Original Article

Causal relationship between cathepsins and major salivary gland neoplasms: a bidirectional Mendelian randomization study

Shiyong Zhuang^, Haoran Ding, Hanyao Huang, Tianyi Wang, Chengyan Li, Xingzhi Zeng, Yi Li

State Key Laboratory of Oral Diseases & National Clinical Research Center for Oral Diseases & Department of Head and Neck Oncology Surgery, West China Hospital of Stomatology, Sichuan University, Chengdu, China

Contributions: (I) Conception and design: Y Li; (II) Administrative support: Y Li; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: S Zhuang, H Ding, T Wang, C Li, X Zeng, H Huang; (V) Data analysis and interpretation: S Zhuang, H Ding, T Wang, C Li, X Zeng, H Huang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Yi Li, PhD. State Key Laboratory of Oral Diseases & National Clinical Research Center for Oral Diseases & Department of Head and Neck Oncology Surgery, West China Hospital of Stomatology, Number 14, Unit 3, Renmin Nan Road, Chengdu 610041, China. Email: Liyi1012@163.com.

Background: Observational studies have suggested a potential link between cathepsins and major salivary gland neoplasms (MSGNs), but the causality of this relationship remains uncertain. Mendelian randomization (MR) is a significant genetic method that employs single nucleotide polymorphisms (SNPs) as instrumental variables (IVs). This approach reduces confounding effects, enabling the analysis of causal relationships between exposure traits and outcome diseases. This study aimed to explore the causal links between cathepsins and MSGNs by utilizing MR analysis.

Methods: In this research, we collected IVs associated with 11 different types of cathepsins (including cathepsins D, L1, B, E, F, G, H, O, S, L2, and Z) from the Medical Research Council (MRC) integrative epidemiology unit (IEU) open genome-wide association studies (GWAS) database. Data for cathepsins D and L1 were sourced from the SCALLOP consortium, which included 21,758 Europeans identified via the Olink proximity extension assay (PEA). Cathepsins B, E, F, G, H, O, S, L2, and Z were obtained from the INTERVAL study involving 3,301 European participants using the SOMAscan assay. We also collected data on benign major salivary gland neoplasms (BMSGNs) from the FinnGen database, consisting of 3,353 cases and 450,380 controls, and information on major salivary gland carcinomas (MSGCs) from the UK Biobank, which included 105 cases and 456,243 controls. Diagnostic criteria for both BMSGNs and MSGCs followed the international statistical classification of diseases and related health problems 10th revision (ICD-10) classification. A comprehensive bidirectional MR study was executed employing diverse methodologies, including inverse variance weighted (IVW), MR-Egger regression, weighted median, and weighted mode. Additionally, sensitivity analyses were conducted to emphasize the solidity of the study.

Results: Increased levels of cathepsin F (CTSF), cathepsin O (CTSO), and cathepsin L2 (CTSL2) were associated with a higher risk of BMSGNs (CTSF: IVW: P=0.01, odds ratio (OR) =1.12, CTSO: IVW: P=0.02, OR =1.14; CTSL2: IVW: P=0.01, OR =1.17). Additionally, no causal association was found between cathepsins and MSGCs. Reverse MR analyses did not establish a causal relationship between BMSGNs and various cathepsins. However, it did reveal that a higher risk of MSGCs was associated with lower levels of CTSL2 (IVW: P=0.01, beta =−0.046).

Conclusions: The study presents compelling evidence of a correlation between elevated CTSF, CTSO, and CTSL2 levels and an increased risk of BMSGNs. Elevated CTSF, CTSO, and CTSL2 levels may serve as significant biomarkers for diagnosing BMSGNs definitively. Conversely, reduced levels of CTSL2 provide a novel foundation for diagnosing MSGCs and differentiating them from BMSGNs. Moreover, CTSF,

^ ORCID: 0009-0005-9242-7224.

CTSO, and CTSL2 represent potential new targets for therapeutic intervention in BMSGNs and MSGCs.

Keywords: Cathepsin; major salivary gland neoplasms; Mendelian randomization

Submitted Aug 27, 2024. Accepted for publication Nov 07, 2024. Published online Nov 26, 2024.
doi: 10.21037/gs-24-374
View this article at: https://dx.doi.org/10.21037/gs-24-374

Introduction

Salivary gland tumors are relatively uncommon, comprising only 3–10% of head and neck tumors (1,2). They are primarily of epithelial origin and tend to develop in the major salivary glands, such as the parotid, submandibular, and sublingual glands (3). Specifically, 64–80% of these tumors occur in the parotid glands, 7–11% in the submandibular glands, and less than 1% in the sublingual glands (1). Based on clinical and pathological features, salivary gland epithelial tumors can be categorized as benign or malignant. Benign tumors commonly include pleomorphic adenoma, myoepithelioma, basal cell adenoma, and Warthin's tumor. Malignant tumors typically include acinic cell carcinoma, mucoepidermoid carcinoma, and adenoid cystic carcinoma (4). While approximately 80% of salivary gland neoplasms are benign, they still pose a risk of recurrence and potential malignant transformation due to their localized invasive characteristics. Pathology data from the Netherlands between 1992 and 2012 indicated a 4.6% first-recurrence rate after at least five years of follow-up and a 6.7% recurrence rate at 20 years of follow-up for salivary gland pleomorphic adenoma. Notably, 3.2% of all recurrences progressed into malignant tumors (5). A recent analysis of treatment efficacy and prognostic factors for carcinoma ex pleomorphic adenoma (CXPA) of the major salivary glands revealed an overall treatment failure rate ranging from 33.3% to 53.0%. Furthermore, the 5-year overall survival rate is only between 30% and 76%. These results highlight the significant need for treatment effectiveness and patient prognosis improvement (6). Salivary gland carcinomas (SGCs) constitute a highly diverse group of tumors, comprising 5–7% of all head and neck cancers (3). Initial symptoms of salivary gland neoplasms (SGNs) may not be immediately noticeable and often manifest as painless swelling in the gland. As the disease progresses, the tumor continues to grow, and malignant tumors in various parts of the salivary glands can result in symptoms such as facial paralysis and pain. In some cases, the cancer can even spread to other parts of the body, significantly impacting the prognosis (7). Many studies have explored the prediction of overall and tumor-specific survival in patients with major salivary gland carcinomas (MSGCs) through nomograms, which have proven effective in estimating patient outcomes. However, a notable scarcity of research on early diagnosis and treatment for these conditions remains (8). The timely identification and subsequent management of SGNs are of utmost significance. It is imperative to expeditiously explore novel diagnostic biomarkers and therapeutic targets in this context. Hence, our study focused on major salivary gland

Highlight box

Key findings

- This study explored the relationship between cathepsins and major salivary gland neoplasms (MSGNs) from genetic insights by using Mendelian randomization (MR) analysis.
- Elevated serum levels of cathepsin F (CTSF), cathepsin O (CTSO), and cathepsin L2 (CTSL2) were connected with an increased risk of benign major salivary gland neoplasms (BMSGNs).
- The development of major salivary gland carcinomas (MSGCs) was associated with lower levels of CTSL2.

What is known and what is new?

- An increasing number of genome-wide association studies (GWAS) have yielded numerous genetic variants associated with complex diseases and traits. MR is a robust approach that utilizes single nucleotide polymorphisms (SNPs) as genetic instrumental variables (IVs) to examine the causal relationship between a specific trait exposure and a disease outcome.
- Our manuscript provided a comprehensive summary of the MR analysis, utilizing genetic data from the FinnGen Biobank, UK Biobank, and GWASs. This analysis aimed to investigate, for the first time, the potential causal relationship between 11 different cathepsins and MSGNs.

What is the implication, and what should change now?

- The research findings have established a direct genetic causal relationship between specific cathepsins and MSGNs. These results underscore the potential of cathepsins as targets for therapy and as diagnostic biomarkers for managing salivary gland neoplasms, offering new research directions for early diagnosis and treatment.

neoplasms (MSGNs), which happened most frequently, to identify diagnostic markers and new treatment targets for the early detection and treatment of both benign major salivary gland neoplasms (BMSGNs) and MSGCs.

Proteases are crucial for both physiological and pathological processes. They are categorized into seven main families: metalloproteinases, serine proteases, threonine proteases, asparagine proteases, aspartate proteases, glutamate proteases, and cysteine proteases (9). Cysteine proteases encompass cathepsin L (L1), B, H, O, S, T, K, L2 (V), and F. Cathepsins, extensively studied in mammals, play a pivotal role in maintaining intracellular and extracellular balance and are linked to tissue differentiation, intracellular protein degradation, hormone maturation, antigen processing, immune responses, and the malignant metastasis of tumors (10,11). Prior research has established a strong link between heightened cathepsin K (CTSK) expression and numerous malignancy indicators in SGCs, including tumor infiltration, lymph node metastasis, distant metastasis, advanced tumor node metastasis (TNM) clinical stage, increased risk of recurrence, and reduced survival rates (12). Furthermore, cathepsin D (CTSD) expression was notably elevated in malignant salivary gland tumors compared to benign tumors. Its increased expression in highly malignant tumors, such as adenoid cystic carcinoma and high-grade mucoepidermoid carcinoma, is regarded as a significant marker of tumor invasiveness, potentially contributing to the perineural invasion of salivary adenoid cystic carcinoma (13-15). These findings emphasized the potential causal relationship between cathepsins and SGNs.

Randomized controlled trials (RCTs) may not always be feasible when investigating the causal relationship between diseases due to trial conditions or ethical considerations. In such instances, using genetic variation as an instrumental variable (IV) in Mendelian randomization (MR) can yield more reliable results that closely resemble those of RCTs than observational or retrospective studies. MR is a robust approach that utilizes single nucleotide polymorphisms (SNPs) as genetic instrumental variables (IVs) to investigate the causal relationship between a specific trait and a disease outcome. The theory behind MR analysis suggests that if these genetic variants are associated with the disease outcome through their impact on the exposure, individuals carrying these variants are more likely to develop the disease (16). Compared to observational studies, MR reduces confounding by using genetic variants consistent from conception and mitigates biases from environmental or lifestyle factors commonly observed in observational studies. Various cathepsin-targeting inhibitors have been developed and utilized in treating clinical diseases in recent decades, showing promising therapeutic effects (17). As a result, exploring the relationship between cathepsins and MSGNs presents a significant opportunity for enhancing treatment strategies for these tumors. Current research on cathepsins in relation to MSGNs is limited, predominantly consisting of retrospective studies that examine the direct association between cathepsin expression levels in tissue samples and SGNs. A primary challenge in this field is that diagnosing salivary gland tumors largely relies on histopathological methods. Additionally, the assessment of cathepsin diversity and the measurement of circulating cathepsin levels are time-consuming and labor-intensive processes, further complicated by the small sample sizes often encountered in studies of salivary gland tumors. This challenge may be addressed through the MR method, which employs genetic IVs to investigate the causal relationship between cathepsins and MSGNs from a genetic standpoint. This approach can potentially provide valuable insights for future basic and clinical research while reducing the time and effort required for subsequent studies. Additionally, genome-wide association studies (GWASs) have significantly improved the methodology by identifying numerous genetic variants associated with complex diseases and traits, expanding the pool of instrumental SNPs for more reliable MR analyses. In this specific context, we conducted two-sample bidirectional MR analyses to investigate the potential causal relationships of different cathepsin types on the risk of MSGNs. This manuscript is written in accordance with the STROBE-MR reporting checklist (available at https://gs.amegroups.com/article/view/10.21037/gs-24-374/rc).

Methods

Study design

In this study, a bidirectional two-sample Mendelian randomization was conducted to systematically evaluate the genetic associations between 11 cathepsins (D, L1, B, E, F, G, H, O, S, L2, and Z) and both BMSGNs and MSGCs. To ensure the study design's credibility, a series of statistical methods were used to validate the results, with the primary approach being inverse variance weighted (IVW) (18,19). The instrumental variables (IVs) chosen for this study were based on relevance, independence, and exclusion restriction, which are three fundamental Mendelian assumptions: (I) genetic variables must significantly correlate with the

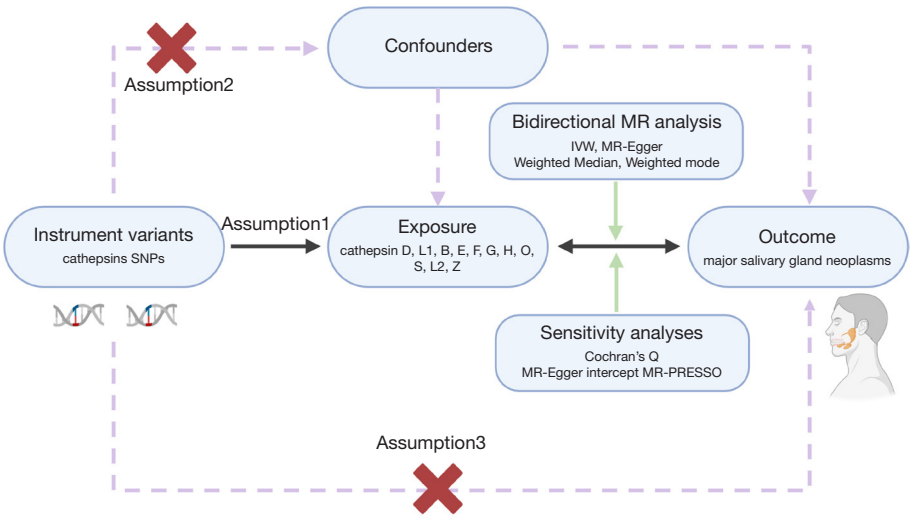

Figure 1 MR design for causal analysis of cathepsins and MSGNs on genetic predisposition. MSGNs, major salivary gland neoplasms; MR, Mendelian randomization; SNPs, single nucleotide polymorphisms; IVW, inverse variance weighted; MR-PRESSO, MR Pleiotropy Residual Sum and Outlier. Created with BioRender.com.

exposure, (II) genetic variables must be independent of confounders, (III) genetic variables can only influence the outcome by affecting the exposure and are not directly associated with the outcome (20,21). The study design is shown in *Figure 1*.

Data source

The summary data for various cathepsins, including D, L1, B, E, F, G, H, O, S, L2, and Z, utilized in this study were sourced from two primary datasets: the Systematic and Combined Analysis of Olink Proteins (SCALLOP) consortium and the INTERVAL study. These datasets are accessible online through the Medical Research Council (MRC) integrative epidemiology unit (IEU) open GWAS database [IEU OpenGWAS project (mrcieu.ac.uk)]. Data for cathepsin D and L1 were obtained from the SCALLOP consortium, which comprised a European population of 21,758 individuals, measured using the Olink proximity extension assay (PEA) technique. In contrast, cathepsins B, E, F, G, H, O, S, L2, and Z were drawn from the INTERVAL study, which included a sample of 3,301 individuals from the European population and employed an expanded version of an aptamer-based multiplex protein assay known as SOMAscan (22,23). All participants provided informed consent, and the National Research Ethics Service granted approval for the original studies from which the summary data were derived. In the context of our study, neither written informed consent nor ethical approval was required. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Meanwhile, the data on BMSGNs were obtained from the FinnGen database R11 version, which consisted of 3,031 cases and 409,150 controls, with 20,094,122 SNPs. Additionally, MSGCs data were acquired from the UK Biobank, which included 105 cases and 456,243 controls, totaling 11,831,932 SNPs (24). The diagnostic criteria for BMSGNs and MSGCs adhered to the international statistical classification of diseases and related health problems, 10th revision (ICD-10), which encompasses both benign and malignant tumors of the parotid, submandibular, and sublingual salivary glands. The phenotypes utilized were accessible online on the FinnGen (FinnGen: an expedition into genomics and medicine | FinnGen) and GWAS Catalog websites [GWAS Catalog (ebi.ac.uk)] (Table S1).

Selection of Ivs

In this study, we chose to apply a more lenient threshold of P<5e−6 to ensure an adequate number of SNPs linked to cathepsins for further analysis (25). In addition, we implemented a clump distance of less than 10,000 kb with r^2<0.001 to mitigate potential linkage disequilibrium

(LD) and excluded palindromic SNPs to minimize bias (26,27). Furthermore, in adherence to the independence assumption of MR, we carefully examined and excluded SNPs strongly associated with the outcome (P<5e–8) to ensure the accuracy of our analysis. We determined the strength of each selected SNP by calculating the *F*-statistic using the formula $F=(\text{beta}/\text{se})^2$ (28,29). An F-statistic of 10 or higher provides substantial evidence against weak instrument bias. Conversely, SNPs with an F-statistic below 10 are considered weak and are recommended for exclusion (30-32). In reverse MR analysis, where cathepsins were the outcome and BMSGN and MSGCs were the exposures, genetic variants instrumental in BMSGN and MSGCs were selected using the same criteria mentioned previously (the supplementary tables are avalable at https://cdn.amegroups.cn/static/public/gs-24-374-1.xlsx; https://cdn.amegroups.cn/static/public/gs-24-374-2.xlsx; https://cdn.amegroups.cn/static/public/gs-24-374-3.xlsx; https://cdn.amegroups.cn/static/public/gs-24-374-4.xlsx).

Mendelian randomization analysis

We employed a range of methods, such as IVW, MR-Egger, weighted median, and weighted mode with the Two-Sample MR package (version 0.5.6) and MR Pleiotropy Residual Sum and Outlier (MR-PRESSO) in R 4.3.3, to conduct a bidirectional MR analysis to investigate the potential causal relationship between cathepsins and BMSGN and MSGCs. The IVW method demonstrated the highest statistical power and was selected as the primary analytical approach for estimating the overall causal effects (18). Although the results from MR-Egger, weighted median, and weighted mode analyses were comparatively less efficient, they remained valuable by providing significant insights and contributing to a comprehensive assessment of the consistency and reliability of the study results (33-35). Statistically, a p value less than 0.05 was considered significant. Moreover, a comprehensive sequence of sensitivity analyses was performed, including Cochran's Q test, MR-Egger intercept test, and MR-PRESSO test. The Cochrane's Q test was used to evaluate the heterogeneity of SNP estimates, with statistical significance set at P<0.05 (19). The MR-Egger intercept test assessed horizontal pleiotropy (P<0.05) (33). In addition, MR-PRESSO analysis was conducted to identify and address potential horizontal pleiotropic outliers (36). Sensitivity analysis is a crucial test that must be conducted on the resulting data to affirm the reliability and precision of the findings.

Results

MR main analysis results

A forward MR analysis investigated the effects of 11 cathepsins (cathepsins D, L1, B, E, F, G, H, O, S, L2, and Z) of MSGNs. The study found that elevated levels of cathepsin F (CTSF), cathepsin O (CTSO), and cathepsin L2 (CTSL2) are associated with a higher risk of BMSGNs (CTSF: IVW: P=0.01, odds ratio (OR) =1.12, 95% confidence interval (CI): 1.02–1.22; CTSO: IVW: P=0.02, OR =1.14, 95% CI: 1.02–1.28; CTSL2: IVW: P=0.01, OR =1.17, 95% CI: 1.03–1.33). Although the methods other than IVW did not show statistical differences, all four methods had a consistent direction of effect, ensuring the credibility of the results (*Figure 2*). No evidence of heterogeneity or horizontal pleiotropy was observed in the sensitivity analysis. The values obtained for Cochran's Q (CTSF: MR-Egger P=0.33, CTSO: MR-Egger P=0.62; CTSL2: MR-Egger P=0.62) and MR-Egger intercept (CTSF:P=0.97, CTSO: P=0.39; CTSL2: P=0.54) tests all exceeded a P value of 0.05, verifying the credibility and reliability of the results. In MR-PRESSO tests, CTSF (P=0.43), CTSO (P=0.66), and CTSL2 (P=0.72) all had P values more than 0.05, further indicating no horizontal pleiotropy (*Figure 3*). However, no causal association was found between cathepsins and MSGCs (*Figure 4*).

Reverse MR analysis results

Reverse MR analyses were conducted to thoroughly investigate the potential for reverse causality, in which cathepsins were the outcomes and BMSGN and MSGCs were the exposures. Our analyses did not establish a causal relationship between BMSGNs and various cathepsins (*Figure 5*). However, our findings revealed an association between the development of MSGCs and reduced levels of CTSL2 (IVW: P=0.01, beta=–0.046, 95% CI: –0.089 to –0.003). Similarly, the weighted median method yielded comparable results (P=0.007, beta=–0.042, 95% CI: –0.073 to –0.012) (*Figure 2*). The reverse analyses showed no heterogeneity or horizontal pleiotropy in Cochran's Q (MR-Egger P=0.45) and MR-Egger intercept (P=0.46) tests. Additionally, the MR-PRESSO (P=0.56) analysis revealed no outliers. Notably, the analysis of MSGCs with cathepsin S (CTSS) produced a negative IVW result (P=0.95). However, Cochran's Q test yielded a P value of 0.006, indicating the presence of heterogeneity. The MR-PRESSO assay identified an outlier, rs141662525, which was removed

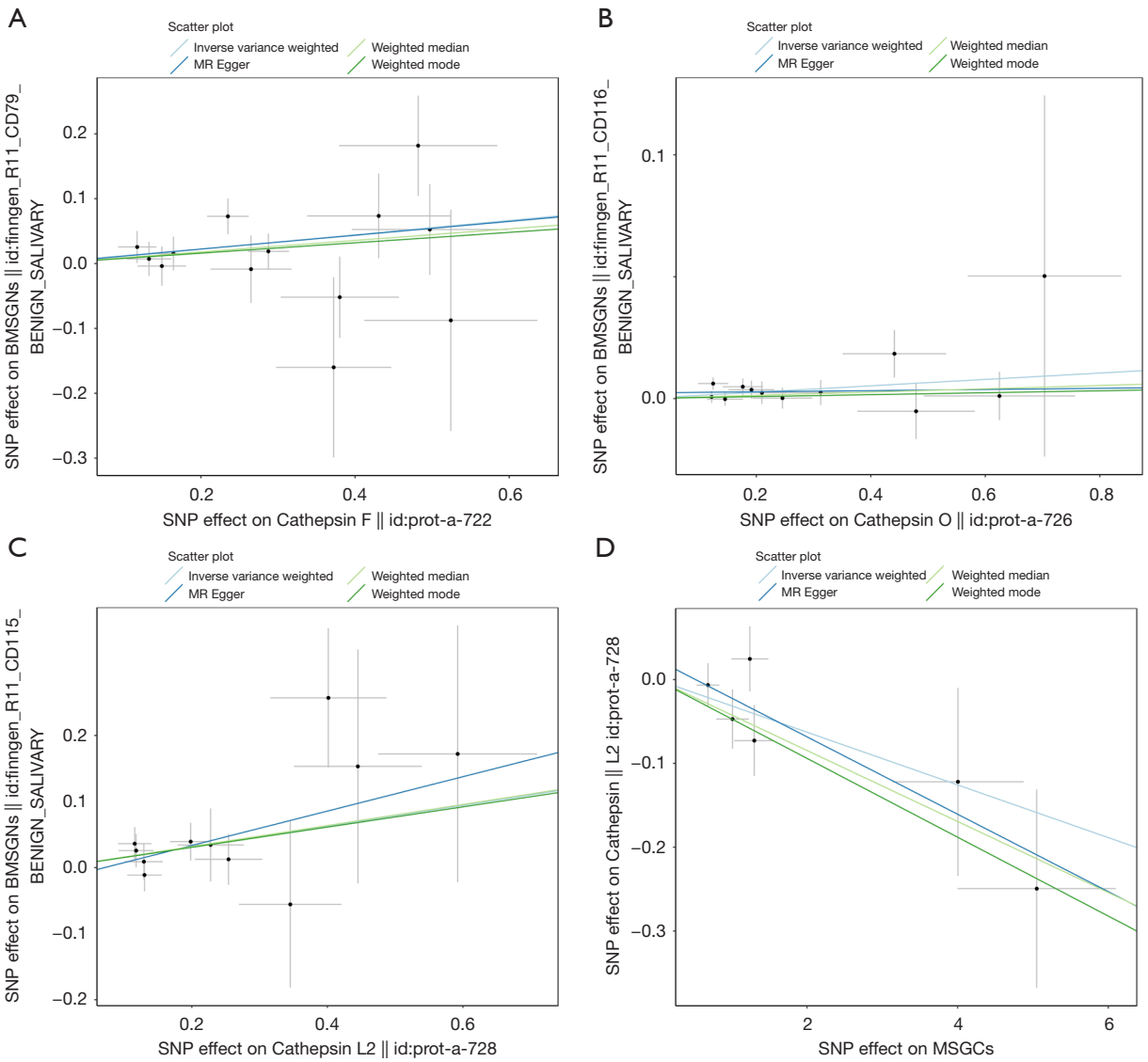

Figure 2 The MR results between cathepsins and MSGNs. (A) Scatter plot showing the causal effect of cathepsin F on BMSGNs. (B) Scatter plot showing the causal effect of cathepsin O on BMSGNs. (C) Scatter plot showing the causal effect of cathepsin L2 on BMSGNs. (D) Scatter plot showing the causal effect of MSGCs on cathepsin L2. MR, Mendelian randomization; BMSGNs, benign major salivary gland neoplasms; MSGCs, major salivary gland carcinomas; SNPs, single nucleotide polymorphisms.

for reanalysis. After removing the outlier, the MR-PRESSO test did not detect any outliers. The MR-PRESSO test (P=0.14) and the MR-Egger intercept (P=0.78) indicated the absence of horizontal pleiotropy. Although Cochran's Q test (P=0.049) still suggested potential heterogeneity, it enhanced the reliability of the negative IVW result (P=0.25), indicating the absence of a potential causal relationship between MSGCs and CTSS. All IVW results were negative for the other cathepsins, and sensitivity analysis did not indicate heterogeneity and horizontal pleiotropy (*Figure 6*).

Discussion

SGNs display a wide range of cytomorphological features, and it is challenging to distinguish between benign and malignant neoplasms due to their diversity and heterogeneity (37). With the increasing occurrence of SGNs and the potential of distant metastasis in malignant

Exposure	Outcome	nSNP	Method	OR(95% CI)	P value	Cochran's Q P	MR−Egger intercept P	MR−PRESSO P
Cathepsin D	BMSGNs	19	MR Egger	0.96 (0.75 to 1.22)	0.73	0.42	0.96	0.51
			IVW	0.96 (0.84 to 1.11)	0.58			
			Weighted median	1.07 (0.88 to 1.31)	0.51			
			Weighted mode	1.07 (0.88 to 1.30)	0.52			
Cathepsin L1	BMSGNs	30	MR Egger	1.29 (0.95 to 1.76)	0.11	0.30	0.24	0.29
			IVW	1.10 (0.96 to 1.25)	0.18			
			Weighted median	1.13 (0.94 to 1.35)	0.19			
			Weighted mode	1.10 (0.88 to 1.36)	0.41			
Cathepsin B	BMSGNs	19	MR Egger	0.89 (0.70 to 1.13)	0.35	0.06	0.32	0.07
			IVW	0.99 (0.90 to 1.10)	0.91			
			Weighted median	0.99 (0.87 to 1.12)	0.84			
			Weighted mode	1.01 (0.88 to 1.16)	0.91			
Cathepsin E	BMSGNs	9	MR Egger	0.94 (0.80 to 1.10)	0.44	0.78	0.26	0.62
			IVW	1.01 (0.92 to 1.12)	0.78			
			Weighted median	1.00 (0.88 to 1.14)	0.96			
			Weighted mode	0.95 (0.79 to 1.14)	0.61			
Cathepsin F	BMSGNs	13	MR Egger	1.11 (0.90 to 1.38)	0.36	0.33	0.97	0.43
			IVW	1.12 (1.02 to 1.22)	0.01			
			Weighted median	1.09 (0.96 to 1.24)	0.16			
			Weighted mode	1.08 (0.92 to 1.28)	0.36			
Cathepsin G	BMSGNs	13	MR Egger	0.86 (0.67 to 1.10)	0.25	0.07	0.65	0.12
			IVW	0.90 (0.79 to 1.02)	0.10			
			Weighted median	0.87 (0.75 to 1.00)	0.06			
			Weighted mode	0.86 (0.74 to 1.00)	0.07			
Cathepsin H	BMSGNs	11	MR Egger	1.00 (0.93 to 1.08)	0.93	0.97	0.82	0.99
			IVW	1.00 (0.94 to 1.05)	0.93			
			Weighted median	1.00 (0.94 to 1.06)	0.97			
			Weighted mode	1.00 (0.94 to 1.06)	0.96			
Cathepsin O	BMSGNs	12	MR Egger	1.02 (0.79 to 1.33)	0.86	0.62	0.39	0.66
			IVW	1.14 (1.02 to 1.28)	0.02			
			Weighted median	1.07 (0.92 to 1.24)	0.38			
			Weighted mode	1.04 (0.85 to 1.27)	0.70			
Cathepsin S	BMSGNs	23	MR Egger	1.17 (1.03 to 1.33)	0.02	0.91	0.06	0.57
			IVW	1.06 (0.98 to 1.13)	0.14			
			Weighted median	1.13 (1.02 to 1.24)	0.02			
			Weighted mode	1.17 (1.05 to 1.31)	0.01			
Cathepsin L2	BMSGNs	11	MR Egger	1.30 (0.92 to 1.83)	0.17	0.62	0.54	0.72
			IVW	1.17 (1.03 to 1.33)	0.01			
			Weighted median	1.17 (0.99 to 1.39)	0.06			
			Weighted mode	1.17 (0.93 to 1.47)	0.22			
Cathepsin Z	BMSGNs	13	MR Egger	0.97 (0.84 to 1.12)	0.65	0.19	0.87	0.28
			IVW	0.96 (0.88 to 1.04)	0.33			
			Weighted median	0.98 (0.88 to 1.09)	0.69			
			Weighted mode	0.99 (0.89 to 1.11)	0.91			

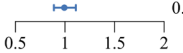

Figure 3 The forest plot of the causal effect of cathepsins on BMSGNs. BMSGNs, benign major salivary gland neoplasms; MR, Mendelian randomization; SNP, single nucleotide polymorphism; IVW, inverse variance weighted; OR, odds ratio; CI, confidence interval; MR-PRESSO, MR Pleiotropy Residual Sum and Outlier.

neoplasms, there is a mounting need to uncover new early diagnostic biomarkers and develop innovative treatments. These efforts are critical for the effective prognosis and treatment of the disease (4). In MR studies conducted previously, a substantial body of evidence has supported a causal relationship between cathepsins and a range of diseases, including cardiovascular diseases (38), Parkinson's disease (39), knee and hip osteoarthritis (25), skin cancers (40), and gastrointestinal tumors (41) from a genetic perspective. Hence, we performed a two-sample bidirectional Mendelian randomization analysis using genetic data from the FinnGen Biobank, UK Biobank,

Exposure	Outcome	nSNP	Method	OR(95% CI)	P value	Cochran's Q P	MR−Egger intercept P	MR−PRESSO P
Cathepsin D	MSGCs	14	MR Egger	0.71 (0.22 to 2.32)	0.59	0.39	0.52	0.54
		14	IVW	0.52 (0.25 to 1.09)	0.08			
		14	Weighted median	0.49 (0.21 to 1.18)	0.11			
		14	Weighted mode	0.53 (0.20 to 1.37)	0.21			
Cathepsin L1	MSGCs	27	MR Egger	2.64 (0.48 to 14.46)	0.27	0.64	0.15	0.60
		27	IVW	0.82 (0.40 to 1.67)	0.58			
		27	Weighted median	1.05 (0.40 to 2.77)	0.93			
		27	Weighted mode	0.97 (0.33 to 2.87)	0.95			
Cathepsin B	MSGCs	18	MR Egger	1.23 (0.35 to 4.36)	0.76	0.14	0.93	0.22
		18	IVW	1.17 (0.70 to 1.95)	0.56			
		18	Weighted median	1.38 (0.73 to 2.63)	0.33			
		18	Weighted mode	1.37 (0.71 to 2.64)	0.37			
Cathepsin E	MSGCs	10	MR Egger	1.55 (0.21 to 11.34)	0.68	0.17	0.56	0.23
		10	IVW	0.90 (0.38 to 2.17)	0.82			
		10	Weighted median	1.03 (0.35 to 3.03)	0.95			
		10	Weighted mode	1.29 (0.27 to 6.25)	0.76			
Cathepsin F	MSGCs	12	MR Egger	0.29 (0.07 to 1.24)	0.13	0.29	0.09	0.14
		12	IVW	1.02 (0.54 to 1.95)	0.94			
		12	Weighted median	1.07 (0.46 to 2.49)	0.88			
		12	Weighted mode	1.25 (0.43 to 3.58)	0.69			
Cathepsin G	MSGCs	12	MR Egger	1.74 (0.39 to 7.83)	0.49	0.46	0.25	0.43
		12	IVW	0.77 (0.38 to 1.55)	0.46			
		12	Weighted median	1.18 (0.43 to 3.23)	0.75			
		12	Weighted mode	1.61 (0.30 to 8.71)	0.59			
Cathepsin H	MSGCs	8	MR Egger	0.90 (0.56 to 1.43)	0.67	0.73	0.74	0.81
		8	IVW	0.85 (0.60 to 1.19)	0.35			
		8	Weighted median	0.82 (0.56 to 1.20)	0.32			
		8	Weighted mode	0.81 (0.57 to 1.16)	0.29			
Cathepsin O	MSGCs	11	MR Egger	0.72 (0.14 to 3.78)	0.71	0.84	0.39	0.84
		11	IVW	1.44 (0.71 to 2.90)	0.31			
		11	Weighted median	1.40 (0.54 to 3.64)	0.49			
		11	Weighted mode	1.57 (0.32 to 7.64)	0.59			
Cathepsin S	MSGCs	21	MR Egger	1.33 (0.67 to 2.61)	0.42	0.29	0.19	0.28
		21	IVW	0.92 (0.60 to 1.39)	0.68			
		21	Weighted median	1.02 (0.60 to 1.72)	0.95			
		21	Weighted mode	1.04 (0.60 to 1.78)	0.90			
Cathepsin L2	MSGCs	8	MR Egger	1.61 (0.23 to 11.50)	0.65	0.45	0.80	0.59
		8	IVW	1.27 (0.55 to 2.96)	0.57			
		8	Weighted median	1.14 (0.37 to 3.55)	0.82			
		8	Weighted mode	1.10 (0.22 to 5.53)	0.91			
Cathepsin Z	MSGCs	11	MR Egger	1.48 (0.74 to 2.99)	0.30	0.78	0.05	0.36
		11	IVW	0.80 (0.50 to 1.29)	0.36			
		11	Weighted median	1.04 (0.56 to 1.92)	0.90			
		11	Weighted mode	1.14 (0.65 to 2.01)	0.65			

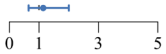

Figure 4 The forest plot of the causal effect of cathepsins on MSGCs. MSGCs, major salivary gland carcinomas; MR, Mendelian randomization; SNP, single nucleotide polymorphism; IVW, inverse variance weighted; OR, odds ratio; CI, confidence interval; MR-PRESSO, MR Pleiotropy Residual Sum and Outlier.

and IEU Open GWAS database to thoroughly investigate potential causal relationships between 11 different cathepsins and MSGNs for the first time to elucidate these relationships systematically. Our findings revealed that elevated CTSF, CTSO, and CTSL2 levels were risk factors for BMSGNs. Additionally, in the reverse MR analysis, MSGCs were found to decrease CTSL2 levels. While CTSD and CTSK have previously been shown to be strongly associated with the malignancy of SGNs (12,14,42), our study did not find a direct causal relationship between SGNs and CTSD levels. This lack of association may be due to complex genetic-environmental interactions.

Exposure	Outcome	nSNP	Method	Beta(95% CI)	P value	Cochran's Q P	MR−Egger intercept P	MR−PRESSO P
BMSGNs	Cathepsin D	13	MR Egger	−0.019 (−0.069 to 0.032)	0.31	0.14	0.41	0.15
		13	IVW	−0.075 (−0.213 to 0.064)	0.47			
		13	Weighted median	−0.027 (−0.088 to 0.034)	0.39			
		13	Weighted mode	−0.007 (−0.122 to 0.108)	0.90			
BMSGNs	Cathepsin L1	13	MR Egger	0.004 (−0.045 to 0.053)	0.59	1.00	0.60	1.00
		13	IVW	0.035 (−0.091 to 0.162)	0.88			
		13	Weighted median	0.002 (−0.057 to 0.060)	0.96			
		13	Weighted mode	−0.007 (−0.104 to 0.090)	0.89			
BMSGNs	Cathepsin B	12	MR Egger	0.001 (−0.106 to 0.108)	0.28	0.89	0.24	0.82
		12	IVW	−0.151 (−0.410 to 0.108)	0.99			
		12	Weighted median	0.016 (−0.123 to 0.155)	0.82			
		12	Weighted mode	0.058 (−0.167 to 0.282)	0.63			
BMSGNs	Cathepsin E	12	MR Egger	−0.100 (−0.210 to 0.010)	0.09	0.42	0.25	0.43
		12	IVW	−0.248 (−0.509 to 0.014)	0.08			
		12	Weighted median	−0.084 (−0.230 to 0.063)	0.26			
		12	Weighted mode	−0.054 (−0.264 to 0.155)	0.62			
BMSGNs	Cathepsin F	12	MR Egger	−0.053 (−0.185 to 0.079)	0.68	0.08	0.90	0.12
		12	IVW	−0.072 (−0.408 to 0.263)	0.43			
		12	Weighted median	−0.085 (−0.241 to 0.071)	0.29			
		12	Weighted mode	−0.102 (−0.345 to 0.142)	0.43			
BMSGNs	Cathepsin G	12	MR Egger	−0.081 (−0.202 to 0.040)	0.83	0.22	0.42	0.25
		12	IVW	0.034 (−0.262 to 0.330)	0.19			
		12	Weighted median	−0.111 (−0.261 to 0.040)	0.15			
		12	Weighted mode	−0.075 (−0.305 to 0.155)	0.54			
BMSGNs	Cathepsin H	12	MR Egger	0.011 (−0.096 to 0.118)	0.47	0.47	0.38	0.47
		12	IVW	−0.098 (−0.357 to 0.160)	0.84			
		12	Weighted median	0.036 (−0.115 to 0.188)	0.64			
		12	Weighted mode	0.156 (−0.137 to 0.449)	0.32			
BMSGNs	Cathepsin O	12	MR Egger	−0.069 (−0.195 to 0.057)	0.09	0.26	0.16	0.17
		12	IVW	−0.271 (−0.559 to 0.016)	0.28			
		12	Weighted median	−0.040 (−0.201 to 0.122)	0.63			
		12	Weighted mode	0.097 (−0.225 to 0.420)	0.57			
BMSGNs	Cathepsin S	12	MR Egger	−0.081 (−0.209 to 0.047)	0.95	0.12	0.64	0.19
		12	IVW	−0.010 (−0.330 to 0.311)	0.22			
		12	Weighted median	−0.087 (−0.241 to 0.066)	0.27			
		12	Weighted mode	−0.092 (−0.310 to 0.126)	0.43			
BMSGNs	Cathepsin L2	12	MR Egger	−0.030 (−0.137 to 0.077)	0.16	0.60	0.18	0.49
		12	IVW	−0.202 (−0.461 to 0.057)	0.58			
		12	Weighted median	−0.005 (−0.148 to 0.139)	0.95			
		12	Weighted mode	0.116 (−0.128 to 0.360)	0.37			
BMSGNs	Cathepsin Z	12	MR Egger	0.007 (−0.109 to 0.122)	0.73	0.25	0.67	0.34
		12	IVW	−0.052 (−0.342 to 0.238)	0.91			
		12	Weighted median	0.001 (−0.151 to 0.153)	0.99			
		12	Weighted mode	−0.010 (−0.263 to 0.243)	0.94			

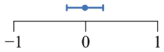

Figure 5 The forest plot of the reverse causal effect of BMSGNs on cathepsins. BMSGNs, benign major salivary gland neoplasms; MR, Mendelian randomization; SNP, single nucleotide polymorphism; IVW, inverse variance weighted; OR, odds ratio; CI, confidence interval; MR-PRESSO, MR Pleiotropy Residual Sum and Outlier.

Limitations in genetic data have impeded further validation of CTSK with salivary gland malignancies using MR.

Strengths and limitations: our study has provided valuable genetic insights into the causal relationships between cathepsins and MSGNs. However, it is essential to acknowledge the study's limitations: (I) the threshold for obtaining sufficient SNPs was P<5e−6 rather than the more stringent P<5e−8. (II) The study focused solely on a European population, highlighting the necessity of including more diverse populations to enhance the study's validity. (III) MSGNs encompass a wide range of diseases, each with distinct manifestations, indicating the need for the study to be further tailored to specific diseases. Additionally, the sample size of MSGCs is relatively small,

Exposure	Outcome	nSNP	Method	Beta(95% CI)	P value	Cochran's Q P	MR−Egger intercept P	MR−PRESSO P
MSGCs	Cathepsin D	6	MR Egger	0.008 (−0.004 to 0.020)	0.45	0.40	0.14	0.26
		6	IVW	−0.009 (−0.029 to 0.012)	0.20			
		6	Weighted median	0.007 (−0.005 to 0.020)	0.25			
		6	Weighted mode	0.005 (−0.012 to 0.021)	0.60			
MSGCs	Cathepsin L1	6	MR Egger	0.000 (−0.011 to 0.011)	0.70	0.88	0.64	0.94
		6	IVW	−0.004 (−0.025 to 0.016)	0.98			
		6	Weighted median	−0.002 (−0.015 to 0.012)	0.81			
		6	Weighted mode	−0.002 (−0.018 to 0.014)	0.80			
MSGCs	Cathepsin B	6	MR Egger	0.023 (−0.001 to 0.047)	0.64	0.86	0.54	0.90
		6	IVW	0.011 (−0.032 to 0.054)	0.06			
		6	Weighted median	0.017 (−0.015 to 0.048)	0.30			
		6	Weighted mode	0.016 (−0.020 to 0.052)	0.43			
MSGCs	Cathepsin E	6	MR Egger	−0.019 (−0.043 to 0.005)	0.48	0.66	0.94	0.82
		6	IVW	−0.017 (−0.060 to 0.026)	0.13			
		6	Weighted median	−0.014 (−0.044 to 0.015)	0.34			
		6	Weighted mode	−0.013 (−0.049 to 0.024)	0.53			
MSGCs	Cathepsin F	6	MR Egger	0.000 (−0.024 to 0.024)	0.99	0.81	0.98	0.91
		6	IVW	−0.000 (−0.043 to 0.043)	0.99			
		6	Weighted median	−0.003 (−0.033 to 0.026)	0.84			
		6	Weighted mode	−0.005 (−0.045 to 0.034)	0.80			
MSGCs	Cathepsin G	6	MR Egger	0.012 (−0.012 to 0.036)	0.56	0.92	0.91	0.97
		6	IVW	0.014 (−0.029 to 0.057)	0.34			
		6	Weighted median	0.009 (−0.021 to 0.039)	0.55			
		6	Weighted mode	0.009 (−0.032 to 0.050)	0.69			
MSGCs	Cathepsin H	6	MR Egger	0.005 (−0.019 to 0.029)	0.50	0.54	0.30	0.52
		6	IVW	−0.016 (−0.059 to 0.027)	0.68			
		6	Weighted median	0.004 (−0.029 to 0.037)	0.80			
		6	Weighted mode	−0.003 (−0.045 to 0.039)	0.90			
MSGCs	Cathepsin O	6	MR Egger	−0.021 (−0.046 to 0.003)	0.23	0.33	0.57	0.48
		6	IVW	−0.033 (−0.079 to 0.013)	0.09			
		6	Weighted median	−0.032 (−0.064 to −0.000)	0.05			
		6	Weighted mode	−0.035 (−0.079 to 0.008)	0.17			
MSGCs	Cathepsin S	5	MR Egger	−0.023 (−0.064 to 0.017)	0.53	0.049	0.78	0.14
		5	IVW	−0.038 (−0.140 to 0.065)	0.25			
		5	Weighted median	−0.018 (−0.063 to 0.027)	0.43			
		5	Weighted mode	−0.002 (−0.065 to 0.060)	0.95			
MSGCs	Cathepsin L2	6	MR Egger	−0.031 (−0.056 to −0.007)	0.10	0.45	0.46	0.56
		6	IVW	−0.046 (−0.089 to −0.003)	0.01			
		6	Weighted median	−0.042 (−0.073 to −0.012)	0.007			
		6	Weighted mode	−0.047 (−0.090 to −0.004)	0.09			
MSGCs	Cathepsin Z	6	MR Egger	0.012 (−0.012 to 0.036)	0.50	0.94	0.84	0.97
		6	IVW	0.016 (−0.027 to 0.059)	0.32			
		6	Weighted median	0.006 (−0.022 to 0.035)	0.66			
		6	Weighted mode	0.003 (−0.039 to 0.045)	0.89			

Figure 6 The forest plot of the reverse causal effect of MSGCs on cathepsins. MSGCs, major salivary gland carcinomas; MR, Mendelian randomization; SNPs, single nucleotide polymorphisms; IVW, inverse variance weighted; OR, odds ratio; CI, confidence interval; MR-PRESSO, MR Pleiotropy Residual Sum and Outlier.

and more cases need to be validated. (IV) Cathepsin levels and the prevalence of MSGNs may be influenced by gender and age. However, the lack of individual genotypic expression data precludes further statistical analyses. (V) The current study only explored the genetic perspective. Further research is necessary to provide specific guidance for clinical application.

CTSF is a critical lysosomal protein degradation system component derived from a skeletal muscle cDNA library. Its expression is prominent in the heart, skeletal muscle, brain, testis, and ovary but not in peripheral leukocytes and the thymus (43-45). While research on CTSF and SGNs is currently lacking, it has been implicated in regulating apoptosis in various cancers and utilized as a prognostic and diagnostic marker for numerous diseases. Elevated levels of CTSF and fibulin-1 have been identified as potential

innovative diagnostic biomarkers for brain metastases in non-small cell lung cancer (NSCLC) (46). Conversely, an alternative study has indicated that CTSF may exert an anti-tumor effect by modulating immune responses in NSCLC (47). Additionally, diminished CTSF expression predicted a poor prognosis in patients with clear-cell renal cell carcinoma and gastric cancer cells (48,49). Although CTSF has received less attention in benign tumors, our study identified it as a risk factor for BMSGNs. This might be connected to its role in protein degradation and modification of the extracellular matrix, which requires further investigation into the specific mechanisms involved.

CTSO was initially identified in the human breast cancer cDNA library in 1994. It is classified as a cysteine protease due to its ability to degrade synthetic cysteine protease peptide substrates. This enzyme is widely distributed in human tissues, with significant levels found in the ovary, kidney, liver, and placenta and lower levels in the thymus and skeletal muscle. It plays a crucial role in protein degradation (50). There are no direct studies on the correlation between CTSO and SGNs. However, it has been reported that human SGNs exhibit some similarities to breast tumors in terms of histology and steroid hormone receptor status (51). CTSO degrades the extracellular matrix and plays a significant role in the development of breast cancer (52). It is thought to influence hormones such as estrogen (ER), progesterone (Pg), and their receptors (53,54). Tamoxifen is commonly used to treat ERα-positive breast cancer. Increased CTSO expression reduces the BRCA1 transcription factor protein level through cysteine proteinase-mediated degradation, leading to tamoxifen resistance in ERα-positive breast cancer patients (55). Additionally, research has shown that the growth of salivary gland tumor cells can be inhibited by introducing Pg receptors and Pg treatment. It is important to note that the study referenced utilized a human adenoid cystic carcinoma cell line (56). Hence, the reason for CTSO being identified as a risk factor for BMSGNs in this MR analysis leads us to speculate that it may be linked to hormone receptors and hormones. This, in turn, affects tumor development and requires further exploration in future studies.

In murine pleomorphic adenomas, CD44 high tumor cells were discovered as tumor-initiating cells, with only 500 CD44 high tumor cells sufficient to induce pleomorphic adenomas in one-third of wild-type mice (57). The interaction of HA with CD44 triggers Na^+-H^+ exchange activity, leading to intracellular acidification and creating an acidic extracellular matrix environment. This environment facilitates hyaluronidase-2-mediated HA breakdown, HA modification, and cathepsin B activation, ultimately promoting invasion by breast tumor cells (58). In colorectal cancer, increased CD44 levels were associated with higher cathepsin D expression (59). In contrast, reducing CD44 inhibited cell proliferation, migration, and invasion in osteosarcoma and lowered cathepsin S expression (60). Thus, CD44 is closely related to cathepsins. Further studies are needed to demonstrate whether cathepsin expression affects tumor development by influencing CD44 expression. The demonstration of elevated levels of CTSF, CTSO, and CTSL2 as a risk factor for BMSGNs raises the question of whether this is related to the promotion of an increased number of CD44 high tumor cells. Further investigation and research are warranted to explore this potential correlation.

CTSL2, or cathepsin V (CTSV), was a cysteine protease identified and cloned from the human brain cDNA library. It was called cathepsin L2 due to its high similarity (78%) with cathepsin L. CTSL2 was mainly expressed in the thymus and testis, and it was also found in colorectal and breast cancer cell lines. Additionally, it was expressed in various tumors, including ovarian and renal cancers, and was involved in tumor progression (61). High expression of CTSL2 was strongly associated with the development of cone cornea, myasthenia gravis, cardiovascular disease, lung disease, and various malignant tumors, including colorectal carcinoma, kidney, ovary, endometrium, liver, and breast cancer (61-67). In CTSL2-deficient mice with squamous cell carcinoma, there was a significant increase in tumor progression and metastasis (68). In our study, elevated levels of CTSL2 might be a risk factor for BMSGNs due to enhanced extracellular matrix degradation in benign tumors. It is worth noting that reverse MR analysis revealed an association between the occurrence of MSGCs and decreased CTSL2 levels. The differing roles of CTSL2 in benign and malignant tumors of the major salivary glands may be attributed to its diverse functions in these environments. In malignant tumors, uncontrolled proliferation and apoptosis of tumor cells lead to rapid growth. The aberrant regulation of the pRB/E2F1 pathway is commonly associated with inappropriate proliferation and apoptosis in human cancers. CTSL2 was a novel E2F1 target involved in E2F1-dependent apoptosis. E2F1 was directly bound to the CTSL2 promoter, causing changes in lysosomal membrane permeability (LMP) and mitochondrial membrane depolarization to promote apoptosis. Therefore, CTSL2, as a target of

E2F1, plays a critical role in regulating apoptosis (69). In MSGCs, reducing CTSL2 levels might be closely linked to decreased tumor cell proliferation and apoptosis. CTSL2 can confidently be considered an oncogene, providing a potential target for future MSGCs treatment, pending further confirmation through additional studies.

The recent study produced encouraging results using various MR methods, and subsequent sensitivity analyses confirmed the reliability of the findings. Elevated CTSF, CTSO, and CTSL2 levels were identified as risk factors for BMSGNs. However, reverse analyses demonstrated that BMSGNs did not cause abnormalities in cathepsin levels. Therefore, elevated CTSF, CTSO, and CTSL2 levels could serve as early diagnostic biomarkers for BMSGNs. Inhibiting these cathepsins might reduce the incidence of BMSGNs, making them a potential new therapeutic target. Similarly, reduced CTSL2 levels could be a diagnostic biomarker for MSGCs. Medications designed to increase CTSL2 levels could promote apoptosis of MSGCs cells for therapeutic purposes. The differing levels of CTSL2 in benign and malignant SGNs suggested that further investigation is needed to understand the regulatory mechanism of CTSL2. The aforementioned findings have been derived from genetic analyses and should be considered for integration into clinical practice. Nevertheless, additional studies are imperative to investigate further and validate these findings.

Conclusions

In summary, CTSF, CTSO, and CTSL2 show promise as diagnostic biomarkers and potential therapeutic targets for the early detection, management, and treatment of MSGNs Notably, further basic and clinical research is needed to validate our findings for translation into clinical practice and to ultimately improve tumor diagnosis, define better prognostic categories, and develop new therapeutic regimens, especially for aggressive tumors.

Acknowledgments

We sincerely thank the MRC IEU Open GWAS project, the FinnGen database, and the UK Biobank for their invaluable contributions to making data available to the public. We are also thankful to the researchers who conducted the GWAS studies for their remarkable contributions and discoveries and their generosity in sharing their data. Their efforts have greatly expanded the scope of scientific discovery and collaboration, significantly advancing the field.

Funding: This work was supported by the National Natural Science Foundation of China (No. 81972546 to Y.L.).

Footnote

Reporting Checklist: The authors have completed the STROBE-MR reporting checklist. Available at https://gs.amegroups.com/article/view/10.21037/gs-24-374/rc

Peer Review File: Available at https://gs.amegroups.com/article/view/10.21037/gs-24-374/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://gs.amegroups.com/article/view/10.21037/gs-24-374/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Barnes L, Eveson JW, Reichart P, et al. Tumours of the salivary glands. World Health Organization Classification of Tumours: Pathology and Genetics of the Head and Neck Tumours. IARC; 2005:209-81.
2. Peel RL, Seethala RR. Pathology of salivary gland disease. In: Myers EN, Ferris RL, editors. Salivary Gland Disorders. Springer; 2007:33-104.
3. Dimitrijevic MV, Boricic IV, Tomanović NR, et al. Clinicopathological Analysis of 907 Major and Minor Salivary Gland Tumors. J Craniofac Surg 2022;33:e507-9.
4. Gatta G, Guzzo M, Locati LD, et al. Major and minor

salivary gland tumours. Crit Rev Oncol Hematol 2020;152:102959.
5. Valstar MH, de Ridder M, van den Broek EC, et al. Salivary gland pleomorphic adenoma in the Netherlands: A nationwide observational study of primary tumor incidence, malignant transformation, recurrence, and risk factors for recurrence. Oral Oncol 2017;66:93-9.
6. Chang YH, Kuo C, Chang TH, et al. Treatment outcome and prognostic factors analysis of carcinoma ex pleomorphic adenoma of major salivary glands. Ther Radiol Oncol 2023;7:14.
7. Geiger JL, Ismaila N, Beadle B, et al. Management of Salivary Gland Malignancy: ASCO Guideline. J Clin Oncol 2021;39:1909-41.
8. Guo Z, Wang Z, Liu Y, et al. Nomograms-based prediction of overall and cancer-specific survivals for patients diagnosed with major salivary gland carcinoma. Ann Transl Med 2021;9:1230.
9. Bond JS. Proteases: History, discovery, and roles in health and disease. J Biol Chem 2019;294:1643-51.
10. Bou-Samra P, Chang A, Guo E, et al. Cathepsin detection to identify malignant cells during robotic pulmonary resection. Transl Lung Cancer Res 2023;12:2370-80.
11. Patel S, Homaei A, El-Seedi HR, et al. Cathepsins: Proteases that are vital for survival but can also be fatal. Biomed Pharmacother 2018;105:526-32.
12. Elhendawy HA, Soliman S. Clinicopathological correlation of Cathepsin K expression in salivary gland carcinomas; relation to patients` outcome. Diagn Pathol 2023;18:66.
13. Vigneswaran N, Müller S, DeRose P, et al. Cathepsin-D and tumor associated antigen DF3 in salivary gland neoplasia. Differential diagnostic and prognostic applications. Pathol Res Pract 1994;190:1174-84.
14. Mohan M, Suresh R, Janardhanan M, et al. Immunohistochemical analysis of cathepsin-D in benign and malignant salivary gland neoplasms. Evaluation of its role as a prognostic indicator. J Oral Maxillofac Pathol 2022;26:589.
15. Zhang M, Wu JS, Yang X, et al. Overexpression Cathepsin D Contributes to Perineural Invasion of Salivary Adenoid Cystic Carcinoma. Front Oncol 2018;8:492.
16. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 2003;32:1-22.
17. Vizovišek M, Vidak E, Javoršek U, et al. Cysteine cathepsins as therapeutic targets in inflammatory diseases. Expert Opin Ther Targets 2020;24:573-88.
18. Burgess S, Scott RA, Timpson NJ, et al. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. Eur J Epidemiol 2015;30:543-52.
19. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol 2013;37:658-65.
20. Burgess S, Davey Smith G, Davies NM, et al. Guidelines for performing Mendelian randomization investigations: update for summer 2023. Wellcome Open Res 2019;4:186.
21. Skrivankova VW, Richmond RC, Woolf BAR, et al. Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization: The STROBE-MR Statement. JAMA 2021;326:1614-21.
22. Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome. Nature 2018;558:73-9.
23. Folkersen L, Gustafsson S, Wang Q, et al. Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. Nat Metab 2020;2:1135-48.
24. Jiang L, Zheng Z, Fang H, et al. A generalized linear mixed model association tool for biobank-scale data. Nat Genet 2021;53:1616-21.
25. Shao Z, Gao H, Han Q, et al. Genetic insights into serum cathepsins as diagnostic and therapeutic targets in knee and hip osteoarthritis. Sci Rep 2024;14:17553.
26. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife 2018;7:e34408.
27. 1000 Genomes Project Consortium; Auton A, Brooks LD, et al. A global reference for human genetic variation. Nature 2015;526:68-74.
28. Huang W, Xiao J, Ji J, et al. Association of lipid-lowering drugs with COVID-19 outcomes from a Mendelian randomization study. Elife 2021;10:e73873.
29. Lv X, Hu Z, Liang F, et al. Causal relationship between ischemic stroke and its subtypes and frozen shoulder: a two-sample Mendelian randomization analysis. Front Neurol 2023;14:1178051.
30. Palmer TM, Lawlor DA, Harbord RM, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. Stat Methods Med Res 2012;21:223-42.
31. Levin MG, Judy R, Gill D, et al. Genetics of height and risk of atrial fibrillation: A Mendelian randomization study. PLoS Med 2020;17:e1003288.
32. Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. Int J

Epidemiol 2011;40:740-52.
33. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol 2015;44:512-25.
34. Bowden J, Davey Smith G, Haycock PC, et al. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol 2016;40:304-14.
35. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol 2017;46:1985-98.
36. Verbanck M, Chen CY, Neale B, et al. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet 2018;50:693-8.
37. Rito M, Fonseca I. Salivary Gland Neoplasms: Does Morphological Diversity Reflect Tumor Heterogeneity. Pathobiology 2018;85:85-95.
38. Zeng R, Zhou Z, Liao W, et al. Genetic insights into the role of cathepsins in cardiovascular diseases: a Mendelian randomization study. ESC Heart Fail 2024;11:2707-18.
39. Yusufujiang A, Zeng S, Li H. Cathepsins and Parkinson's disease: insights from Mendelian randomization analyses. Front Aging Neurosci 2024;16:1380483.
40. Ma X, Zhuang H, Xu M, et al. Association between cathepsins and skin cancers: A bidirectional two-sample Mendelian randomization study. Skin Res Technol 2024;30:e13905.
41. Huang X, Deng H, Zhang B, et al. The causal relationship between cathepsins and digestive system tumors: a Mendelian randomization study. Front Oncol 2024;14:1365138.
42. Sivaranjini Y, Angadi PV, Kumar GS. Immuno-histochemical evaluation of Cathepsin D in malignant salivary gland carcinomas. Gulf J Oncolog 2012;(11):38-44.
43. Wang B, Shi GP, Yao PM, et al. Human cathepsin F. Molecular cloning, functional expression, tissue localization, and enzymatic characterization. J Biol Chem 1998;273:32000-8.
44. Wex T, Levy B, Wex H, et al. Human cathepsins F and W: A new subgroup of cathepsins. Biochem Biophys Res Commun 1999;259:401-7.
45. Santamaría I, Velasco G, Pendás AM, et al. Molecular cloning and structural and functional characterization of human cathepsin F, a new cysteine proteinase of the papain family with a long propeptide domain. J Biol Chem 1999;274:13800-9.
46. Wei S, Liu W, Xu M, et al. Cathepsin F and Fibulin-1 as novel diagnostic biomarkers for brain metastasis of non-small cell lung cancer. Br J Cancer 2022;126:1795-805.
47. Song L, Wang X, Cheng W, et al. Expression signature, prognosis value and immune characteristics of cathepsin F in non-small cell lung cancer identified by bioinformatics assessment. BMC Pulm Med 2021;21:420.
48. Ji C, Zhao Y, Kou YW, et al. Cathepsin F Knockdown Induces Proliferation and Inhibits Apoptosis in Gastric Cancer Cells. Oncol Res 2018;26:83-93.
49. Zhou X, Chen H, Huang D, et al. Reduced expression of cathepsin F predicts poor prognosis in patients with clear cell renal cell carcinoma. Sci Rep 2024;14:13556.
50. Velasco G, Ferrando AA, Puente XS, et al. Human cathepsin O. Molecular cloning from a breast carcinoma, production of the active enzyme in Escherichia coli, and expression analysis in human tissues. J Biol Chem 1994;269:27136-42.
51. Wick MR, Ockner DM, Mills SE, et al. Homologous carcinomas of the breasts, skin, and salivary glands. A histologic and immunohistochemical comparison of ductal mammary carcinoma, ductal sweat gland carcinoma, and salivary duct carcinoma. Am J Clin Pathol 1998;109:75-84.
52. Bond HM, Scicchitano S, Chiarella E, et al. ZNF423: A New Player in Estrogen Receptor-Positive Breast Cancer. Front Endocrinol (Lausanne) 2018;9:255.
53. Clarke CL, Sutherland RL. Progestin regulation of cellular proliferation. Endocr Rev 1990;11:266-301.
54. Horwitz KB. Mechanisms of hormone resistance in breast cancer. Breast Cancer Res Treat 1993;26:119-30.
55. Cairns J, Ingle JN, Wickerham LD, et al. SNPs near the cysteine proteinase cathepsin O gene (CTSO) determine tamoxifen sensitivity in ERα-positive breast cancer through regulation of BRCA1. PLoS Genet 2017;13:e1007031.
56. Yoshimura T, Sumida T, Liu S, et al. Growth inhibition of human salivary gland tumor cells by introduction of progesterone (Pg) receptor and Pg treatment. Endocr Relat Cancer 2007;14:1107-16.
57. Shen S, Yang W, Wang Z, et al. Tumor-initiating cells are enriched in CD44(hi) population in murine salivary gland tumor. PLoS One 2011;6:e23282.
58. Bourguignon LY, Singleton PA, Diedrich F, et al. CD44 interaction with Na+-H+ exchanger (NHE1) creates acidic microenvironments leading to hyaluronidase-2 and cathepsin B activation and breast tumor cell invasion. J Biol Chem 2004;279:26991-7007.
59. Ioachim E, Goussia A, Agnantis NJ. Glycoprotein

CD44 expression in colorectal neoplasms. An immunohistochemical study including correlation with cathepsin D, extracellular matrix components, p53, Rb, bcl-2, c-erbB-2, EGFR and proliferation indices. Virchows Arch 1999;434:45-50.
60. Kong L, Ji H, Gan X, et al. Knockdown of CD44 inhibits proliferation, migration and invasion of osteosarcoma cells accompanied by downregulation of cathepsin S. J Orthop Surg Res 2022;17:154.
61. Santamaría I, Velasco G, Cazorla M, et al. Cathepsin L2, a novel human cysteine proteinase produced by breast and colorectal carcinomas. Cancer Res 1998;58:1624-30.
62. Brookes NH, Loh IP, Clover GM, et al. Involvement of corneal nerves in the progression of keratoconus. Exp Eye Res 2003;77:515-24.
63. Tolosa E, Li W, Yasuda Y, et al. Cathepsin V is involved in the degradation of invariant chain in human thymus and is overexpressed in myasthenia gravis. J Clin Invest 2003;112:517-26.
64. Forouzanfar MH, Liu P, Roth GA, et al. Global Burden of Hypertension and Systolic Blood Pressure of at Least 110 to 115 mm Hg, 1990-2015. JAMA 2017;317:165-82.
65. Wolters PJ, Chapman HA. Importance of lysosomal cysteine proteases in lung disease. Respir Res 2000;1:170-7.
66. Skrzypczak M, Springwald A, Lattrich C, et al. Expression of cysteine protease cathepsin L is increased in endometrial cancer and correlates with expression of growth regulatory genes. Cancer Invest 2012;30:398-403.
67. Jing J, Wang S, Ma J, et al. Elevated CTSL2 expression is associated with an adverse prognosis in hepatocellular carcinoma. Int J Clin Exp Pathol 2018;11:4035-43.
68. Haider AS, Peters SB, Kaporis H, et al. Genomic analysis defines a cancer-specific gene expression signature for human squamous cell carcinoma and distinguishes malignant hyperproliferation from benign hyperplasia. J Invest Dermatol 2006;126:869-81.
69. Wong CH, Wu Z, Yu Q. CTSL2 is a pro-apoptotic target of E2F1 and a modulator of histone deacetylase inhibitor and DNA damage-induced apoptosis. Oncogene 2014;33:1249-57.

Cite this article as: Zhuang S, Ding H, Huang H, Wang T, Li C, Zeng X, Li Y. Causal relationship between cathepsins and major salivary gland neoplasms: a bidirectional Mendelian randomization study. Gland Surg 2024;13(11):2148-2162. doi: 10.21037/gs-24-374