#### RESEARCH



# Genetic risk factors for periodontitis: a genome-wide association study using UK Biobank data

Chenyi Gao<sup>1,2</sup> · Mark M. Iles<sup>3,4</sup> · David Timothy Bishop<sup>5</sup> · Harriet Larvin<sup>1</sup> · David Bunce<sup>6</sup> · Bei Wu<sup>7</sup> · Huabin Luo<sup>8</sup> · Luigi Nibali<sup>10</sup> · Susan Pavitt<sup>2</sup> · Jianhua Wu<sup>1</sup> · Jing Kang<sup>9</sup>

Received: 5 November 2024 / Accepted: 3 February 2025 / Published online: 14 February 2025 © The Author(s) 2025, corrected publication 2025

#### Abstract

**Objectives** Periodontitis is linked with many health conditions, but its genetic basis is not yet understood. This genome-wide association study (GWAS) aimed to investigate the genetic variants associated with periodontitis.

**Materials and methods** This study utilised UK Biobank participants of European descent. Individuals were categorised as "having periodontitis" if they self-reported having 'painful gums', 'bleeding gums' or 'loose teeth' (n=68,482), or as "controls" for those without these symptoms (n=307,342). We conducted GWAS of this binary periodontitis phenotype using logistic regression models with PLINK2.0 adjusting for age, sex and the first 15 principal components to account for population stratification.

**Results** There were 376,611 participants (mean baseline  $age=57\pm7.9$  SD) included in the GWAS, and four significant loci were identified: rs775476621 on chromosome 11 (Odds Ratio, OR[T]: 3.08,  $p=1.01 \times 10^{-8}$ ), rs751014048 on chromosome 11 (OR[G]: 3.07,  $p=1.04 \times 10^{-8}$ ), rs149922301 on chromosome 4 near gene *RP11-61G19.1* (OR[A]: 1.18,  $p=2.71 \times 10^{-8}$ ) and rs368467810 on chromosome 6 near gene *HIST1H3L* (OR[TTTA]: 0.96,  $p=3.88 \times 10^{-8}$ ).

**Conclusions** Within the current limitations, such as self-reported phenotype and older age of the study population, four loci were detected for periodontitis that have not previously been linked with this condition. Further exploration of the function of these loci may contribute to improved understanding of periodontitis aetiology and subsequent drug development.

**Clinical relevance** These findings offer new targets for future research to investigate the genetic impact on periodontitis and aid the future understanding of periodontitis pathology and the disease's progression.

Keywords Periodontal disease · Genome · Genetic · GWAS

# Introduction

Periodontitis is a common inflammatory condition that affects the tissue surrounding and supporting the teeth [1]. Periodontitis is recognised as a major cause of tooth loss [1] and is associated with various long-term chronic conditions such as diabetes [2], cognitive impairment and dementia [3, 4], and cardiovascular disease [5], resulting in a worsening quality of life [6] and increased risk of mortality [7, 8]. Periodontitis affects 20–50% of the global population [9] with severe cases impacting more than 1 billion adults worldwide [10]. The pathology and aetiology of periodontitis are multifaceted, which includes the complex interplay of microorganisms, pathogens, environmental factors such as nutritional intake and smoking, and genetic factors [11-13]. While there have been numerous experimental studies into the role of microorganisms, pathogens and environmental factors in the development of periodontitis, there are also emerging studies exploring the genetic influences with 65 genes putatively associated with periodontitis [13].

Recent analyses have produced estimates of the heritability of periodontitis ranging from 7 to 38% [14]. To date, 16 genome-wide association studies (GWAS) of various forms of periodontitis have been conducted. 15 GWASs identified in the previous systematic review [15], and another GWAS of periodontitis published recently [16]. Here, 12 SNPs were collectively found that reached the conventional level for genome-wide statistical significance ( $p < 5 \times 10^{-8}$ ).

Extended author information available on the last page of the article

However, no common loci have been identified across these studies, even amongst those of similar ethnicity [15]. This may be partly due to heterogeneity across studies and small sample size. Improved knowledge of the genetics of periodontitis could potentially benefit not only our understanding of disease aetiology, but also risk assessment and personalised treatment plans [17].

Further GWAS investigations using larger samples are warranted. The aim of this study, therefore, was to use data from the UK Biobank (UKB) to identify the genetic risk factors for periodontitis in a population of European descent.

### Method

#### Sample and data resources

UKB [18] is a prospective cohort study of over 500,000 participants aged from 40 to 69 years at initial recruitment (2006–2010). A wide range of demographic and health-related variables and lifestyle data were recorded for participants through questionnaires, physical examination, imaging and genotyping at baseline. UKB continued to follow-up participants' health-related outcomes through continued contact and access to medical records [19] (https: //www.ukbiobank.ac.uk/).

In the current study, the sample was drawn from the final release of UKB genetic data (Sample size n=488,377), out of which 409,548 (84% of total genotyped sample) persons have been classified as of European ancestry by integration of self-reported ethnicities and principal components analysis (PCA) results of population structure [18]. At the time of analysis, 139 UKB recruits had subsequently withdrawn their consent and so were excluded from this study.

The study report follows the STrengthening the REporting of Genetic Association Studies(STREGA)—An Extension of the STROBE Statement [20].

#### Phenotype: periodontitis

The UKB variable "self-reported dental health" (variable code: 100538) was used as a surrogate for periodontitis. Participants who reported "Painful gums", "bleeding gums", or "loose teeth" were classified as having periodontitis; whereas, participants who reported other or no dental health issue were classified as control. Previous validation study suggested that self-reported measure of periodontitis has acceptable validity as surrogate for periodontitis [21]. Participants choosing "prefer not to say" were coded as missing for this variable. The same definition has been used in previous UKB studies [7, 22] in the absence of a clinical diagnosis.

#### **Genotyping and imputation**

Detailed DNA extraction and quality filtering are described elsewhere (https://biobank.ctsu.ox.ac.uk/crystal/ukb/docs/u kb\_dna\_processing.pdf).

Two genotyping arrays were employed: (1) the Affymetrix UK Biobank Axiom array and (2) the Affymetrix UK BiLEVE Axiom Array (Fig. 1). The genotyped samples shared 95% common single nucleotide polymorphisms (SNP) and produced 825,927 SNPs for genotyping in total. The imputation of missing genotypes was implemented in IMPUTE4 (https://jmarchini.org/software/), which was developed from IMPUTE2 undertaken by UKB using genotype reference data from the Haplotype Reference Consortium (HRC). The UK10K and 1000G genotype reference panels were employed to impute SNPs not included in HRC. This study performed quality control filtering and genomewide analysis on Version 3 of the imputed data, containing 487,409 participants for 93,095,623 autosomal SNPs. Details of DNA extraction, genotyping and imputation are documented elsewhere [18].

UKB also provide the values of principal components (PC) for each participant to adjust for population stratification; as suggested by UKB, the first 15 are used in most analyses.

#### **Statistical analysis**

Sample and variant exclusions were based on the quality control metrics provided by UKB. Only participants of European ancestry were included in this study. For sample quality control, UKB performed heterozygosity checks by fitting a linear regression model with the first six principal components and outlined 968 outliers with unusually high heterozygosity rate and missing call rate >5%. The heterozygosity rate and missing call rate outliers were removed from the current study. Participants were also excluded for a discrepancy in biological sex and self-reported sex. Related pairs of participants were randomly removed one from every pair of participants with a kinship coefficient>0.08838835 [23] which was calculated and provided by UKB. Based on the imputation information (INFO) scores and minor allele frequency (MAF) information provided by UKB, an exclusion list containing SNPs with INFO score  $\leq 0.5$  was created. Duplicated SNPs were also removed before GWA analyses, resulting in 58,443,190 included in the final GWAS.

GWA analyses were conducted on autosomal chromosomes using a logistic regression model that assumed an additive mode of inheritance, implemented in PLINK 2.0 (Linux) (https://www.cog-genomics.org/plink/2.0/) [24]. Two distinct association analyses were performed: (1) An association analysis without any covariates, and (2) An **Fig. 1** Flow chart showing the sample size selection by applying the quality control filter and ethnicity filter. Sample characteristics were also shown. SD, Standard Deviation; QC, quality control



association analysis with age, sex and the first 15 principal components (PC1-15) included as covariates to adjust for baseline age, sex and population stratification. In the analysis, we defined the threshold for genome-wide significance as the conventional level of  $p < 5 \times 10^{-8}$  and a suggestive level of significance as  $p < 5 \times 10^{-6}$ . Manhattan plots and Q-Q plots were created in R 4.2.2 (https://www.r-proj ect.org/about.html), using the packages "qqman" [25] and "fastqq" (https://github.com/gumeo/fastqq). In addition, SNPs that were significant in previous GWASs reviewed by Gao, Iles [15] plus SNPs reported in [16] were also examined in our study to confirm replication of the results. The efficacy of genome-wide significant SNPs ( $p < 5 \times 10^{-8}$ ) in predicting periodontitis were evaluated using Receiver Operator Characteristic (ROC) curves [26] in the package "ROCR" in R4.2.2 [27].

#### **Functional annotation using FUMA 1.6.0**

SNPs with MAF>=0.001 from the model including covariates were followed up for functional annotation in FUMA v.1.6.1 [28] using function SNP2GENE. The default parameter in the SNP2GENE function was used with 1000 Genome phase 3 as the reference panel for its largest number of reference SNPs. SNPs with MAF>=0.05 were entered into FUMA as sensitivity analysis.

SNPs were declared as independent significant loci if they reached  $p < 5 \times 10^{-8}$  and were in no more than weak linkage disequilibrium (r<sup>2</sup><0.6). Lead SNPs at a locus were

further defined as the most significant amongst those with  $r^2 > 0.1$ . ANNOVAR (2017-07-17) was applied to estimate the functional consequence of SNPs. MAGMA (v1.08) was implemented to conduct gene-based tests, gene-set tests and gene-property tests with the Genotype-Tissue Expression (GTEx) v8 53 tissue types. Details of the function of FUMA and analysis of SNP2GENE are available in [28].

#### Results

Following initial quality control, 378 sex mismatches between biological sex and self-reported sex, 968 participants with high SNP missing rates (>5%) and unusually high heterozygosity rates as documented by UKB, plus 40,196 participants from related pairs were removed. Finally, 376,611 European participants (Mean age=57.6 years old, 53.76% women, 18.2% periodontitis cases) remained, of which 787 participants were dropped from the association analysis due to missing phenotype data. The sample characteristics are presented in Fig. 1.

# SNPs—periodontitis association: null model and covariates model

Manhattan plots are presented in Fig. 2 to show the association signals across the genome. In the null model (no covariates), 8 SNPs reached genome-wide significance  $(p < 5 \times 10^{-8})$  in relation to periodontitis (Table 1). After



Fig. 2 (a) Manhattan plot of the association analysis model without adjusting for any covariates, and (b) Corresponding Q-Q plot. (c) Association analysis model with adjustment for age, sex and first 1–15 PCs, and (d) Q-Q plots of corresponding model (GC lambda=0.98).

adjusting for 17 covariates (age, sex, PC1-15), three of the 8 SNPs remained genome-wide significant: rs775476621 (Odds Ratio<sub>[T]</sub>=3.08,  $p=1.01 \times 10^{-8}$ ) and rs751014048 (OR<sub>[G]</sub>=3.07,  $p=1.04 \times 10^{-8}$ ) on chromosome 11; rs149922301 (OR<sub>[A]</sub>=1.18,  $p=2.71 \times 10^{-8}$ ) on chromosome 4. An additional SNP also reached significance in the covariates model: rs368467810 (OR<sub>[TTTA]</sub>=0.96,  $p=3.88 \times 10^{-8}$ ) on chromosome 6. Overall, 674 SNPs (including four GWAS significant SNPs) reached our suggestive level of significance ( $p < 5 \times 10^{-6}$ ) in the covariates model, so may also contribute to periodontitis, and the results for these SNPs can be seen in Supplementary Table 1.

QQ plots compared the observed p-value and expected p-value under the null hypothesis providing an indication of inflation in the GWAS results. A genomic inflation factor



Note. The red line in (a) and (c) Manhattan plot represents the genomewide significant threshold ( $p < 5 \times 10^{-8}$ ); while the blue represents the suggestive line of significance ( $p < 5 \times 10^{-6}$ )

lambda=0.98 for both models indicated no inflation in the results.

To cross-validate the accuracy and performance of our GWA models, significant SNPs from both models and covariates (age, sex, first 15 PCs) were entered into the ROC curve estimate, respectively. Better performance and model accuracy were observed for the covariates model (Area under curve (AUC)=0.572) compared to the no covariates model (AUC=0.503). Figure 3 shows the ROC curve with area under curve for both association models.

In addition to the results from the current study, SNPs reported as associated with periodontitis in 13 previous studies reviewed by Gao, Iles [15], were also investigated. Based on the results from our GWAS with covariates, none of the previously reported SNPs reached  $p < 5 \times 10^{-8}$  in our study (Supplementary Table 2).

**Table 1** Association results for SNPs with  $p < 5 \times 10^{-8}$  for both models with and without covariates adjusted

RS ID	Chr: Pos	A1	A2	A1 frequency	Model	OR	95%CI	p
rc775476621	11, 24802004	т	٨	0 000240221	no covariato model	2.05	(2.09.4.47)	1.04v10-8
13773470021	11: 54692994	1	A	0.000249321		2.05	(2.00, 4.47)	1.04X10
rs751014048	11: 34991008	G	А	0.000250787	no covariate model	3.08	(2.08, 4.45)	1.09x10 <sup>-8</sup>
	11.0.001000				covariates model	3.07	(2.09, 4.51)	1.04x10 <sup>-8</sup>
rs185910248	5: 124199710	А	G	0.0148132	no covariate model	0.86	(0.82, 0.91)	1.78x10 <sup>-8</sup>
					covariates model	0.86	(0.82, 0.91)	6.37x10 <sup>-8</sup>
rs28381639	6: 31837774	G	А	0.0435935	no covariate model	1.08	(1.05, 1.11)	2.12x10 <sup>-8</sup>
					covariates model	1.08	(1.05, 1.11)	5.47x10 <sup>-8</sup>
rs149922301	4: 10809690	А	G	0.00955164	no covariate model	1.18	(1.12, 1.26)	2.62x10 <sup>-8</sup>
					covariates model	1.18	(1.12, 1.26)	2.71x10 <sup>-8</sup>
rs34465217	6: 31880807	С	CA	0.0620776	no covariate model	1.07	(1.05, 1.1)	4.02x10 <sup>-8</sup>
					covariates model	1.07	(1.05, 1.10)	5.99x10 <sup>-8</sup>
rs35664695	6: 31929808	т	С	0.0418201	no covariate model	1.08	(1.05, 1.11)	4.31x10 <sup>-8</sup>
					covariates model	1.08	(1.05, 1.11)	1.15x10 <sup>-7</sup>
rs41315812	6: 31943232	т	С	0.0416518	no covariate model	1.08	(1.05, 1.11)	4.76x10 <sup>-8</sup>
					covariates model	1.08	(1.05, 1.11)	1.25x10 <sup>-7</sup>
rs368467810	6: 27849721	TTTA	Т	0.359797	no covariate model	0.96	(0.95, 0.98)	1.59x10 <sup>-7</sup>
					covariates model	0.96	(0.95, 0.98)	3.88x10 <sup>-8</sup>

RS ID, SNP reference identifier; Chr: Pos, Chromosome: Position; A1, effect allele; A2, non-effect allele; OR, Odds Ratio; CI, Confidence Interval; p, p-value. Note. The SNPs significant in the covariates model were highlighted where dark green represents the SNPs with MAF >0.005, and light green represents MAF<0.00



#### **FUMA** analysis

In the covariates model, there were 18,849,429 SNPs with (MAF < 0.001) imported to FUMA for further downstream functional analysis. Two leading and independent genome-wide significant SNPs (rs149922301, rs368467810) were identified, both located in intergenic regions, where 132 genes mapped onto leading SNPs, 31 of which are pseudogenes. Four genes were annotated to the two leading SNPs: (1) rs149922301: *RP11-61G19.1* (distance=110405)

and *MIR572* (distance=560761); and (2) rs368467810: *HIST1H4L* (distance=8432) and *HIST1H3L* (distance=8472). The regional plot can be viewed in Supplementary Fig. 1.

In the gene set analysis, there were no genes found to be significant after Bonferroni correction, with p = 1 after Bonferroni correction for all gene sets (Supplementary Table 3). The MAGMA tissue expression analysis in this study found no significant association between gene expressed in all 53 tissue and periodontitis (Supplementary Fig. 2).

The results from the sensitivity analysis (MAF>=0.05) also showed no significant association with all 53 tissue and any gene-sets. Results can be viewed in supplementary file.

#### Discussion

The current study used UKB data to investigate the genetic risk variants for self-reported periodontitis and found four significant SNPs with two mapped onto gene *RP11-61G19.1* (rs149922301[A]), and *HIST1H3L* (rs368467810[TTTA]) associated with the self-reported periodontitis phenotype based on the fully adjusted model. The gene analysis of GWAS results prioritised 132 genes.

When compared to prior GWAS investigations of periodontitis, only two SNPs reported in previous studies reached nominal significance (p < 0.05/59 = 0.001) after Bonferroni correction in our results: rs12461706 [T] identified in Shungin, Haworth [17] (our study  $p=5.34 \times 10^{-4}$ ), and rs11084095 [A] identified in Munz, Richter [29] ( $p=4.58 \times 10^{-4}$ ) (Supplementary Table 2). None of the genome-wide significant ( $p < 5 \times 10^{-8}$ ) SNPs in our study were previously associated with periodontitis. Additionally, most of the previously reported SNPs did not achieve statistical significance (p < 0.05) in our GWAS.

There are several reasons for the lack of significance of the previously reported SNPs and why the SNPs highlighted here have not been reported before. One reason is the differences in the ethnic groups studied. For example, SNPs (rs2392520[C]  $p=4.17 \times 10^{-6}$ ) discovered in Japanese sample [30] were not statistically significant (p<0.05) in our sample (p=0.52). Similarly, most of the genome-wide significant ( $p<5 \times 10^{-8}$ ) and suggestive significant SNPs ( $p<5 \times 10^{-6}$ ) discovered in Korean sample from Hong, Shin [31] were not significant in this study. The two SNPs reached nominal significance both included European sample [17, 29], although there is sample overlap between current study and Shungin, Haworth [17].

A second reason is the limited sample size in most prior studies, which may not be sufficient for a GWAS to find "true risk/protective SNPs". For example, Petty, Silva [32] included only 879 mixed ethnicity participants (333 periodontitis cases) and did not find any genome-wide significant SNPs. Their suggestive SNPs (rs12800372 [C],  $p < 5 \times 10^{-6}$ ) did not reach statistical significance in our study. The two SNPs reached the nominal significance in the current study were from two large scale studies using European sample [17, 29], although the genome-wide significant SNP (rs12461706 [T]) from Shungin, Haworth [17] cannot be considered as replication due to overlapped sample used.

The third reason for the previously reported SNPs not reaching the same level of significance or nonsignificant in this study might be due to phenotype definition differences. For Shungin, Haworth [17], Munz, Richter [29] and the present study, although we all included European samples and had relatively larger sample sizes, different measures were used for periodontitis and varying case definitions employed. Shungin, Haworth [17] used mixed periodontitis measures and definitions as multiple datasets were included. Periodontitis measures included self-reported periodontitis, clinically diagnosed periodontitis and clinical examination measures, defined using criteria such as the CDC-AAP definition [33] and the 1999 international workshop for classification of periodontal diseases [34]. Munz, Richter [29] used mixed definitions as well which included the CDC-AAP definition [33] and definition based on bone loss ( $\geq 30\%$ ) or loss of attachment (>=4 mm) solely. Our study utilised a self-reported periodontitis definition encompassing bleeding gums, painful gums and loose teeth. This definition varied from Shungin, Haworth [17] who identified UKB participants with tooth loss only as a surrogate for periodontitis. These differences in case definitions are likely to have affected case and control classification in the final study population, thereby impacting the results.

Despite the differences, the significance  $(p < 10^{-3})$  observed on the rs12461706 [T] locus from Shungin, Haworth [17], and rs11084095 [A] from Munz, Richter [29] partially support the important role of gene *SIGLEC5*. Both rs12461706 [T] and rs11084095 [A] were annotated on *SIGLEC5* which functioning innate immune systems and contributes to periodontitis [35].

The function of the four SNPs (i.e., rs775476621, rs751014048, rs149922301, rs368467810) identified in this study are not fully understood, especially SNPs rs775476621 and rs751014048 which had no functionality associated with them using FUMA. There are four genes annotated for the other two identified SNPs: (1) rs149922301: RP11-61G19.1 (distance=110405) and Micro RNA 572 (MIR572) (distance=560761), and (2) rs368467810: HIST1H4L (distance=8432) and HIST1H3L (distance=8472). Gene MIR572 is attributed to microRNA class which are short non-coding regulatory RNAs involved in the expression of more than 60% of human genes [36, 37]. Previous studies have found MIR572 is upregulated in several types of cancer/malignancy such as non-small cell lung cancer [38], ovarian cancer [39], renal cell carcinoma [40], contributing to malignant development, poor prognosis, and shortened survival time. However, the regulatory role of MIR572 in periodontitis is not clear and requires further investigation to establish how changes in expression of MIR572 is related to periodontitis progression.

In addition to *MIR572, HIST1H4L* and *HIST1H3L* are histone genes within the family of protein coding genes for regulating the DNA binding. Histone modification is essential to many biological processes including brain development and mental illness [34] such as schizophrenia, bipolar disorder, autistic spectrum disorder and depression [41]. This corresponds to the previous findings on associations between periodontitis and mental illness [22], and raises the possibility that the mental illness and periodontitis share similar pathological pathways. However, further exploration of the exact biological impact these genes have on periodontitis development is still necessary and especially, little is known about the function of *RP11-61G19.1*. Future study of the function of these four genes could improve understanding of the pathology and underlying mechanisms of periodontitis.

According to the ROC curve, both of our association analysis models provided some prediction. These results suggest the discovered eight SNPs in null models and four SNPs together with covariates (i.e., age, sex and PC1-15) provide modest to poor performance on predicting periodontitis. The no covariate model has AUC=0.503 and covariate model has AUC=0.572 respectively. The possibility of low to moderate prediction and performance is that periodontitis is a multi-facetted disease involving environmental factors, microorganisms and genetic influences. The improvement in model prediction and performance by adding covariates (i.e., age, sex, PC1-15) again also suggests the importance of non-genetic factors in periodontitis.

The current study has several strengths. First, the use of high-quality genetic data with standardised quality control increases the reliability of the results. The sample size was larger than most previous GWAS investigations into periodontitis and drew from an ethnically homogeneous population, to maximise power and minimise the likelihood of false positive results. Second, our functional analysis suggested some promising avenues for further exploration of the genetic causes of periodontitis.

This study also has some limitations. Firstly, participants with missing data were not utilised in the association analysis, and we cannot assess whether they are missing at random. Secondly, this study used self-reported oral health status as a proxy for periodontitis, which is not as ideal as clinical measures. Additionally, our results are based on population with European ancestry, not including ethnicities minorities. Another limitation of this study may impact on the significance of findings is that the current study included an older population where the impact of lifestyle, comorbidities or any other potential risk factors may contribute more than genetic factor, while for younger population genetic variants may play a much stronger role in periodontitis. Therefore, interpretation with caution is needed.

# Conclusion

The current study has identified four significant loci associated with periodontitis in a European population using a large high-quality dataset from the UKB. Future studies are needed to further explore the loci identified here to better understand the pathology of periodontitis.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00784-0 25-06205-8.

**Acknowledgements** This article has not been published or being considered for publication elsewhere, in whole, in part, except as an abstract. This research has been conducted using the UK Biobank Resource under Application Number: 54633.

Author contributions CG contributed to data analysis, interpretation, and manuscript write-up. JK, HL [Larvin], MI, DTB contributed to data analysis. JK, JW, SP, DTB, MI and DB contributed to the conceptualizations, funding acquisition, study design and data interpretation. All the author (CG, JK, HL[Larvin], MI. DTB, SP, DB, HL[Luo], LN, BW, JW) contributed to the critically review the manuscript and final version approval and submission.

**Funding** Chenyi Gao is supported by the Alzheimer's Society (Alzheimer's Society Heather Corrie Ph.D. studentship: 546 (AS-PhD-19b-012)) and Barts Charity (MGU0504). The views expressed are those of the author(s) and not necessarily those of the Alzheimer's Society. The Alzheimer's Society did not involve in design and conduct this study, as well as data collection or preparation, review, approval, and decision to submit the manuscript. Jianhua Wu and Harriet Larvin are supported by Barts Charity (MGU0504). Mark Iles is supported in part by the National Institute for Health and Care Research (NIHR) Leeds Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

**Data availability** The SNPs reached suggestive significance level (5e-6) is provided within the manuscript or supplementary information files. UK Biobank data were used in the current study and application is required for access.

#### Declarations

Competing interests The authors declare no competing interests.

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

## References

- Eke PI et al (1970) Update of the Case Definitions for Population-Based Surveillance of Periodontitis. Journal of periodontology 2012. 83(12):1449–1454. https://doi.org/10.1902/jop.2012.1106 64
- Larvin H et al (2022) Periodontitis and risk of immune-mediated systemic conditions: a systematic review and meta-analysis. Commun Dent Oral Epidemiol. https://doi.org/10.1111/cdoe.128 12
- Larvin H et al (2023) The impact of study factors in the association of periodontal disease and cognitive disorders: systematic review and meta-analysis. Age Ageing 52(2):afad015. https://doi. org/10.1093/ageing/afad015
- Kang J et al (2020) Bidirectional relations between cognitive function and oral health in ageing persons: a longitudinal cohort study. Age Ageing 49(5):793–799. https://doi.org/10.1093/ageing /afaa025
- Larvin H et al (2021) Risk of incident cardiovascular disease in people with periodontal disease: a systematic review and metaanalysis. Clin Experimental Dent Res 7(1):109–122. https://doi.o rg/10.1002/cre2.336
- Mao W et al (2023) Experiences of discrimination and oral health-related quality of life among foreign-born older Chinese americans: does resilience play a mediating role? Community dentistry and oral epidemiology. 51(2):187–193. https://doi.org/ 10.1111/cdoe.12723
- Larvin H et al (2020) The impact of Periodontal Disease on Hospital Admission and Mortality during COVID-19 pandemic. Front Med 7:604980–604980. https://doi.org/10.3389/fmed.2020 .604980
- Larvin H et al (2024) All-cause and cause-specific mortality in US adults with periodontal diseases: a prospective cohort study. J Clin Periodontol 51(9):1157–1167. https://doi.org/10.1111/jcpe.1 4002
- Nazir MA (2017) Prevalence of periodontal disease, its association with systemic diseases and prevention. Int J Health Sci 11(2):72–80
- WHO. Oral health Periodontal (gum) disease (2023) [cited 2023 15/08/2023]
- Page RC et al (1997) Advances in the pathogenesis of periodontitis: summary of developments, clinical implications and future directions. Periodontol 2000 14(1):216–248. https://doi.org/10.11 11/j.1600-0757.1997.tb00199.x
- Preshaw PM et al (2011) Periodontitis and diabetes: a two-way relationship. Diabetologia 55(1):21–31. https://doi.org/10.1007/s 00125-011-2342-y
- Loos B.G., Van Dyke T.E. (2020) The role of inflammation and genetics in periodontal disease. Periodontol 2000 83(1):26–39. ht tps://doi.org/10.1111/prd.12297
- Nibali L et al (2019) What is the heritability of Periodontitis? A systematic review. J Dent Res 98(6):632–641. https://doi.org/10. 1177/0022034519842510
- Gao C et al (2024) Genome-Wide Association studies on periodontitis: a systematic review. PLoS One 19(9):e0306983. https:// /doi.org/10.1371/journal.pone.0306983
- 16. De Almeida SD 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. J Clin Periodontol 51(4):431–440. https://doi.org/10.1111/jcpe.13922
- Shungin D et al (2019) Genome-wide analysis of dental caries and periodontitis combining clinical and self-reported data. Nat Commun 10(1):2773. https://doi.org/10.1038/s41467-019-1063 0-1

- Bycroft C et al (2018) The UK Biobank resource with deep phenotyping and genomic data. Nat (London) 562(7726):203–209. h ttps://doi.org/10.1038/s41586-018-0579-z
- Sudlow C et al (2015) Biobank: an Open Access Resource for identifying the causes of a wide range of Complex diseases of Middle and Old Age. PLoS Med 12(3):e1001779–e1001779. http s://doi.org/10.1371/journal.pmed.1001779
- Little J et al (2009) STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. PLoS Med 6(2):e22–e22. https://doi.org/10.1371/jou rnal.pmed.1000022
- Abbood HM et al (2016) Validity of self-reported Periodontal Disease: a systematic review and Meta-analysis. J Periodontology (1970) 87(12):1474–1483. https://doi.org/10.1902/jop.2016 .160196
- Kang J et al (2022) Periodontal disease in people with a history of psychosis: results from the UK biobank population-based study. Commun Dent Oral Epidemiol. https://doi.org/10.1111/cdoe.127 98
- Manichaikul A et al (2010) Robust relationship inference in genome-wide association studies. Bioinformatics 26(22):2867– 2873. https://doi.org/10.1093/bioinformatics/btq559
- Chang CC et al (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4(1):7–7. htt ps://doi.org/10.1186/s13742-015-0047-8
- 25. Stephen, D.T., qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv, 2014: p. 005165. https://doi.org/10.1101/005165
- Fawcett T (2006) An introduction to ROC analysis. Pattern Recognit Lett 27(8):861–874. https://doi.org/10.1016/j.patrec.20 05.10.010
- Sing T et al (2005) ROCR: visualizing classifier performance in R. Bioinformatics 21(20):3940–3941. https://doi.org/10.1093/bio informatics/bti623
- Watanabe K et al (2017) Functional mapping and annotation of genetic associations with FUMA. Nat Commun 8(1):1826–1811. https://doi.org/10.1038/s41467-017-01261-5
- Munz M et al (2019) Meta-analysis of genome-wide association studies of aggressive and chronic periodontitis identifies two novel risk loci. Eur J Hum Genet 27(1):102–113. https://doi.org/1 0.1038/s41431-018-0265-5
- Shimizu S et al (2015) A genome-wide association study of periodontitis in a Japanese population. J Dent Res 94(4) 555–61. http s://doi.org/10.1177/0022034515570315
- Hong KW et al (2015) Genomewide association study on chronic periodontitis in Korean population: results from the Yangpyeong health cohort. J Clin Periodontol 42(8):703–710. https://doi.org/1 0.1111/jcpe.12437
- Petty LE et al (2023) Genome-wide association study identifies novel risk loci for apical periodontitis. J Endod. https://doi.org/10 .1016/j.joen.2023.07.018
- Page RC, Eke PI (2007) Case definitions for Use in Populationbased surveillance of Periodontitis. J Periodontol 78(7):1387– 1399. https://doi.org/10.1902/jop.2007.060264
- Armitage GC (1999) Development of a classification system for periodontal diseases and conditions. Ann Periodontol 4(1):1–6. h ttps://doi.org/10.1902/annals.1999.4.1.1
- Mueller R et al (2022) Periodontitis Risk variants at SIGLEC5 impair ERG and MAFB binding. J Dent Res 101(5):551–558. htt ps://doi.org/10.1177/00220345211049984
- Suer I et al (2019) MicroRNAs as prognostic markers in prostate cancer. Prostate 79(3):265–271. https://doi.org/10.1002/pros.237 31
- Seven M et al (2014) The role of miRNAs in cancer: from pathogenesis to therapeutic implications. Future Oncol 10(6):1027– 1048. https://doi.org/10.2217/fon.13.259

- Sun B, Zhao J, Shao ZY (2022) MiR-572 promotes the development of non-small cell lung cancer by targeting KLF2. Eur Rev Med Pharmacol Sci 26(9):3083–3090. https://doi.org/10.26355/e urrev\_202205\_28725
- Zhang X et al (2015) Upregulation of miR-572 transcriptionally suppresses SOCS1 and p21 and contributes to human ovarian cancer progression. Oncotarget 6(17):15180–15193. https://doi. org/10.18632/oncotarget.3737
- Pan X et al (2018) miR-566 functions as an oncogene and a potential biomarker for prognosis in renal cell carcinoma. Biomed

# **Authors and Affiliations**

Pharmacother 102:718–727. https://doi.org/10.1016/j.biopha.20 18.03.072

41. Wu Y et al (2020) Multi-trait analysis for genome-wide association study of five psychiatric disorders. Translational Psychiatry 10(1):209–209. https://doi.org/10.1038/s41398-020-00902-6

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Chenyi Gao<sup>1,2</sup> · Mark M. Iles<sup>3,4</sup> · David Timothy Bishop<sup>5</sup> · Harriet Larvin<sup>1</sup> · David Bunce<sup>6</sup> · Bei Wu<sup>7</sup> · Huabin Luo<sup>8</sup> · Luigi Nibali<sup>10</sup> · Susan Pavitt<sup>2</sup> · Jianhua Wu<sup>1</sup> · Jing Kang<sup>9</sup>

🖂 Jing Kang

j.kang@kcl.ac.uk Chenyi Gao c.gao@qmul.ac.uk

Mark M. Iles m.m.iles@leeds.ac.uk

David Timothy Bishop d.t.bishop@leeds.ac.uk

Harriet Larvin h.larvin@qmul.ac.uk

David Bunce d.bunce@leeds.ac.uk

Bei Wu bei.wu@nyu.edu

Huabin Luo LUOH@ECU.EDU

Luigi Nibali luigi.nibali@kcl.ac.uk

Susan Pavitt s.pavitt@leeds.ac.uk Jianhua Wu

jianhua.wu@qmul.ac.uk

- <sup>1</sup> Wolfson Institute of Population Health, Queen Mary University of London, London, UK
- <sup>2</sup> School of Dentistry, University of Leeds, Leeds, UK
- <sup>3</sup> Leeds Institute for Data Analytics, University of Leeds, Leeds, UK
- <sup>4</sup> NIHR Leeds Biomedical Research Centre, Leeds Teaching Hospitals NHS Trust, Leeds, UK
- <sup>5</sup> Leeds Institute of Medical Research, School of Medicine, University of Leeds, Leeds, UK
- <sup>6</sup> School of Psychology, University of Leeds, Leeds, UK
- <sup>7</sup> Rory Meyers College of Nursing, New York University, New York, US
- <sup>8</sup> Department of Public Health, East Carolina University, Greenville, US
- <sup>9</sup> Oral Clinical Research Unit, Faculty of Dentistry Oral Craniofacial Sciences, King's College London, London, UK
- <sup>10</sup> Periodontology Unit, Centre for Host Microbiome Interactions, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College, London, UK