Research Article

BASCD Because healthy mouths matter

Community Dental Health
2025, Vol. 0(0) 1–8

DOI: 10.1177/0265539X251381496
journals.sagepub.com/home/cdh
Sage

Introduction to genome-wide association studies: Exploring their utility in caries and periodontitis

Lisa Kastenbom[1], Anders Esberg[1] and Simon Haworth[2]

Abstract

Genome-Wide Association Studies (GWAS) are the primary method used to investigate the effects of common genetic variants on health and disease and are increasingly used as an arena for applied epidemiological methods. There are now a growing number of applications of GWAS in oral health research, yielding exciting findings and holding great potential for future discoveries. While exciting, the GWAS approach has inherent limitations, and the success of these experiments depends on achieving adequate statistical power. In the context of dental diseases, where multiple genetic variants have small effects, very large samples are needed to unlock the full potential of GWAS. This article discusses the motivation for undertaking these studies, introduces the GWAS method, and highlights the potential of GWAS to help unravel the complex factors influencing oral health and diseases. Alongside the promises of GWAS, this article also summarizes the key challenges that need to be addressed for this method to be deployed usefully and suggests criteria to consider during the appraisal of a GWAS paper.

Keywords

periodontitis, dental caries, review, genome-wide association study, genetic variation

Introduction

Nearly all human complex traits are partially influenced by genetic factors (Polderman et al., 2015). There is natural scientific interest in understanding which genetic factors influence human traits or diseases, and various designs have been used to help map these associations.

Towards the end of the 20th century, advances in genetic sequencing and bioinformatic tools enabled the entire human genome to be sequenced, unlocking the development of novel study designs capable of capturing genetic variation across the entire genome (Lander et al., 2001; Venter et al., 2001). The genome-wide association study (GWAS) emerged as a method (Ozaki et al., 2002), with successful large-scale GWAS studies published from 2005 onwards (Klein et al., 2005).

General characteristics of GWAS

Testing for genetic association is conceptually similar to any other test for association between exposure and outcome, which might be carried out in classical epidemiology. Markers of host genetic variation, such as a single nucleotide polymorphism (SNP), are treated as exposure variables. A disease or trait is treated as an outcome, and a suitable method, such as linear or logistic regression, is used to test the for an association between them. In GWAS, this basic analysis is repeated many times until markers of genetic variation across the whole genome have been tested. By convention, GWAS are characterized by testing at least 500,000 markers, and modern studies often test more than 10,000,000 markers. Usually, these markers are focused to SNPs, the most common type of genetic variation in humans. GWAS does not require any prior information about the possible genetic basis of the disease since genetic markers across the whole genome are tested (Mayhew and Meyre, 2017; Pe'er et al., 2008). Therefore, the method is often described as an *unbiased* or *agnostic* approach. Because of the large amount of tests applied in GWAS, a stringent threshold is necessary to reduce false positives. The generally used threshold for genome-wide significance, $p < 5 \times 10^{-8}$ is derived from the Bonferroni correction for about a million independent tests. More precisely, it is the α of 0.05 divided with the estimated number of common variants in the human genome, which is approximately one million (Pe'er et al., 2008). Due to the stringent thresholds

[1]Department of Odontology, Umeå University, Umeå, Sweden
[2]Bristol Dental School, University of Bristol, Bristol, UK

Corresponding author:
Lisa Kastenbom, Department of Odontology, Umeå University, Norrlands Universitetssjukhus 1, Umeå, Sweden.
Email: lisa.kastenbom@umu.se

for genome-wide significance very large samples are needed. Since it can be difficult to achieve such large samples in single studies, genome-wide meta-analyses can be used to enhance statistical power. By combining data from different independent studies, the meta-analysis increases the sample size and statistical power. A separate advantage of meta-analysis is that the study can include more diverse populations and allows comparison of genetic effects in different studies. Ultimately, genome-wide meta-analysis is a useful approach which can lead to more reliable findings in GWAS (Nakaoka and Inoue, 2009). The original motivation for undertaking GWAS was to identify genes associated with a trait or disease since this can yield information about the molecular biology underlying it (Abdellaoui et al., 2023). This remains an important motivation, but there are now more new applications of GWAS in both clinical settings and applied epidemiology.

In clinical settings, there is growing interest in using polygenic score approaches to identify individuals at high risk of disease, providing opportunities for targeted prevention or personalized interventions (Zheng et al., 2024). These scores, calculated by aggregating the effects of many genetic variants, aim to capture the individual's genetic predisposition to a specific disease or trait (Torkamani et al., 2018). Pipelines and frameworks for clinical implementation of these scores have now been described for multiple traits (Lennon et al., 2024), but to date, there have been few applications to dental diseases (Fries et al., 2024). A key feature of these approaches is that the predictive ability of the score depends on the performance of the underlying GWAS study used to train the score, meaning high-quality and adequately powered GWAS studies are required. Other potential clinical applications involve using genetic data to predict drug response and guide prescribing (Duarte and Cavallari, 2021) or to prioritize possible protein therapeutic targets (De Cesco et al., 2020). However, they are yet to be adopted widely in dentistry.

In research, a growing range of methods apply GWAS results to elucidate modifiable risk factors for diseases. A key method is Mendelian randomization (MR), which utilizes genetic variants as instrumental variables to test for potential causal associations between exposures and outcomes (Davies et al., 2018; Smith and Ebrahim, 2024). This method is versatile and is widely applied, but again, it relies on an adequately powered underlying GWAS. A description of the MR approach is outside the scope of this article; however, it should be noted that it makes strict assumptions that are difficult to satisfy faithfully, meaning that applications of this approach should be carefully considered and compared with other sources of evidence (Haworth et al., 2024). There is legitimate concern about the application of this method (Smith and Ebrahim, 2024).

Challenges in undertaking GWAS

While GWAS are potentially useful with multiple downstream applications, they can also be challenging to undertake and have specific requirements for statistical power and control of confounding.

Statistical power is difficult to obtain in GWAS due to the small genetic effect sizes anticipated for a complex trait and the high multiple testing burden (discussed further below). The pattern of genetic contribution (i.e. the number of associated genetic variants and their effect sizes) differs between different traits (Timpson et al., 2018). Some traits have relatively simple genetic architecture, where a handful of genetic variants with large effects determine the observed phenotype. However, many traits are highly polygenic, influenced by multiple genetic variants, each with small individual effects. Interactions between genes and environmental factors may further modify these small effects. Detecting genuine (but small and possibly varying) genetic effects is inherently challenging given the background noise in almost all complex phenotypes (Lappalainen et al., 2024; Watanabe et al., 2019). The strong prior expectation is that common genetic variants associated with a complex disease will have small effects, meaning that the reported effect size in a study should always be examined for face validity (Table 1).

Since the GWAS method demands that genetic markers across the entire genome are tested the number of statistical tests is very high. There will be random fluctuation in the test statistics (e.g., p-values) even when the null hypothesis of no association is correct. Conducting many statistical tests risks false positive results, sometimes termed *multiple testing burden*. A stringent threshold accounting for the large number of multiple tests is needed to address this issue. By convention, most GWAS studies use a threshold of $p < 5 \times 10^{-8}$ to determine statistical significance, although some authors advocate an even stricter threshold, such as $p < 6.6 \times 10^{-9}$ (Morris et al., 2019). The use of an appropriately strict threshold is essential and should be appraised when reading a GWAS article (Table 1). Most studies use imputed genetic data, where analysis of measured genetic variants is supplemented by inferred unmeasured genetic variants based on a reference panel. This increases the genomic coverage and strengthens the ability to identify associations with genome-wide significance (Lau et al., 2024). Where imputation has been carried out, it is important that the reference panel is appropriate to the study population and that steps have been included to account for low confidence imputed variants, for example by removing poorly imputed variants prior to analysis.

While strict thresholds are important for reducing false positive associations, the trade-off is that many tests will produce false negatives (Gao et al., 2008; Risch and Merikangas, 1996; Sabatti et al., 2003; Storey and Tibshirani, 2003; Tam et al., 2019). To address the stringent threshold criteria for significance, studies need very *large samples*, ideally in the hundreds of thousands to millions of participants, for optimal power. Only a single study has convincingly shown that additional participants would not result in additional discovery of associated common genetic variants, and that included over five million participants (Yengo et al., 2022). Saturated discovery of the full spectrum of genetic contributions to a trait (i.e both common and rare variants) may require even larger samples. Obtaining such large samples is difficult, requiring

Table 1. Main considerations when reading a GWAS article.

Consideration	What to appraise	Rationale
Statistical significance	Appropriate threshold used, e.g., $p < 5 \times 10^{-8}$ or stricter	The multiple testing burden brings a high risk of false positives unless a strict threshold is used. It is not appropriate to report 'suggestive' associations, which can be explained by chance
Sample size	Large sample (ideally hundreds of thousands to millions of participants)	Much larger sample sizes are needed in GWAS as the effects of individual genetic variants are very small, and the multiple testing burden is very high
Population stratification	Steps used to minimize effects of population stratification, e.g., adjustment for genetic principal components or use of a mixed model	Population stratification can confound the observed association between a genotype and disease. Steps must always be taken to minimize this bias's effects
Face validity of reported effect sizes	Studies claiming to identify novel and large genetic effects on caries and periodontitis need to be interpreted with caution	Large genetic effects should have been easily visible during the first 15 years. The current consensus is that caries and periodontitis are polygenic diseases where common genetic variants have small effects
Constancy of genetic effect	Where possible. Consistency of genetic effect should be examined and reported	False positive results are unlikely to replicate consistently across different populations. Consistency can be examined by replication or by comparing effect sizes in different underlying studies of a meta-analysis
Phenotypic definition	Does the phenotype adequately capture variation in the target dental disease trait? Does the phenotype capture other heritable traits?	Random measurement errors may reduce power in analysis but are unlikely to confound analysis since these random errors are uncorrelated with genotypes Genetic effects on measurement (for example if a heritable biomarker is used to define periodontitis cases) can induce association

collaborative efforts, often across multiple research institutions and cohorts. Sample size is a key factor to consider when appraising GWAS studies (Table 1). Studies claiming to identify novel associations in small studies should be treated with great caution since any real associations should be readily visible in existing studies.

Similar to confounding in classical epidemiological studies, GWAS can be confounded if any unobserved variable is associated with both the exposure (genotype) and outcome (disease). This problem is termed *population stratification*, and several approaches are used to try and address this, including the use of genetic principal components, mixed model approaches and Bayesian modelling approaches. Modelling genetic similarity as random effects in mixed models is one approach to deal with this problem, which accounts for genetic similarity due to ancestry and relatedness, helping to avoid false positives while preserving statistical power (Liu et al., 2023). The challenges of accurately accounting for population stratification are discussed elsewhere (Lawson et al., 2020). While there is ongoing debate about the most effective approach, it is important that all studies include some attempt to account for confounding by population stratification (Table 1).

A genetic variant with consistent effects in different samples, increases the strength of evidence that the association is genuine. This can be carried out by undertaking targeted replication in an independent cohort or (where the GWAS included multiple participating cohorts), by testing the stability of effect across those cohorts, for example, using a test for heterogeneity. Regardless of the approach used, tests for consistency in genetic effect are useful when interpreting a reported association.

It can be difficult to obtain detailed measurement in the large samples needed for a GWAS, meaning that the disease measures used are often less precise than those that would be used in other epidemiological study in smaller samples. To collect a large enough sample, the researcher must sometimes make compromises towards including a simpler phenotype than would be used in other study designs. Decisions on phenotypic definitions need to consider, how the trait is being measured, the measure's precision and the statistical power required to detect meaningful genetic associations (Table 1). GWAS may sometimes use proxy phenotypes, typically a related characteristic that is easier to measure or more widely available in large datasets but has similar underlying genetic determinants. Where used, proxy phenotypes should ideally be presented with data supporting their validity. Care is needed when measurement for GWAS includes other partially heritable traits in constructing disease measures. For example, using BMI as a measure for obesity could lead to finding GWAS associations at loci which influence height rather than obesity, since height is partially heritable and is used to derive the BMI score. In oral health GWAS, including traits like bacterial colonization and local inflammatory response in a phenotype may increase the discovery yield of genetic associations (Agler et al., 2019). However, analogous to the height/BMI example above, it is possible these associations are driven by genetic effects on the measured score rather than the oral disease itself. Care is needed when using results from such experiments in downstream applications such as Mendelian randomization (Haworth et al., 2024).

In summary, GWAS studies can be challenging since they require very large samples and must account for

potential effects of population stratification. Factors to consider when appraising these studies are summarized in Table 1.

Genetic contributions to periodontitis and caries

Periodontal diseases and caries are two of the most prevalent diseases around the world and the most prevalent oral diseases. Both are potentially ideal targets for GWAS, since their aetiology includes behavioral, environmental, and host genetic components. A variety of methods have been used for more than 70 years to explore host genetic effects on oral health (Boomsma et al., 2002; Hunt et al., 1944; Krasse, 2001; Michalowicz et al., 2000; Rijsdijk and Sham, 2002; Vieira, 2021). Over time it has become possible to quantify the relative contribution of host and environmental factors for caries (Bretz et al., 2005; Haworth et al., 2020; Shaffer et al., 2012; Wang et al., 2010) and periodontitis (Haworth et al., 2021; Nibali et al., 2020) with greater precision. Despite the potential usefulness for understanding the genetic contribution to disease, these studies cannot identify genetic variants associated with the trait and lack the capacity to distinguish between different genetic mechanisms underlying the trait.

Around 2010, researchers in oral diseases adopted GWAS methodology (Kim et al., 2009; Mangold et al., 2010; Pillas et al., 2010), including Shaffer's pioneering work in periodontitis and dental caries (Schaefer et al., 2010, Shaffer et al., 2011).

State of the literature

Schaefer's (2010) work led to the discovery of the first genetic risk locus for periodontal disease, underscoring the role of predisposition for this inflammatory disease. However, early reports were limited by the methods and size of samples available at the time (Table 2). In the following years, a number of GWAS were relatively small studies targeting traits such as chronic periodontitis (CP) (Sanders et al., 2017; Shaffer et al., 2014; Shimizu et al., 2015; Teumer et al., 2013), aggressive periodontitis (AgP) (Hashim et al., 2015; Munz et al., 2017), and host genetic effects on putative periodontal pathogens (Divaris et al., 2013; Rhodin et al., 2014).

Recent years have seen a shift towards including multiple cohorts in meta-analyses, which both increases the power of GWAS and provides opportunities to test the consistency of effects across different populations. This approach allows for larger samples and the ability to identify additional risk loci that might not be detectable in smaller, individual studies. By combining data from various cohorts, researchers can better understand the genetic architecture of periodontitis and identify more robust genetic associations.

Adopting the meta-analysis approach associations between periodontitis and specific nucleotide variants at two loci were identified. These findings support the involvement of these genetic regions in the susceptibility to periodontitis, aligning with other studies that have identified the same loci (Munz et al., 2017). Another group identified a novel locus associated with aggressive periodontitis (De Almeida et al., 2024; Munz et al., 2017) which aligned with previous research identifing the same locus in a different cohort. Schaefer's work similarly highlighted this region, reinforcing its potential role in the genetic susceptibility to aggressive periodontitis.

Although there are now some consensus genetic associations, the overall discovery yield for periodontitis traits from the first 15 years of GWAS studies has been low. The possible reasons for this are discussed below.

Soon after the first GWAS of caries in primary teeth (Shaffer et al., 2011), Wang et al. (2012) published the first on adult caries, in a relatively small sample. These early studies generally had modest samples and relatively low discovery yields (Shaffer et al., 2011, 2013; Wang et al., 2012; Zeng et al., 2013).

Like studies of periodontitis, there has been a shift over time towards larger samples, partially driven by the formation of consortia and the use of genome-wide meta-analyses. One example is the GLIDE (Gene-Lifestyle Interactions in Dental Endpoints) consortium, which is a collaborative research effort focused on understanding the genetic and lifestyle factors influencing dental diseases, specifically dental caries, and periodontitis. This consortium (now in the second round of analysis) reported on more than half a million individuals in 2019, combining detailed clinical data with genetically validated proxy phenotypes to investigate dental caries and periodontitis. By leveraging self-reported dental disease proxies and clinical measures in a meta-analysis, the study identified novel risk loci for dental caries that were not detectable in smaller, single-cohort studies (Shungin et al., 2019). Other examples of researchers combining multiple cohorts to enhance power include studies by Sofer et al. (2016), Alotaibi et al. (2021), Haworth et al. (2018) and Orlova et al. (2022).

Parallel to studies of adult populations, applications of GWAS for childhood caries traits remain challenging due to generally small samples with few consensus association signals identified to date (Ballantine et al., 2018; Olatosi et al., 2021). While a small number of genetic variants are reportedly associated with childhood caries, the genes near these variants often lack known functions in caries development (Alotaibi et al., 2021). Like adult populations, there has been a move towards consortia (Haworth et al., 2018; Orlova et al., 2022).

In addition to enhancing statistical power, researchers have made efforts to understand ancestry-specific genetic effects using GWAS (Alotaibi et al., 2021; Orlova et al., 2019). These studies aim to uncover genetic associations that might be missed in European-centric studies. To date, however, most of the available literature relates to populations of European ancestry (Table 2).

Table 2. Examples of genome-wide studies of periodontitis and caries.

Author	Publication year	Trait	Genetic ancestry of participants	Sample size total (cases/controls)*	Significant GWAS findings
Periodontitis					
Schafer	2010	AgP[1]	European	641 (141/500)	Yes
Teumer	2013	CP[2] as: PAL[3] or attachment loss > 4 mm, CDC/AAP definition[4] or DPAL[5]	European	4032 (2215/1817)	No
Divaris	2013	CP	European	2867 (958/1909)	No
Shaffer	2014	Probing dept ≥5.5 mm	European	673 (176/497)	No
Rhodin	2014	CP & periodontal pathogen colonization	European	2584 (1823/761)	No
Shimizu	2015	Local or generalized periodontitis	Japanese	9573 (1593/7980)	No
Munz	2017	AgP	European	AgP 8000 (896/7104) CP 2412 (993/1419)	Yes
Sanders	2017	Interproximal clinical attachment level	Admixed Hispanic/Latino	10,935*	Yes
Bevilacqua	2018	CP	European	602 (442/160)	Yes
Shungin	2019	CP (clinical or self-reported)	Multiethnic	45,563 (17,353/28,210)	Yes
Ishigaki	2020	Periodontitis (not specified)	Japanese	3,219/209,234	No
Almeida	2024	Stage III/IV grade C periodontitis	European	9123 (1306/7817)	Yes
Caries					
Shaffer	2011	Primary dentition	European	1264 (618/646)	No
Wang	2012	Permanent dentition, DMFS[6], DMFT[7], DFS[8], severity	Multi-ethnic	7443*	No
Shaffer	2013	Permanent dentition, DMFS	European	920*	Yes
Morrison	2016	DMFT, DMFS	Hispanic/Latino	12 803*	Yes
Haworth	2018	Primary teeth, permanent teeth. With/without treated caries	Multi-ethnic	17,037 (6,922/10,115)	Yes
Shungin	2019	DMFS/DFS, Nteeth	Multi-ethnic	26,792*	Yes
Alotaibi	2021	Primary and permanent dentition	Multi-ethnic	3856 (1050/2806)	No
Olatosi	2021	ECC[9] yes/no	African	691 (338/353)	No
Alotaibi	2021	Primary and permanent dentition (DFT[10]/dft[11])	Multi-ethnic	Primary 1,116vbn Permanent 2570	Yes
Orlova	2023	ECC yes/no	European	2967 (843/2124)	Yes

*Not specified or quantitative measure is used. [1] = Aggressive periodontitis, [2] = Chronical periodontitis, [3] = probing attachment levels, [4] = Case definitions Proposed for Population-Based Surveillance of Periodontitis (mild, moderate and severe disease), [5] = difference in mean PAL between 5-year follow-up and baseline [6] = Decayed Missed Filled Surfaces [7] = Decayed Missed Filled Teeth [8] = Decayed Filled Surfaces [9] = Early Childhood Caries [10] = Decayed Filled Teeth [11] = Decayed Filled Primary Teeth.

Summary of GWAS in caries and periodontal disease

GWAS has contributed to the knowledge of thousands of genetic variants associated with different diseases and traits. Compared to other complex systemic traits and diseases (Andrews et al., 2023; Benafif et al., 2018; Caliskan et al., 2021; IPDC, 2020; Mahajan et al., 2022; Walsh et al., 2023) there is a clear gap between GWAS evidence on caries and periodontitis. The discovery yield for caries and periodontitis remains low compared to other common diseases. The lack of GWAS results in diverse populations for caries and periodontitis, is a further limitation. There are several possible explanations for this evidence gap. Sample size remains the most likely explanation, since oral examination is impractical in large epidemiological studies. Oral health traits are also difficult to measure compared to other complex traits (e.g., height) in very large samples and considerable investment is needed to undertake dental examinations or develop systems which allow retrieval of records from primary dental care. Such investment has not always been available since funding for oral health research has been limited compared to GWAS on general health problems. In addition to the practical difficulties of oral disease measurement, there are also methodological challenges with measurement, since (unlike a trait such as height), the phenotypic definition for caries or periodontitis contains effects of both the disease and treatment. Determining the appropriate sample size for GWAS in caries and periodontitis depends on multiple factors and considerations, such as the complexity of the oral disease, the genetic diversity and the population being studied. While it may be possible to detect some associations in studies with tens of thousands of individuals (Haworth et al., 2018; Morrison et al., 2016), experience from other complex traits suggests sample sizes of millions of participants will be needed to saturate association discovery for common variants.

Researchers apply different approaches to the phenotypic definitions of caries, such as, continuous measures of Decayed, Missing and Filled teeth (DMFT) and binary traits. Periodontitis definitions often include probing depth or self-reported periodontitis but definitions vary even among these variables (Table 1).

Population stratification presents a potential confounding factor in GWAS of caries and periodontitis, because individuals with different demographic or ancestral histories, such as those influenced by immigration patterns, often exhibit differences in dental health outcomes due to environmental and socioeconomic factors, including variations in diet, income and access to dental care. These factors, shaped by demographic history, influence the prevalence and severity of caries and periodontitis (Cruz et al., 2009; Gómez-Costa et al., 2021; Julihn et al., 2010). Allele frequencies vary by ancestral background, creating a scenario where non-causal associations between single nucleotide polymorphisms (SNPs) and caries or periodontitis might be observed. Such spurious associations result from the correlation between genetic variants and ancestry-related environmental factors, rather than a direct causal relationship with the dental condition under study.

Finally, the genetic architecture of caries and periodontitis appears highly polygenic, and the small genetic effects of a polygenic trait are inherently harder to identify than a trait with more blended or oligogenic architecture. Further concerted efforts are needed to address this evidence gap.

In general, GWAS for caries and periodontitis have provided some insight into the genetic basis of these important diseases, with some highlights presented in Table 2. When looking at the overall picture of multiple studies on caries and periodontitis, it becomes clear that many studies face similar challenges due to these complexities of genetic architecture, small effect sizes and small sample sizes available, which often results in insufficient statistical power. Continued research efforts, emphasizing larger studies, more diverse populations, and interdisciplinary collaboration, will be essential for unlocking the full potential of GWAS in this field. Ongoing efforts which aim to address these problems include the second round of analysis in the GLIDE2 consortium (Divaris et al., 2022).

In conclusion, GWAS have proved to be powerful tools in identifying genetic variants associated with complex traits. However, conducting GWAS for caries and periodontitis presents several challenges, primarily because very large samples are needed to detect small genetic effects. To date, few genetic variants have been identified in GWAS for caries and periodontitis compared to other prevalent diseases. Further large-scale collaborative efforts are required to better understand the genomic basis of these important diseases.

ORCID iD

Lisa Kastenbom https://orcid.org/0000-0002-4702-1306

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: We gratefully acknowledge financial support provided by Wellcome Trust [227534/Z/23/Z].

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Abdellaoui A, Yengo L, Verweij KJH, et al. (2023) 15 years of GWAS discovery: realizing the promise. *The American Journal of Human Genetics* 110: 179–194.

Agler CS, Shungin D, Ferreira Zandoná AG, et al. (2019) *Protocols, Methods, and Tools for Genome-Wide Association Studies (GWAS) of Dental Traits. Methods in Molecular Biology*. New York: Springer, 493–509.

Alotaibi RN, Howe BJ, Chernus JM, et al. (2021) Genome-Wide Association Study (GWAS) of dental caries in diverse populations. *BMC Oral Health* 21: 377.

Andrews SJ, Renton AE, Fulton-Howard B, et al. (2023) The complex genetic architecture of Alzheimer's disease: novel insights and future directions. *EBioMedicine* 90: 104511.

Ballantine JL, Carlson JC, Ferreira Zandona AG, et al. (2018) Exploring the genomic basis of early childhood caries: a pilot study. *International Journal of Paediatric Dentistry* 28: 217–225.

Benafif S, Kote-Jarai Z, et al. (2018) A review of prostate cancer Genome-Wide Association Studies (GWAS). *Cancer Epidemiology, Biomarkers & Prevention* 27: 845–857.

Boomsma D, Busjahn A and Peltonen L (2002) Classical twin studies and beyond. *Nature Reviews Genetics* 3: 872–882.

Bretz WA, Corby PM, Schork NJ, et al. (2005) Longitudinal analysis of heritability for dental caries traits. *Journal of Dental Research* 84: 1047–1051.

Caliskan M, Brown CD and Maranville JC (2021) A catalog of GWAS fine-mapping efforts in autoimmune disease. *The American Journal of Human Genetics* 108: 549–563.

Cruz GD, Chen Y, Salazar CR, et al. (2009) The association of immigration and acculturation attributes with oral health among immigrants in New York City. *American Journal of Public Health* 99: S474–S480.

Davies NM, Holmes MV and Davey Smith G (2018) Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* 362: k601.

De Almeida SD, Richter GM, de Coo A, et al. (2024) A genome-wide association study meta-analysis in a European sample of stage III/IV grade C periodontitis patients ≤35 years of age identifies new risk loci. *Journal of Clinical Periodontology* 51: 431–440.

De Cesco S, Davis JB and Brennan PE (2020) TargetDB: a target information aggregation tool and tractability predictor. *PLoS One* 15: e0232644.

Divaris K, Monda KL, North KE, et al. (2013) Exploring the genetic basis of chronic periodontitis: a genome-wide association study. *Human Molecular Genetics* 22: 2312–2324.

Divaris K, Haworth S, Shaffer JR, et al. (2022) Phenotype harmonization in the GLIDE2 oral health genomics consortium. *Journal of Dental Research* 101: 1408–1416.

Duarte JD and Cavallari LH (2021) Pharmacogenetics to guide cardiovascular drug therapy. *Nature Reviews Cardiology* 18: 649–665.

Fries N, Haworth S, Shaffer JR, et al. (2024) A polygenic score predicts caries experience in elderly Swedish adults. *Journal of Dental Research* 103: 502–508.

Gao X, Starmer J and Martin ER (2008) A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genetic Epidemiology* 32: 361–369.

Gómez-Costa D, San-Roman-Montero J, Rojo R, et al. (2021) Self-reported prevalence of periodontal disease among the Spanish population and immigrants: 2006, 2011/12 and 2017: a population-based study. *BMC Oral Health* 21: 215.

Hashim NT, Linden GJ, Ibrahim ME, et al. (2015) Replication of the association of GLT6D1 with aggressive periodontitis in a Sudanese population. *Journal of Clinical Periodontology* 42: 319–324.

Haworth S, Shungin D, van der Tas JT, et al. (2018) Consortium-based genome-wide meta-analysis for childhood dental caries traits. *Human Molecular Genetics* 27: 3113–3127.

Haworth S, Esberg A, Lif Holgerson P, et al. (2020) Heritability of caries scores, trajectories, and disease subtypes. *Journal of Dental Research* 99: 264–270.

Haworth S, Esberg A, Kuja-Halkola R, et al. (2021) Using national register data to estimate the heritability of periodontitis. *Journal of Clinical Periodontology* 48: 756–764.

Haworth S, Timpson NJ and Divaris K (2024) Mendelian randomization studies of periodontitis: understanding benefits and natural limitations in an applied context. *Journal of Clinical Periodontology* 51: 1258–1266.

Hunt HR, Hoppert CA and Erwin WG (1944) Inheritance of susceptibility to caries in albino rats (mus norvegicus). *Journal of Dental Research* 23: 385–401.

Julihn A, Ekbom A and Modéer T (2010) Migration background: a risk factor for caries development during adolescence. *European Journal of Oral Sciences* 118: 618–625.

Kim H, Ramsay E, Lee H, et al. (2009) Genome-wide association study of acute post-surgical pain in humans. *Pharmacogenomics* 10: 171–179.

Klein RJ, Zeiss C, Chew EY, et al. (2005) Complement factor H polymorphism in age-related macular degeneration. *Science* 308: 385–389.

Krasse B (2001) The vipeholm dental caries study: recollections and reflections 50 years later. *Journal of Dental Research* 80: 1785–1788.

Lander ES, Linton LM, Birren B, et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860–921.

Lappalainen T, Li YI, Ramachandran S, et al. (2024) Genetic and molecular architecture of complex traits. *Cell* 187: 1059–1075.

Lau W, Ali A, Maude H, et al. (2024) The hazards of genotype imputation when mapping disease susceptibility variants. *Genome Biology* 25: 7–17.

Lawson DJ, Davies NM, Haworth S, et al. (2020) Is population structure in the genetic biobank era irrelevant, a challenge, or an opportunity? *Human Genetics* 139: 23–41.

Lennon NJ, Kottyan LC, Kachulis C, et al. (2024) Selection, optimization and validation of ten chronic disease polygenic risk scores for clinical implementation in diverse US populations. *Nature Medicine* 30: 480–487.

Liu Z, Turkmen AS and Lin S (2023) Population stratification correction using Bayesian shrinkage priors for genetic association studies. *Annals of Human Genetics* 87: 302–315.

Mahajan A, Spracklen CN, Zhang W, et al. (2022) Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nature Genetics* 54: 560–572.

Mangold E, Ludwig KU, Birnbaum S, et al. (2010) Genome-wide association study identifies two susceptibility loci for non-syndromic cleft lip with or without cleft palate. *Nature Genetics* 42: 24–26.

Mayhew AJ and Meyre D (2017) Assessing the heritability of complex traits in humans: methodological challenges and opportunities. *Current Genomics* 18: 332–340.

Michalowicz BS, Diehl SR, Gunsolley JC, et al. (2000) Evidence of a substantial genetic basis for risk of adult periodontitis. *Journal of Periodontology* 71: 1699–1707.

Morris JA, Kemp JP, Youlten SE, et al. (2019) An atlas of genetic influences on osteoporosis in humans and mice. *Nature Genetics* 51: 258–266.

Morrison J, Laurie CC, Marazita ML, et al. (2016) Genome-wide association study of dental caries in the Hispanic communities health study/study of Latinos (HCHS/SOL). *Human Molecular Genetics* 25: 807–816.

Munz M, Willenborg C, Richter GM, et al. (2017) A genome-wide association study identifies nucleotide variants at SIGLEC5 and DEFA1A3 as risk loci for periodontitis. *Human Molecular Genetics* 26: 2577–2588.

Nakaoka H and Inoue I (2009) Meta-analysis of genetic association studies: methodologies, between-study heterogeneity and winner's curse. *Journal of Human Genetics* 54: 615–623.

Nibali L, Almofareh SA, Bayliss-Chapman J, et al. (2020) Heritability of periodontitis: a systematic review of evidence from animal studies. *Archives of Oral Biology* 109: 104592.

Olatosi OO, Li M, Alade AA, et al. (2021) Replication of GWAS significant loci in a sub-saharan African cohort with early childhood caries: a pilot study. *BMC Oral Health* 21: 274.

Orlova E, Carlson JC, Lee MK, et al. (2019) Pilot GWAS of caries in African-Americans shows genetic heterogeneity. *BMC Oral Health* 19: 215.

Orlova E, Dudding T, Chernus JM, et al. (2022) Association of early childhood caries with bitter taste receptors: a meta-analysis of genome-wide association studies and transcriptome-wide association study. *Genes* 14: 59.

Ozaki K, Ohnishi Y, Iida A, et al. (2002) Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. *Nature Genetics* 32: 650–654.

Pe'er I, Yelensky R, Altshuler D, et al. (2008) Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genetic Epidemiology* 32: 381–385.

Pillas D, Hoggart CJ, Evans DM, et al. (2010) Genome-wide association study reveals multiple loci associated with primary tooth development during infancy. *PLoS Genetics* 6: e1000856.

Polderman TJ, Benyamin B, de Leeuw CA, et al. (2015) Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics* 47: 702–709.

Rhodin K, Divaris K, North KE, et al. (2014) Chronic periodontitis genome-wide association studies: gene-centric and gene set enrichment analyses. *Journal of Dental Research* 93: 882–890.

Rijsdijk FV and Sham PC (2002) Analytic approaches to twin data using structural equation models. *Briefings in Bioinformatics* 3: 119–133.

Risch N and Merikangas K (1996) The future of genetic studies of complex human diseases. *Science* 273: 1516–1517.

Sabatti C, Service S and Freimer N (2003) False discovery rate in linkage and association genome screens for complex disorders. *Genetics* 164: 829–833.

Sanders AE, Sofer T, Wong Q, et al. (2017) Chronic periodontitis genome-wide association study in the Hispanic community health study/study of latinos. *Journal of Dental Research* 96: 64–72.

Schaefer AS, Richter GM, Nothnagel M, et al. (2010) A genome-wide association study identifies GLT6D1 as a susceptibility locus for periodontitis. *Human Molecular Genetics* 19: 553–562.

Shaffer JR, Wang X, Feingold E, et al. (2011) Genome-wide association scan for childhood caries implicates novel genes. *Journal of Dental Research* 90: 1457–1462.

Shaffer JR, Feingold E, Wang X, et al. (2012) Heritable patterns of tooth decay in the permanent dentition: principal components and factor analyses. *BMC Oral Health* 12: 7.

Shaffer JR, Feingold E, Wang X, et al. (2013) GWAS of dental caries patterns in the permanent dentition. *Journal of Dental Research* 92: 38–44.

Shaffer JR, Polk DE, Wang X, et al. (2014) Genome-wide association study of periodontal health measured by probing depth in adults ages 18–49 years. *Genes Genomes Genetics* 4: 307–314.

Shimizu S, Momozawa Y, Takahashi A, et al. (2015) A genome-wide association study of periodontitis in a Japanese population. *Journal of Dental Research* 94: 555–561.

Shungin D, Haworth S, Divaris K, et al. (2019) Genome-wide analysis of dental caries and periodontitis combining clinical and self-reported data. *Nature Communications* 10: 2773.

Smith GD and Ebrahim S (2024) Mendelian randomisation at 20 years: how can it avoid hubris, while achieving more? *Lancet Diabetes & Endocrinology* 12: 14–17.

Sofer T, Shaffer JR, Graff M, et al. (2016) Meta-analysis of genome-wide association studies with correlated individuals: application to the Hispanic community health study/study of Latinos (HCHS/SOL). *Genetic Epidemiology* 40: 492–501.

Storey JD and Tibshirani R (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America* 100: 9440–9445.

Tam V, Patel N, Turcotte M, et al. (2019) Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics* 20: 467–484.

Teumer A, Holtfreter B, Völker U, et al. (2013) Genome-wide association study of chronic periodontitis in a general German population. *Journal of Clinical Periodontology* 40: 977–985.

The International Parkinson Disease Genomics Consortium (IPDGC) (2020) Ten years of the international parkinson disease genomics consortium: progress and next steps. *Journal of Parkinson's Disease* 10: 19–30.

Timpson NJ, Greenwood CMT, Soranzo N, et al. (2018) Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nature Reviews Genetics* 19: 110–124.

Torkamani A, Wineinger NE and Topol EJ (2018) The personal and clinical utility of polygenic risk scores. *Nature Reviews Genetics* 19: 581–590.

Venter JC, Adams MD, Myers EW, et al. (2001) The sequence of the human genome. *Science* 291: 1304–1351.

Vieira AR (2021) Individual susceptibility to dental caries: the vipeholm study. *Monographs in Oral Science* 30: 36–44.

Walsh R, Jurgens SJ, Erdmann J, et al. (2023) Genome-wide association studies of cardiovascular disease. *Physiological Reviews* 103: 2039–2055.

Wang X, Shaffer JR, Weyant RJ, et al. (2010) Genes and their effects on dental caries may differ between primary and permanent dentitions. *Caries Research* 44: 277–284.

Wang X, Shaffer JR, Zeng Z, et al. (2012) Genome-wide association scan of dental caries in the permanent dentition. *BMC Oral Health* 12: 57.

Watanabe K, Stringer S, Frei O, et al. (2019) A global overview of pleiotropy and genetic architecture in complex traits. *Nature Genetics* 51: 1339–1348.

Yengo L, Vedantam S, Marouli E, et al. (2022) A saturated map of common genetic variants associated with human height. *Nature* 610: 704–712.

Zeng Z, Shaffer JR, Wang X, et al. (2013) Genome-wide association studies of pit-and-fissure- and smooth-surface caries in permanent dentition. *Journal of Dental Research* 92: 432–437.

Zheng Z, Liu S, Sidorenko J, et al. (2024) Leveraging functional genomic annotations and genome coverage to improve polygenic prediction of complex traits within and between ancestries. *Nature Genetics* 56: 767–777.