], April Kennedy[1], Mathew Gene[1], Joel Goodman[1]

[1]AI Genetics, Mississauga, Ontario, Canada

[2] Absolute Towers Dental Centre, Mississauga, Ontario, Canada

* Correspondence: Dr. Steven Moore, smoore@aigenetics.ca

Word Count: 2901

Pages: 10

Keywords: polygenic risk score, periodontal disease, dental caries, oral health, relative risk

Abstract

This single center observational study measured a relative polygenic risk score for two oral health conditions from 27 dental patients. Calculating an individual's polygenetic risk score is an emerging tool in genetics. Its potential to focus on preventative healthcare in populations has allowed it to be implemented by health systems like the UK's National Health Services (Genome UK). There are several oral health conditions that have a genetic basis including dental caries and periodontal disease. Despite good oral hygiene habits, some individuals may have an increased genetic predisposition to certain dental problems. The Canadian Dental Association reported that an estimated 2.26 million school-days are missed each year due to dental-related illness and tooth decay accounts for one-third of all day surgeries performed on children between the ages of 1 and 5. In the United States, a child is five times more likely to seek emergency room treatment for dental problems than for asthma, often because they are unable to see a dentist, are uninsured or cannot afford routine dental care. Upon review of the literature, we identified common genetic variants with evidence for association with periodontal disease and dental caries and developed a genotyping panel coupled with a relative polygenic risk score. We assessed the performance of this assay in a cohort of 27 dental clinic patients by running polygenic risk scores against a baseline derived from the publicly available 1000 Genomes Project dataset as a reference population. The baseline score distribution was used to define categories of relative risk. Evaluation of the relative-polygenic risk score in larger case control cohorts should be considered to weigh the utility of the proposed relative risk scoring model; allowing for the stratification of patients who may benefit from enhanced monitoring or proactive oral health care regimens at the discretion of dental healthcare providers.

1 Introduction

It is well established that many diseases and conditions have an underlying genetic component, however, determination of genetic etiology can be challenging especially for multifactorial diseases (Dudbridge, 2016; Chasioti et al., 2019). Multifactorial, or complex diseases, arise from a combination of genetic predisposition and extrinsic factors such as environmental exposures or lifestyle choices. Unlike Mendelian (monogenic) disorders where variation in a single gene can give

rise to a particular phenotype, multifactorial (polygenic) disorders are often influenced by many genes, each of small effect (Manolio et al., 2009; Golan et al., 2014; Chasioti et al., 2019; Cano-Gamez and Trynka, 2020). The development of polygenic risk scores enables a deeper understanding of the impact genetics has on the development of diseases and conditions (Torkamani et al., 2018; Chasioti et al., 2019; Klarin and Natarajan, 2022). A polygenic risk score (PRS) illustrates how an individual's risk of developing a disease or condition compares to the broader population baseline. A PRS is comprised of genetic variants that have been identified to be associated with a particular disease or condition, typically as a result of large Genome-Wide Association Studies (GWAS) (Choi et al., 2021). Using statistical models and scoring algorithms, it is possible to evaluate how an individual's unique genetic profile contributes to their overall risk for a disease or condition of interest (Dudbridge, 2013; Chang et al., 2015; Choi et al., 2021).

PRS is an emerging tool in genetics. Its potential to focus on preventative healthcare in populations has allowed it to be implemented by health systems such as the UK's National Health Services (Genome UK). PRS for common diseases like coronary artery disease and type 2 diabetes have been developed and analyzed (Khera et al., 2018). Despite good oral hygiene habits, some individuals may have an increased genetic predisposition to certain dental problems. There are several oral health conditions that have a genetic basis including periodontal disease, dental caries (cavities) resistance, and oral cancers (Michalowicz et al., 2000; Bretz et al., 2005; Sarode et al., 2018; Shungin et al., 2019). Two of these conditions, periodontitis and dental caries, were estimated to be the eleventh and first most prevalent diseases respectively in a global study (Vos et al., 2017). Global economic impact of dental diseases amounted to $442 billion USD in 2010 (Listl et al., 2015).

Periodontal disease (PD) is a chronic inflammatory disease that leads to the degradation of tooth-supporting structures (Shungin et al., 2019). At first a patient might present with gingivitis, which is characterized by swollen and red gums that tend to bleed. Gingivitis can progress to periodontitis, which can result in bone or tooth loss as the gum detaches from the tooth (Kinane et al., 2017). It has been found that PD may increase the risk of cardiovascular disease by 19%. In addition, Type 2 diabetic individuals with severe PD have three times greater mortality risk compared to those with no or mild periodontitis. Periodontal therapy has also been shown to improve glycemic control in type 2 diabetic individuals (Nazir, 2017). PD is the main cause of tooth loss and is one of the most common oral conditions in the human population (Nazir, 2017; Vos et al., 2017).

Tooth decay results from destruction of the tooth's enamel. Tooth decay can be caused by the acid produced by the bacteria responsible for breaking down food in the mouth. The acid-induced enamel erosion creates a hole (cavity) in the tooth. If left untreated, infection or more severe outcomes such as tooth loss can occur (National Institute of Dental and Craniofacial Research, 2019). Worldwide, more than two billion people have cavities of the permanent teeth, and 520 million children have cavities in their primary teeth (Vos et al., 2017). There has been an increase in prevalence in developing countries due to the growing consumption of sugary foods, poor tooth brushing habits, and absence of adequate dental services (Teshome et al., 2021).

Oral disease prevention strategies should be incorporated in chronic systemic disease preventative initiatives to lessen the burden of disease in the population. Mitigating the incidence and prevalence of PD can reduce its associated systemic diseases (Liccardo et al., 2019). Overall improvements in oral health may lead to significant economic benefit with respect to decreased cost of treatment and labour resource allocation in dental clinics (Listl et al., 2015). Based on the existing literature, we

identified common genetic variants with evidence for association with PD and tooth decay/cavities and developed a genotyping panel coupled with a relative polygenic risk score. In this study we assessed the performance of the assay in a cohort of 27 dental clinic patients. A genotyping panel that integrates a relative-polygenic risk score can be utilized by dental professionals to identify individuals' genetic contribution to the overall predisposition to common oral health conditions.

2 Materials and Methods

2.1 Patient Enrolment

Participants were required to be 18 years old or older for inclusion in this study. Participant selection occurred independent of sex, ethnicity, and of any diagnosis or epidemiological indices. Informed consent was obtained from each of the 29 participants. The cohort consisted of 11 males, 15 females, and 3 individuals whose biological sex was not reported. Participants ranged in age from 21-87 years old. Enrolment took place over the course of one day (December 11th, 2021).

2.2 Study Design

Participant recruitment occurred at a single site (private dental office) in Ontario, Canada. The study was designed as a cohort study whereby individuals were approached during routine practice, under the dental clinic's natural settings. Written informed consent was obtained from study participants. Consented participants provided a buccal (cheek) swab for genetic testing. Select demographic information including biological sex and age was recorded. Genetic data was deidentified. Individual genetic testing reports were generated based on genetic variants associated with the two oral health indications (periodontal disease and cavities/tooth decay). Reports were issued to the study investigator and unblinded for the dental healthcare provider for discussion at a subsequent regularly scheduled visit. The study protocol was reviewed and approved by an independent ethics review board.

2.3 Variant Selection/Panel Design

Genetic variants to be included in an oral health-based genetic panel were first identified following a thorough review of the literature encompassing genetic risk and susceptibility for PD and dental cavities/tooth decay. Risk-associated variants were selected from a combination of case-control, meta-analyses, and GWAS publications. The list of candidate variants was narrowed down by considering study design/phenotype criteria, statistical significance, availability of effect alleles and weights in the original publications, and availability of coordinates in the Genome Reference Consortium Human GRCh38.p13 (GRCh38) reference genome assembly. After quality control (QC), a total of 25 single nucleotide polymorphisms (SNPs) were selected for PD, and 35 SNPs for cavities/tooth decay. These SNPs were included in the design of the proprietary Oral Health Panel for downstream application with the Agena Bioscience MassARRAY® System with Chip Prep Module (CPM) 96 (CP1603036).

2.4 Relative-Polygenic Risk Score (R-PRS) Development

The candidate SNPs were subsequently evaluated for utility in the development of a PRS in a background (baseline) control population. The 1000 Genomes Project (1000G) reference data set was selected as the control cohort for this purpose. The 1000G cohort consists of genetic data from over 2500 consented subjects from 26 global populations (Auton et al., 2015). The distribution of risk scores obtained from the control cohort thus served as the baseline against which all subsequent

genotyping results can be compared to, enabling the derivation of an individual's polygenic risk score relative to the general population, referred to as the relative polygenic risk score (R-PRS). An assessment of specificity and sensitivity was not performed as the intent of the relative polygenic risk score outlined herein is to discern where an individual's risk score lies with respect to the distribution observed in a population of self-reported healthy individuals.

A subset of candidate variants was selected for inclusion in the construction of the R-PRS based on the following criteria: (1) presence of the SNP in the 1000 Genomes Phase 3 Integrated Variant Calls dataset and (2) presence of effect (risk) alleles in either the reference or the alternative allele in the 1000 Genomes dataset. A separate R-PRS baseline was constructed for each oral health condition. Bioinformatic processing was achieved using PLINK 1.9, an open-source whole genome association analysis toolkit (Purcell et al., 2007). Development of each R-PRS followed a clumping and thresholding approach adapted from Choi et al. (Choi et al., 2021). The subset of SNPs that passed QC criteria were used in the final risk score calculations to establish the baseline distribution. Score calculation followed an additive model, whereby the SNP effect size (logarithm of the reported odds ratio), $S_{i,}$ was multiplied by the dosage (copies of the effect allele), G_{ij}, and summed across all SNPs. To account for missing genotypes, the sum is divided by the number of non-missing SNPs, $M_{j,}$ multiplied by the ploidy, P.

$$PRS_j = (\sum_i^N S_i * G_{ij}) \,/\, (P * M_j)$$

Where G_{ij} is the genotype for the i^{th} individual and j^{th} SNP

Risk scores for each of the 1000G subjects were plotted on a curve and the data assessed for normality for each condition tested. The risk score distributions were divided into percentiles, corresponding to categories of relative risk. These categories were used to assign relative risk to the participants of the study based on their personal risk scores.

2.5 Genotyping and Quality Control

Buccal swabs were obtained from 29 subjects using the ORAcollect•DNA (OCR-100) kit (DNA Genotek). Genomic DNA was extracted using the prepIT®•L2P protocol for 0.5mL of sample (DNA Genotek). Quantity and purity of DNA was determined using absorbance at wavelengths of 260 and 280nm (NanoDrop™ One, Thermo Scientific™). Sample identification and authentication was performed using the iPLEX® Pro Sample ID Panel (Agena Bioscience). The Sample ID Panel is comprised of 44 SNPs that are used to generate a unique genetic fingerprint for each sample, in addition to three biological sex markers and five copy number quality markers. This panel also serves as a secondary metric for DNA quality assessment. Samples are flagged as QC failures if any of the following criteria are met: (i) gender mismatch identified (discrepancy between detected and reported gender), (ii) ≥ 14 unsuccessful SNP calls, (iii) >11 low quality calls, (iv) ≤ 500 amplifiable copies of DNA, or (v) an unexpected match between two presumably unrelated patient samples. Samples that passed QC were prepared for genotyping with the oral health panel and QC failures were re-processed.

Forward and reverse primers targeting the candidate SNPs were designed in multiplex by Agena Bioscience for the Oral Health Panel. A total of 2 µl of genomic DNA (20 ng/µl concentration) was loaded in 96 well PCR plates along with the PCR master mix. The PCR amplification steps were performed on the T100 Thermal Cycler (Bio-Rad). Post-PCR processing with shrimp alkaline phosphatase (SAP), single base extension reactions, and SNP genotyping was performed as per the

manufacturer's protocol for custom MassARRAY® panels using iPLEX Pro chemistry (Agena Bioscience). The 29 samples were genotyped in two separate batches. A subset (n=3) of DNA samples were randomly selected to test inter-run reproducibility. Three non-template, negative controls were loaded onto each plate. Biological samples were flagged as QC failures if the high-quality genotype call rate was ≤ 80%. Samples failing QC criteria were not assigned an R-PRS.

3 Results

3.1 Oral Health Panel PRS Show Normal Population Distribution

The original genotyping panel design included 25 and 35 SNPS for PD and cavities/tooth decay respectively. These SNPs were selected on the basis of having published evidence for statistically significant disease association and coordinates in the GRCh38 reference genome assembly. Summary statistics were obtained from the corresponding publications for downstream application in the development of the PRS algorithm.

PRS analyses and baseline scoring was performed with PLINK 1.9, using the publicly available 1000G dataset as a reference population baseline (Purcell et al., 2007; Chang et al., 2015; Fairley et al., 2020). The baseline data (summary statistics) was first subjected to QC, whereby SNPs with two or more alleles or ambiguous alleles were removed. For the target data (1000G cohort data), SNPs were removed if (i) the minor allele frequency is <0.01, (ii) Hardy-Weinberg equilibrium is <1E-6, or (iii) if the SNP exhibited a rate of missingness >0.01 in the 1000G cohort. Additional QC filters were applied to remove samples with a high rate of genotype miss-calls, sex discordance, or close biological relatedness. After performing QC, 11 PD-associated SNPs and 8 cavities-associated SNPs were removed from the PRS algorithm. A total of 2548 and 1863 subjects were retained for PD and cavities respectively. Next, clumping was performed to remove SNPs if the r^2 linkage disequilibrium threshold is >0.1 with the index SNP. After clumping, 11 SNPs for PD and 27 SNPs for cavities remained for baseline scoring. The oral health conditions were scored separately using the additive dosage weighted model (Choi et al., 2021).

The resulting distribution of polygenic risk scores was approximately normally distributed (Figures 1, 2). PD showed stronger evidence of normality with points following a nearly linear line while cavities showed a poorer fit with skewing towards its low and high ranges. (Figures 1 and 2 in Supplementary Material). The median polygenic risk score was 0.1462235 for PD and -0.00203701 for cavities. The risk categories were calculated based on the 5^{th}, 15^{th}, 50^{th}, 85^{th}, and 95^{th} percentiles as the cut-offs for low, intermediate-low, average, intermediate-high, and high-risk categories respectively (Figures 3, 4, and Table 1 in Supplementary Material). The risk categories observed in the control cohort thus serve as the baseline against which an individual's risk can be assessed. All subsequent individual risk assessment is hence relative to that of the general population.

3.2 Independent Cohort Exhibits Normal Distribution

An assessment of the MassARRAY genotyping call rate revealed four underperforming SNP assays across the cohort. These SNPs had significant low quality or missing genotype calls and/or indistinguishable genotype call clusters. These SNPs are to be removed from the PRS algorithm and the 1000G baseline moving forward. The assessment of inter-run reproducibility revealed a mean genotype call concordance rate of 98% across the three tested samples. A total of 27 out of 29 DNA samples passed QC, with a 93.5% mean high quality call rate, and were assigned an R-PRS.

Relative risk scores were calculated for each subject using the same approach as the 1000G control cohort such that each individual score could be placed in the context of the baseline distribution. The poor performing SNPs were labelled as missing genotypes and accounted for in the PRS calculation using the conservative substitute formula ($2 * minor\ allele\ frequency * effect\ size$), enabling a direct comparison to the baseline risk scores. The distribution of risk scores observed in the study cohort closely followed the distributions of the 1000G cohort for both conditions. The quantile plots exhibited data normality (Figures 3 and 4 in Supplementary Material). As expected, most relative risk scores fell in the 50^{th} percentile (average relative risk) for both conditions (Figure 5). One high risk categorization was obtained for each condition. Despite the small sample size, this preliminary evaluation of the distribution of risk scores and relative risk categorizations suggest that the SNPs included in the final PRS algorithm can effectively stratify individuals into distinct relative risk bins.

4 Discussion

The availability of GWAS data has made it possible to better understand and model genetic susceptibility for polygenic diseases and traits such as PD and cavities (Chasioti et al., 2019; Lambert et al., 2019; Shungin et al., 2019). Through the curation of disease-associated SNPs and development of PRS algorithms in a large, ethnically diverse control population, we demonstrated an approach that can be used to infer relative genetic risk for two common oral health conditions. The results of this study demonstrate the utility of an oral health genotyping panel coupled with a relative-PRS to stratify patients into one of five categories of genetic risk relative to that of the general population.

One of the most challenging aspects of PRS is ensuring that the generated scores are equally applicable across all ethnic groups. Most existing data available within genome wide association studies are from individuals of European ancestry, as a result the current scores are most predictive for individuals within this population (Duncan et al., 2019; Lewis and Vassos, 2020). This issue needs to be highlighted as minority ethnic groups may be under-represented in research studies. This may possibly lead to a less predictive score for the under-represented ethnic groups (Lewis and Vassos, 2020). We (and others) within this space are addressing this gap as we continuously grow and diversify our population database (Morales et al., 2018). By doing so, our algorithm will in parallel be continuously updated to provide increasingly more accurate relative polygenic risk scores. We also acknowledge that our study lacked the incorporation of individuals' clinical background, which was beyond the scope of the study. We are currently building our algorithm to include the effects of environmental and behavioral factors to better estimate individual risk. Future work will be centered on evaluating the sensitivity and specificity of the proposed risk scoring algorithm. These metrics can be used to assess the ability of the risk score to reliably stratify patients who are high risk for the two dental conditions tested. While this assessment could not be performed in the current analysis, the approach outlined provides a means to contextualize the effect of risk associated SNPs across individuals in the absence of an independent validation cohort of sufficient sample size.

The need for genetic curriculum and education has been raised in the dental community in order to better understand how genetic testing can benefit patient care (Behnke and Hassell, 2004; Hart and Hart, 2016). It will be important to engage with the dental community and other healthcare providers to equip them with the knowledge needed to make informed decisions based on the outcomes of genetic testing, whether it be for Mendelian diseases or multifactorial conditions (Zimani et al., 2021). The results of the relative-PRS assessment can be used at the discretion of dentists to identify patients that might benefit from enhanced surveillance. It will be important to weigh the genetic risk in the context of other clinical and lifestyle factors that play a role in these multifactorial conditions. While the R-PRS is not intended to be diagnostic, it offers a more comprehensive insight to the

personal genetic susceptibility for PD and cavities that cannot be gained from a report on singular risk variants at common polymorphic loci. Extensive testing and validation will be an essential prerequisite to adoption of polygenic risk scoring models in the clinic.

5 Conflict of Interest

SM, JG, PQ, AK, MG are shareholders in AI Genetics. SM, PQ, AK, MG, and JG are employees of AI Genetics.

6 Author Contributions

SM designed the study protocol, supported patient enrolment, and oversaw the manuscript development. PQ and AK conducted experimentation. PQ, AK, and MG drafted the manuscript. SG provided clinical access and subject expertise. JG provided subject expertise and oversight. All authors contributed to manuscript review and editing.

7 Funding

The study was funded by the sponsor, AI Genetics.

8 Acknowledgments

The authors would like to thank the patients that contributed to the study.

Author 1: Contributed to conception, design, data acquisition and critically revised the manuscript.

Author 2: Contributed to data acquisition and critically revised the manuscript.

Author 3: Contributed to conception, design, data acquisition and interpretation, drafted and critically revised the manuscript.

Author 4: Contributed to conception, design, data acquisition and interpretations, performed statistical analyses, drafted and critically reviewed the manuscript.

Author 5: Performed statistical analyses, drafted and critically revised the manuscript.

Author 6: Contributed to conception, design and critically revised the manuscript.

9 Supplementary Material

The Supplementary Material for this article can be found online at:

10 Data Availability Statement

Further inquiries regarding generated datasets can be directed to the corresponding author.

11 References

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., et

al.(2015). A global reference for human genetic variation. *Nature* 526, 68–74. doi: 10.1038/nature15393.

Behnke, A. R., and Hassell, T. M. (2004). Need for genetics education in U.S. dental and dental hygiene programs. *J. Dent. Educ.* 68, 819–822.

Bretz, W. A., Corby, P. M., Schork, N. J., Robinson, M. T., Coelho, M., Costa, S., et al. (2005). Longitudinal analysis of heritability for dental caries traits. *J. Dent. Res.* 84, 1047–1051. doi: 10.1177/154405910508401115.

Cano-Gamez, E., and Trynka, G. (2020). From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Front. Genet.* 11, 1–21. doi: 10.3389/fgene.2020.00424.

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4, 1–16. doi: 10.1186/s13742-015-0047-8.

Chasioti, D., Yan, J., Nho, K., and Saykin, A. J. (2019). Progress in Polygenic Composite Scores in Alzheimer’s and Other Complex Diseases. *Trends Genet.* 35, 371–382. doi: 10.1016/j.tig.2019.02.005.

Choi, S. W., Shin, T., Mak, H., and Reilly, P. F. O. (2021). A guide to performing Polygenic Risk Score analyses Introduction to Polygenic Risk Scores. *Nat. Protoc.* 15, 2759–2772. doi: 10.1038/s41596-020-0353-1.A.

Dudbridge, F. (2013). Power and Predictive Accuracy of Polygenic Risk Scores. *PLoS Genet.* 9. doi: 10.1371/journal.pgen.1003348.

Dudbridge, F. (2016). Polygenic Epidemiology. *Genet. Epidemiol.* 40, 268–272. doi: 10.1002/gepi.21966.

Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., et al. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* 10. doi: 10.1038/s41467-019-11112-0.

Fairley, S., Lowy-Gallego, E., Perry, E., and Flicek, P. (2020). The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Res.* 48, D941–D947.doi: 10.1093/nar/gkz836.

Golan, D., Lander, E. S., and Rosset, S. (2014). Measuring missing heritability: Inferring the contribution of common variants. *Proc. Natl. Acad. Sci. U. S. A.* 111, E5272–E5281. doi: 10.1073/pnas.1419064111.

Hart, P. S., and Hart, T. C. (2016). Invited commentary: The need for human genetics and genomics in dental school curricula. *Mol. Genet. Genomic Med.* 4, 123–125. doi: 10.1002/mgg3.216.

Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, H. S., et al. (2018). 乳鼠心肌提取 HHS Public Access. *Physiol. Behav.* 176, 139–148. doi: 10.1038/s41588-018-0183-z.Genome-wide.

Kinane, D. F., Stathopoulou, P. G., and Papapanou, P. N. (2017). Periodontal diseases. *Nat. Rev. Dis. Prim.* 3, 1–14. doi: 10.1038/nrdp.2017.38.

Klarin, D., and Natarajan, P. (2022). Clinical utility of polygenic risk scores for coronary artery disease. *Nat. Rev. Cardiol.* 19, 291–301. doi: 10.1038/s41569-021-00638-w.

Lambert, S. A., Abraham, G., and Inouye, M. (2019). Towards clinical utility of polygenic risk scores. *Hum. Mol. Genet.* 28, R133–R142. doi: 10.1093/hmg/ddz187.

Lewis, C. M., and Vassos, E. (2020). Polygenic risk scores: From research tools to clinical instruments. *Genome Med.* 12, 1–11. doi: 10.1186/s13073-020-00742-5.

Liccardo, D., Cannavo, A., Spagnuolo, G., Ferrara, N., Cittadini, A., Rengo, C., et al. (2019). Periodontal disease: A risk factor for diabetes and cardiovascular disease. *Int. J. Mol. Sci.* 20. doi: 10.3390/ijms20061414.

Listl, S., Galloway, J., Mossey, P. A., and Marcenes, W. (2015). Global economic impact of dental diseases. *J. Dent. Res.* 94, 1355–1361. doi: 10.1177/0022034515602879.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., et al. (2009). Finding the missing heritability of complex diseases. *Nature* 461, 747–753. doi: 10.1038/nature08494.

Michalowicz, B. S., Diehl, S. R., Gunsolley, J. C., Sparks, B. S., Brooks, C. N., Koertge, T. E., et al. (2000). Evidence of a Substantial Genetic Basis for Risk of Adult Periodontitis. *J. Periodontol.* 71, 1699–1707. doi: 10.1902/jop.2000.71.11.1699.

Morales, J., Welter, D., Bowler, E. H., Cerezo, M., Harris, L. W., McMahon, A. C., et al. (2018). A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. *Genome Biol.* 19, 1–10. doi: 10.1186/s13059-018-1396-2.

Nazir, M. A. (2017). Prevalence of periodontal disease, its association with systemic diseases and prevention. *Int. J. Health Sci. (Qassim).* 11, 72–80. Available from: https://pubmed.ncbi.nlm.nih.gov/28539867.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795.

Sarode, G. S., Sarode, S. C., Maniyar, N., Anand, R., and Patil, S. (2018). Oral cancer databases: A comprehensive review. *J. Oral Pathol. Med.* 47, 547–556. doi: 10.1111/jop.12667.

Shungin, D., Haworth, S., Divaris, K., Agler, C. S., Kamatani, Y., Keun Lee, M., et al. (2019). Genome-wide analysis of dental caries and periodontitis combining clinical and self-reported data. *Nat. Commun.* 10. doi: 10.1038/s41467-019-10630-1.

Teshome, A., Muche, A., and Girma, B. (2021). Prevalence of Dental Caries and Associated Factors in East Africa, 2000–2020: Systematic Review and Meta-Analysis. *Front. Public Heal.* 9, 1–15. doi: 10.3389/fpubh.2021.645091.

Torkamani, A., Wineinger, N. E., and Topol, E. J. (2018). The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* 19, 581–590. doi: 10.1038/s41576-018-0018-x.

Vos, T., Abajobir, A. A., Abbafati, C., Abbas, K. M., Abate, K. H., Abd-Allah, F., et al. (2017). Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 390, 1211–1259. doi: 10.1016/S0140-6736(17)32154-2.

Zimani, A. N., Peterlin, B., and Kovanda, A. (2021). Increasing Genomic Literacy Through National Genomic Projects. *Front. Genet.* 12. doi: 10.3389/fgene.2021.693253.

Figure 1 PRS distribution for periodontal disease in the reference population (n=2548) calculated using PLINK 1.9 PRS pipeline.

Figure 2 PRS distribution for cavities in the reference population (n=1863) calculated using PLINK 1.9 PRS pipeline.

Figure 3 Risk score bins derived from the reference population used to define relative risk categories for cavities.

Figure 4 Risk score bins derived from the reference population used to define relative risk categories for periodontal disease.

Figure 5. Distribution of relative risk categorizations associated with polygenic risk scores observed in the study cohort (n=27). **(A)** Cavities distribution. **(B)** Periodontal disease distribution.