Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Cathepsins and drug target genes in chronic gastritis: Deciphering potential therapeutic mechanisms via Mendelian randomization analysis

Lin Wang
The First Clinical College of Shanxi Medical University
Peifang Li
The First Clinical College of Shanxi University of Traditional Chinese Medicine
Jianwei Yuan
Department of Gastroenterology, the First People's Hospital of Xiaogan City
Yujia Feng
The First Clinical College of Shanxi University of Traditional Chinese Medicine
Dan Yang
The First Clinical College of Shanxi University of Traditional Chinese Medicine
Yanrong He
The First Clinical College of Shanxi University of Traditional Chinese Medicine
Tingquan Li
dghshanxi@163.com
The First Clinical College of Shanxi Medical University
Zheng Qi
The First Clinical College of Shanxi University of Traditional Chinese Medicine

Research Article

Keywords: Drug targets, Mendelian randomization, Genetics, Chronic gastritis, Cathepsins

Posted Date: June 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9701035/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Cathepsins and drug target genes in chronic gastritis: Deciphering potential therapeutic mechanisms via Mendelian randomization analysis

Lin Wang[1] Peifang Li[2] Jianwei Yuan[3] Yujia Feng[2] Dan Yang[2] Yanrong He[2] Tingquan Li[1]* Zheng Qi[2]*

[1]The First Clinical College of Shanxi Medical University,Taiyuan, People's Republic of China; [2]The First Clinical College of Shanxi University of Traditional Chinese Medicine,Taiyuan, People's Republic of China;[3]Department of Gastroenterology, the First People's Hospital of Xiaogan City, Xiaogan, People's Republic of China

Corresponding authors：Tingquan Li,Email: dghshanxi@163.com; Zheng Qi,Email: lqqizheng@126.com

Abstract

Background: Chronic gastritis is a highly prevalent gastrointestinal disorder characterized by persistent gastric mucosal inflammation. In approximately 10–30% of cases, this inflammation triggers a progression toward precancerous states, such as mucosal atrophy and intestinal metaplasia, which substantially increases the risk of gastric malignancy. Despite notable advances in diagnostic and therapeutic strategies, over 40% of patients fail to achieve adequate symptom relief.This therapeutic gap is driven by the disease's complex etiology and a critical lack of genetically validated targets.

Methods: To identify actionable therapeutic targets, we leveraged summary statistics from the largest genome-wide association studies (GWAS) to date: drug targets from the MiBioGen Consortium

(n=13,157) and chronic gastritis data from the FinnGen study (12,441 cases; 423,785 controls). Through bidirectional Mendelian randomization (MR), we evaluated the causal influence of potential drug targets on chronic gastritis. To further clarify the underlying biology, we employed two-step and multivariate MR to identify mediating metabolites that bridge these genetic signals to clinical outcomes.

Results: MR analysis identified 11 drug targets with genome-wide significance ($P<5\times10^{-8}$) across both cohorts, supported by robust colocalization evidence (PP.H4>75%). Functional annotation revealed these genes are tightly linked to innate immune signaling (e.g., NF-κB pathway, Toll-like receptor cascades) and mucosal barrier regulation. These findings suggest that the identified targets are not merely markers, but key drivers of the disease's pathogenesis.

Conclusions: This study nominates 11 high-priority targets for the treatment of chronic gastritis. By leveraging genetic evidence, which is known to double or triple the success rate of clinical trials, these results provide a rigorous foundation for drug development. Prioritizing these targets offers a strategic pathway to reduce research costs and improve outcomes for a high-burden patient population.

Keywords: Drug targets; Mendelian randomization; Genetics; Chronic gastritis; Cathepsins

Introduction

Chronic gastritis(CG) is defined as a persistent inflammatory state of the gastric mucosa that progresses to gastric gland loss, often accompanied by intestinal metaplasia (IM) or dysplasia, well-established precursor lesions in the cascade of gastric carcinogenesis[1]. The pathogenesis of CG is a multifaceted process involving a complex interplay of environmental and molecular signaling pathways, with *Helicobacter pylori* ($H. pylori$) infection serving as the primary etiological driver[2].

The clinical significance of CG lies in its intimate link to gastric cancer. The internationally validated Correa pathway, “normal mucosa → chronic inflammation → atrophy → intestinal metaplasia → dysplasia → gastric cancer,” illustrates this progression. Data suggest an annual gastric cancer incidence of 0.1% in patients with atrophic gastritis, rising to 0.25% in those with concurrent IM, and reaching 0.6% in high-grade intraepithelial neoplasia[3]. Globally, gastric cancer ranks 4th in incidence and 2nd in mortality among malignancies, with >90% of cases preceded by chronic gastritis, underscoring its role as a pivotal node for gastric cancer prevention and control[4].

As gastritis progresses toward atrophic states and potential malignancy, various protein families facilitate the transition. Specifically, matrix metalloproteinases (MMPs), tissue inhibitors of metalloproteinases (TIMPs), and "a disintegrin and metalloproteinases" (ADAMs) have been implicated in the disease’s progression[5]. These proteins regulate critical processes such as extracellular matrix remodeling, cell signaling, immune responses, angiogenesis, inflammation, and epithelial-mesenchymal transition, key events in mucosal damage and malignant transformation.

Among these molecular drivers, cysteine proteases—particularly cathepsin B (CATB) and cathepsin L (CATL)—have emerged as central mediators. Since ECM proteolysis is a rate-limiting step in cellular invasion[6] [7], the elevated activity of these proteases has been linked to various gastrointestinal malignancies. For instance, studies have demonstrated CATB’s pro-invasive activity in pancreatic cancer[8], while significantly higher levels of CATB and CATL have been observed in colorectal cancer tissues compared to healthy mucosa[9].

Despite these insights, establishing a definitive causal link between these targets and chronic gastritis in humans remains challenging due to confounding factors in observational studies. To overcome this, we employed Mendelian randomization (MR), a robust genetic epidemiological

approach that uses genetic variants as instrumental variables to infer causality [10]. By using single nucleotide polymorphisms (SNPs) associated with expression quantitative trait loci (eQTLs) as proxies for lifelong protein exposure, MR mimics the design of randomized controlled trials[11].

In this study, we integrated summary statistics from the largest available genome-wide association studies (GWAS) to date, covering 13,157 drug target genes and chronic gastritis data. By applying a bidirectional and two-step MR framework, we sought to dissect the causal relationships between these targets and CG, specifically investigating whether cathepsins serve as key mediating variables in the pathogenesis of the disease.

Materials and Methods

GWAS data of cathepsins and chronic gastritis

Cathepsin data were obtained from the open GWAS study published by Folkersen et al[13]. (https://gwas.mrcieu.ac.uk/). Cathepsins were used as exposures. Chronic gastritis data were retrieved from an open GWAS (https://gwas.mrcieu.ac.uk/), specifically from the FinnGen study, including 12,441 cases and 423,785 controls. GWAS data were used as exposures, with the same genetic instrument criteria: $P < 5 \times 10^{-8}$, $r^2 < 0.001$, MAF > 0.01, and kb < 10,000. The F-statistic $F = R^2(n-2)/(1-R^2)$ was used to measure the strength of each genetic instrument, where N represents the effective sample size of the GWAS used for SNP association.

Mendelian randomization (MR) analysis

In this study, MR analysis was first performed using the "TwoSampleMR" package, with 13,157 drug target genes and 11 cathepsins as exposure factors and cathepsins as outcome indicators, respectively. Inverse variance weighting (IVW) was employed as the main method to assess the causal relationship between exposure and outcome. Subsequently, cathepsins were used

as exposure factors and chronic gastritis as outcome indicators, and the causal relationship between them was similarly analyzed. MR-Egger regression, weighted median method, weighted mode method, Bayesian colocalization, and MR-Presso method were utilized for analysis. IVW was primarily used to assess the causal relationship between exposure factors and outcome factors. When the statistical P-value of MR-Egger is less than 0.05, heterogeneity between exposure and outcome can be considered. Weighted median and weighted mode were used to assess horizontal pleiotropy between exposure and outcome. The inverse variance weighted fixed-effects model was used to estimate the effect size and its 95% confidence interval.

Bayesian colocalization analysis

Bayesian colocalization analysis was used to assess the possibility that two traits share the same causal variable. The "coloc" package (https://github.com/chr1swallace/coloc) was used to verify the arguments in[14]. Bayesian colocalization provides posterior probabilities for five hypotheses: Hypothesis 0 (PPH0), Hypothesis 1 (PPH1), Hypothesis 2 (PPH2), Hypothesis 3 (PPH3), and Hypothesis 4 (PPH4). We verified the posterior probabilities of these hypotheses in the colocalization analysis of drug targets and chronic gastritis. SNPs were defined based on shared variants associated with the region. A gene was considered to have gene-based colocalization evidence if PPH4 > 75% as determined by at least one algorithm[15]-[16]. In this study, SNPs in the drug target instrumental variables were colocalized one by one with SNPs of chronic gastritis.

Phenotype Scanning

This study also performed a phenotypic scan to look for associations between previously identified SNPs and other traits through genome-wide association studies. Phenotypic scans were

performed using a phenotypic scanner[17]. SNPs that met the following criteria were considered pleiotropic: (i) they had significant genome-wide associations ($P < 5 \times 10^{-8}$); (ii) they were associated with corresponding GWAS in European ancestral populations; and (iii) these SNPs were associated with any known risk factors, such as gut microbiota, inflammatory cells, inflammatory proteins, low-density lipoprotein cholesterol, non-alcoholic fatty liver disease, acute pancreatitis, chronic gastritis, etc. In addition, we calculated the LD r^2 values between preferred SNPs to reveal potential associations.

Internal validation of drug targets eqtl, cathepsins, and chronic gastritis MR.

To further verify cathepsins as mediators and assess the existence of a direct causal relationship between exposure and outcome, this study performed MR analysis with drug target genes as exposure and chronic gastritis as outcome. The main analysis methods included IVW, MR-Egger, weighted median, and weighted mode.

Enrichment analysis

To explore the functional characteristics and biological relevance of the identified candidate therapeutic target genes, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were implemented via the R packages clusterProfiler[18] and Pathview[19]. The GO annotation system encompasses three core categories: biological process (BP), molecular function (MF), and cellular component (CC). Meanwhile, the KEGG pathway database provides detailed information on metabolic pathways and functional cascades involved in biological processes.

Protein interaction network construction

Analyzing protein-protein interaction (PPI) networks enables researchers to gain deeper insights into the intracellular interaction patterns among proteins[21]. In the present study, the PPI network was constructed with the STRING database, where a confidence score of 0.4 was set as the minimum required

interaction threshold, and all other parameters were retained as default[20]. Additionally, GeneMANIA (https://genemania.org/) was employed as a complementary tool for PPI network analysis[22].

Candidate drug prediction

Assessing protein-drug interactions is crucial for determining whether target genes can serve as actual drug targets. This study utilized the Drug Signatures Database (DSigDB, http://dsigdb.tanlab.org/DSigDBv1.0/) for this purpose. Specifically, DSigDB contains 22,527 gene sets and 17,389 different compounds involving 19,531 genes, making it a comprehensive database that links drugs and other chemicals to their target genes[23]. The identified target genes were uploaded to DSigDB, and candidate drugs were predicted to evaluate the medicinal activity of the target genes.

Data availability

The original study provides a summary statistic of genome-wide exposure[24][25]-[26]. Results data are available at: https://gwas.mrcieu.ac.uk/. All data are publicly available and do not require ethics committee review.

Result

Overview of Two-Sample MR Results

Using cis-eQTL data provided by the eQTLGen Consortium, we initially screened 13,157 genes identified as having potential for drug development. These genetic instruments were then analyzed against chronic gastritis summary statistics from the FinnGen database (12,441 cases; 423,785 controls). Our primary two-sample MR analysis identified 1880 genes with significant causal associations to chronic gastritis risk (Supplementary Table S1). Notably, **Cathepsin H** levels demonstrated a modest correlation with disease risk (Supplementary Table S4, S7).

Specifically, the IVW analysis indicated a subtle association (P=0.023, OR=1.053, 95%CI=1.005~1.104), which remained consistent across weighted models (P=0.041, OR=1.095, 95%CI=1.004~1.195). To account for potential variability across instruments, we utilized the random-effects IVW method[27], ensuring that our estimates remained robust against underlying heterogeneity (Supplementary Table S1-9).

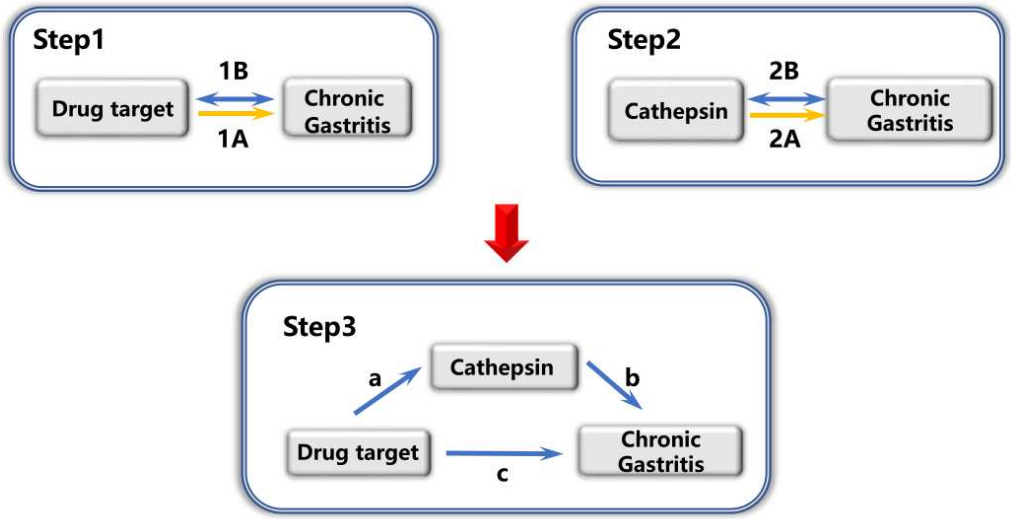


Figure 1. Overview of the study design.

As shown in Figure 1,step 1A represents the causal effects of drug target on CG (Supplementary Figure S1-3), step 1B represents the bi-directional causal effects between drug target and CG.Step 2A represents the causal effects of cathepsin on CG (Supplementary Figure S4-6), step 2B represents the bi-directional causaleffects between cathepsin and CG, step3 represents the mediating analysis of cathepsin in the pathway from the drug target to CG: path c was the total effect of drug target on CG; path b was the causal effect of cathepsin on CG; path a was the causal effect of drug target on cathepsin (Supplementary Figure S7-9).

Colocalization analysis

To differentiate between true causal relationships and genetic signals arising from simple linkage disequilibrium (LD), we performed Bayesian colocalization analysis on the 1,880 significant genes identified in the MR screening. Using the coloc R package with default prior probabilities (p_1 = 1E-4, p2 = 1E-4, p12 = 1E-5), we integrated discovery-phase data to calculate the posterior probability of a shared causal variant.

28 proteins demonstrated strong colocalization evidence with chronic gastritis, yielding a posterior probability of Hypothesis 4 (PP.H4) greater than 0.75. These 28 proteins represent high-confidence candidates for therapeutic targeting, as the genetic evidence suggests that the same underlying variant influences both protein expression and disease risk.

Table 1. Colocalization result of the eQTLs for 28 genes with CG-associated SNPs.

Gene	PP.H0	PP.H1	PP.H2	PP.H3	PP.H4
MRS2	0.000	0.179	0.000	0.000	0.820
TCF19	0.000	0.064	0.000	0.020	0.916
PTK7	0.000	0.219	0.000	0.000	0.781
CUL7	0.000	0.008	0.000	0.000	0.992
RBM26	0.000	0.108	0.000	0.000	0.892
ATF6B	0.000	0.154	0.000	0.051	0.795
TFDP2	0.000	0.022	0.000	0.000	0.978
C4B	0.000	0.120	0.000	0.001	0.879
ZNF664	0.000	0.208	0.000	0.000	0.792
C1orf174	0.000	0.056	0.000	0.000	0.944

SDC4	0.000	0.209	0.000	0.000	0.791
ZNF236	0.000	0.143	0.000	0.000	0.857
LINC00243	0.000	0.245	0.000	0.001	0.754
SP140	0.000	0.018	0.000	0.000	0.982
HLA-DQB2	0.000	0.022	0.000	0.002	0.976
SLC35F2	0.000	0.163	0.000	0.000	0.837
KCTD2	0.000	0.193	0.000	0.000	0.806
DNAJC18	0.000	0.053	0.000	0.000	0.947
TICAM1	0.000	0.185	0.000	0.000	0.815
TMEM173	0.000	0.013	0.000	0.000	0.987
HCP5	0.000	0.081	0.000	0.000	0.919
MMAB	0.000	0.210	0.000	0.001	0.789
RP11-1220K2.2	0.000	0.175	0.000	0.000	0.825
DDX39B	0.000	0.143	0.000	0.000	0.857
NFKBIL1	0.000	0.031	0.000	0.000	0.969
SKIV2L	0.000	0.026	0.000	0.020	0.954
AGPAT1	0.000	0.041	0.000	0.000	0.959
VWA7	0.000	0.009	0.000	0.007	0.984

Enrichment analysis

To elucidate the biological roles of the identified candidate genes and their corresponding proteins, we conducted GO and KEGG enrichment analyses. The results revealed that these targets are primarily integrated into three functional domains: inflammatory immune modulation, mucosal structural integrity, and lipid metabolism (Figure 2).

At the cellular level, the identified genes were significantly enriched in the MHC class II protein complex, the cullin-RING ubiquitin ligase complex, and integral components of the

endoplasmic reticulum membrane. Biological process mapping highlighted a strong involvement in lipopolysaccharide-mediated signaling and responses to bacterial-derived molecules, underscoring the interaction between the host immune system and the gastric microbiome. Interestingly, pathways related to epithelial tube formation and neural tube closure were also identified, suggesting roles in mucosal structural regulation and tissue development.

GO pathway analysis further confirmed a heavy concentration of targets within infectious disease pathways, such as pertussis, hepatitis B, and *Staphylococcus aureus* infection, reflecting their broader roles in pathogen-driven inflammation (Figure 3). Beyond immunity, metabolic signatures such as glycerophospholipid and glyceride metabolism were significantly enriched, pointing to a potential metabolic component in the pathogenesis of chronic gastritis.

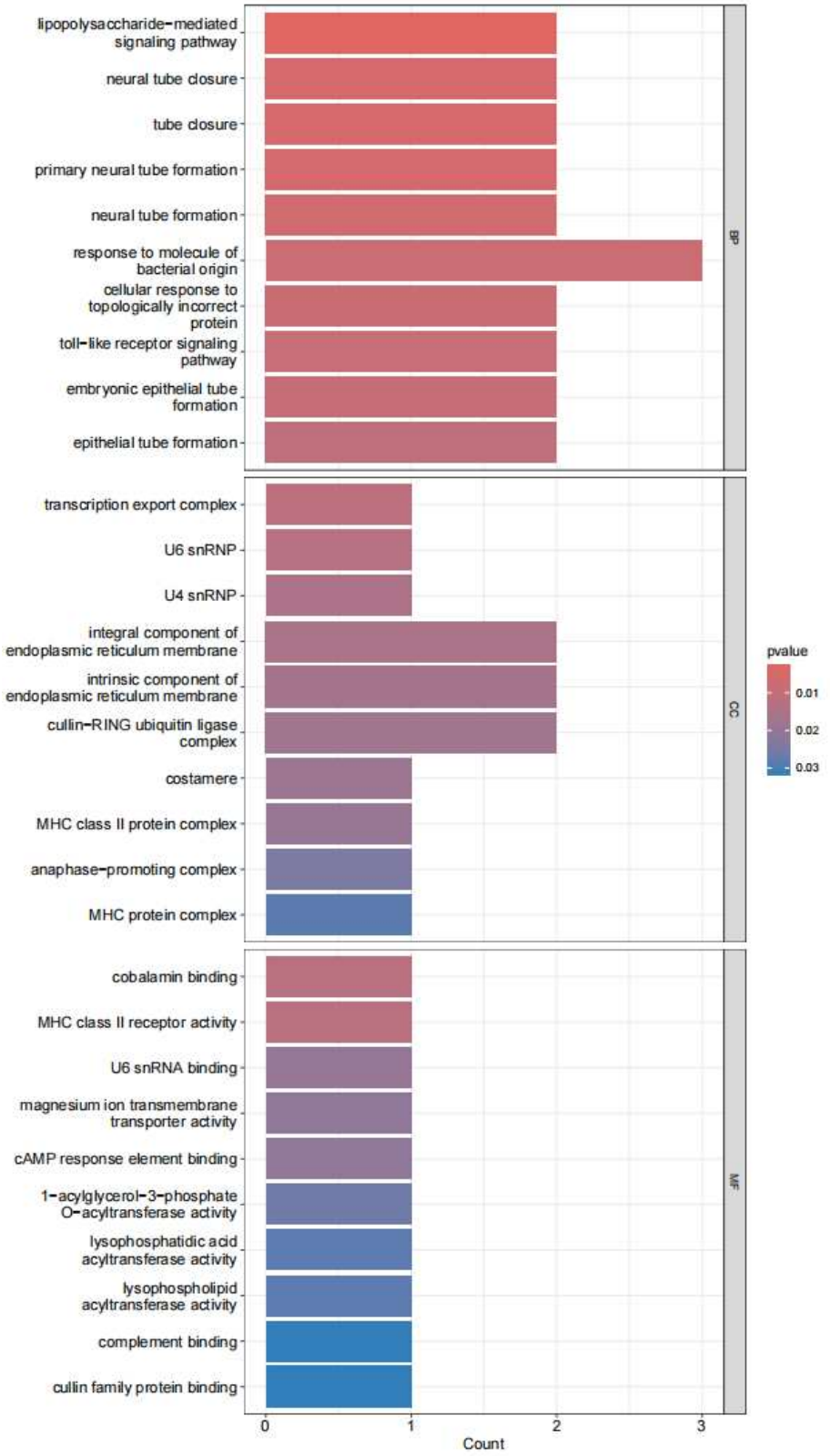


Figure 2. KEGG enrichment results.

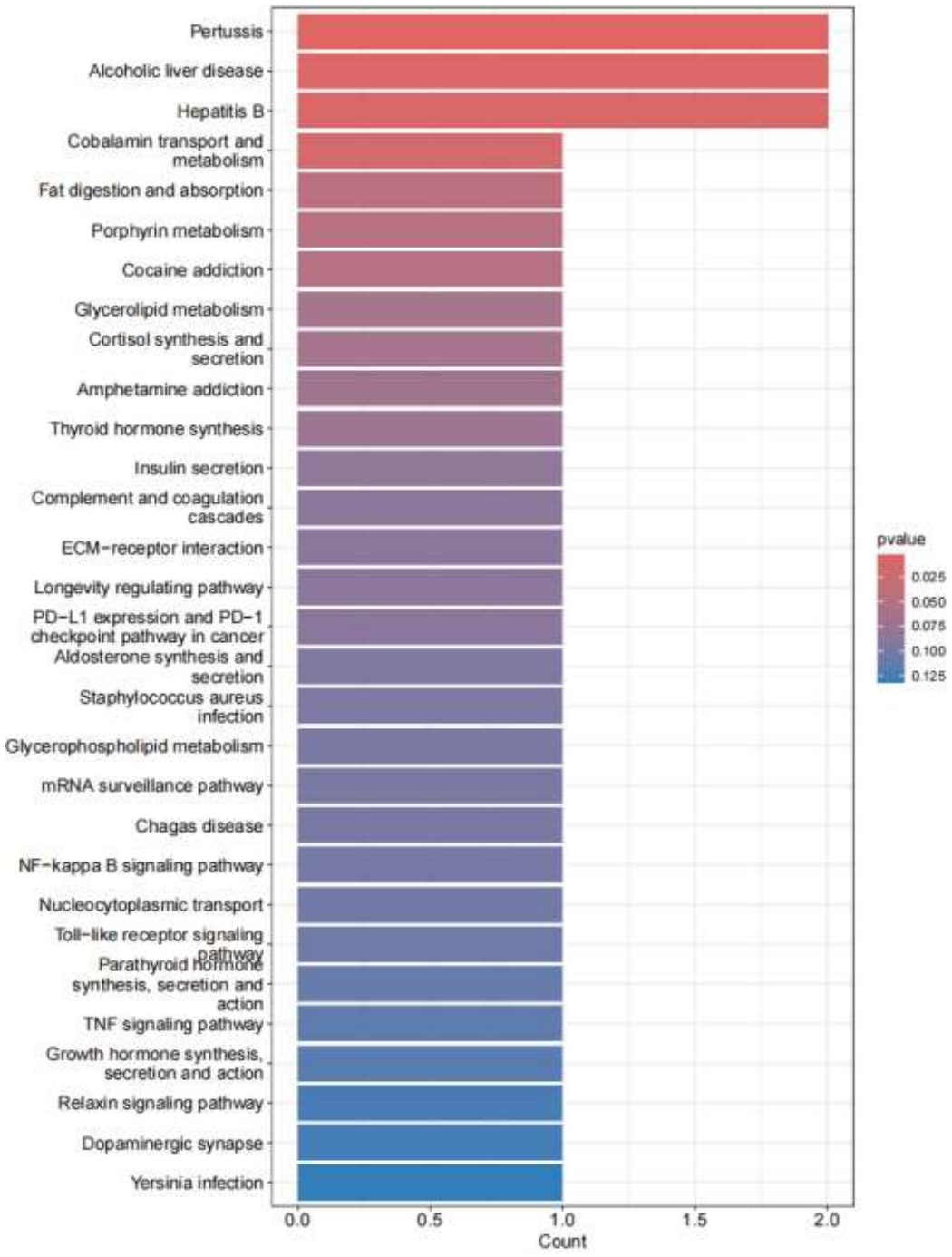


Figure 3. GO enrichment results for three terms.

Protein-protein interaction analysis

To characterize the functional crosstalk between the 28 identified drug target candidates and their broader molecular partners, we mapped their interactions using the STRING database. The resulting network was visualized and analyzed within the Cytoscape environment to identify core regulatory clusters.

Our analysis revealed a highly interconnected PPI network centered around a core cluster of five high-confidence proteins. This functional module consists of 45 nodes and 160 edges (Figure 4). The high degree of connectivity within this network suggests a coordinated biological response, further supporting the hypothesis that these targets operate within shared pathological pathways—particularly those related to immune signaling and mucosal homeostasis—rather than as isolated genetic signals.

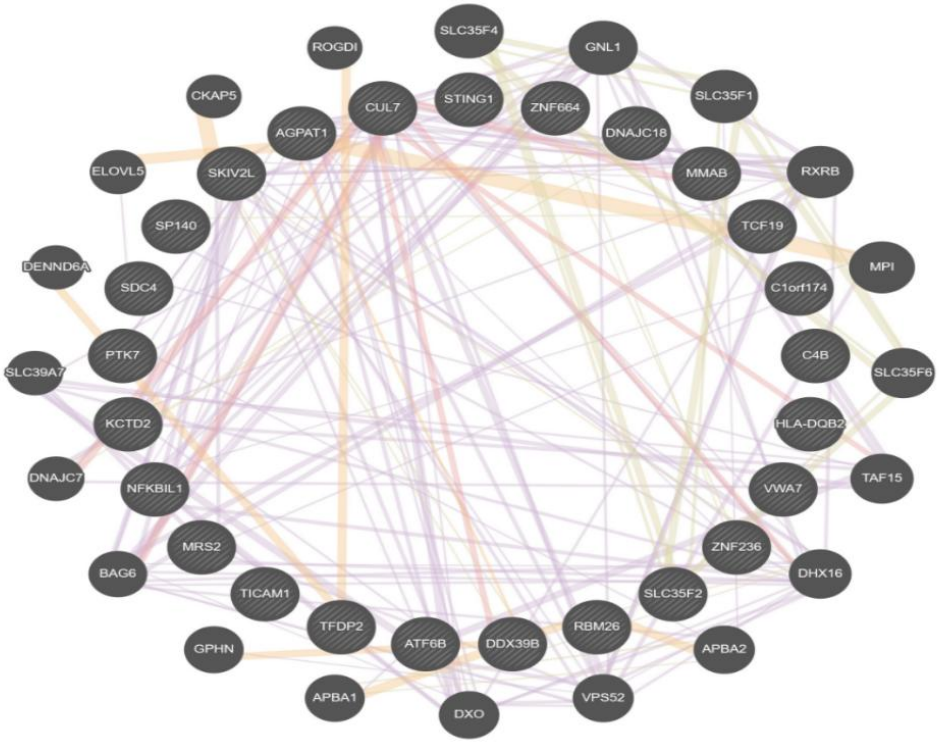


Figure 4. PPI network built with STRING.

Validation of the Causal Relationship between Drug Target Genes, Cathepsin H, and Chronic Gastritis

To substantiate the mediating role of Cathepsin H identified in our initial screening, we conducted multivariate Mendelian randomization (MVMR). By quantifying the indirect effects and calculating mediation proportions, we confirmed that Cathepsin H acts as a significant intermediary between several high-priority genes and the development of chronic gastritis.

As summarized in Table 2, the proportion of the total effect mediated by Cathepsin H varied across different genetic drivers. The most pronounced mediation effect was observed for SDC4, where Cathepsin H accounted for 49.71% of the causal impact on chronic gastritis risk. Substantial mediation ratios were also identified for NFKBIL1 (31.41%) and AGPAT1 (28.80%). Additional moderate mediation effects were observed for several other immune and structural regulators, including ATF6B (17.35%), TICAM1 (16.62%), VWA7 (14.50%), SP140 (12.77%), RP11-1220K2.2 (12.68%), KCTD2 (11.61%), HLA-DQB2 (11.28%), and TMEM173 (11.21%).

These results suggest that these 11 genes do not merely act in isolation (Figure 5); rather, a significant portion of their pathological influence is channeled through the regulation or activity of Cathepsin H, highlighting its potential as a central node in the molecular pathogenesis of chronic gastritis.

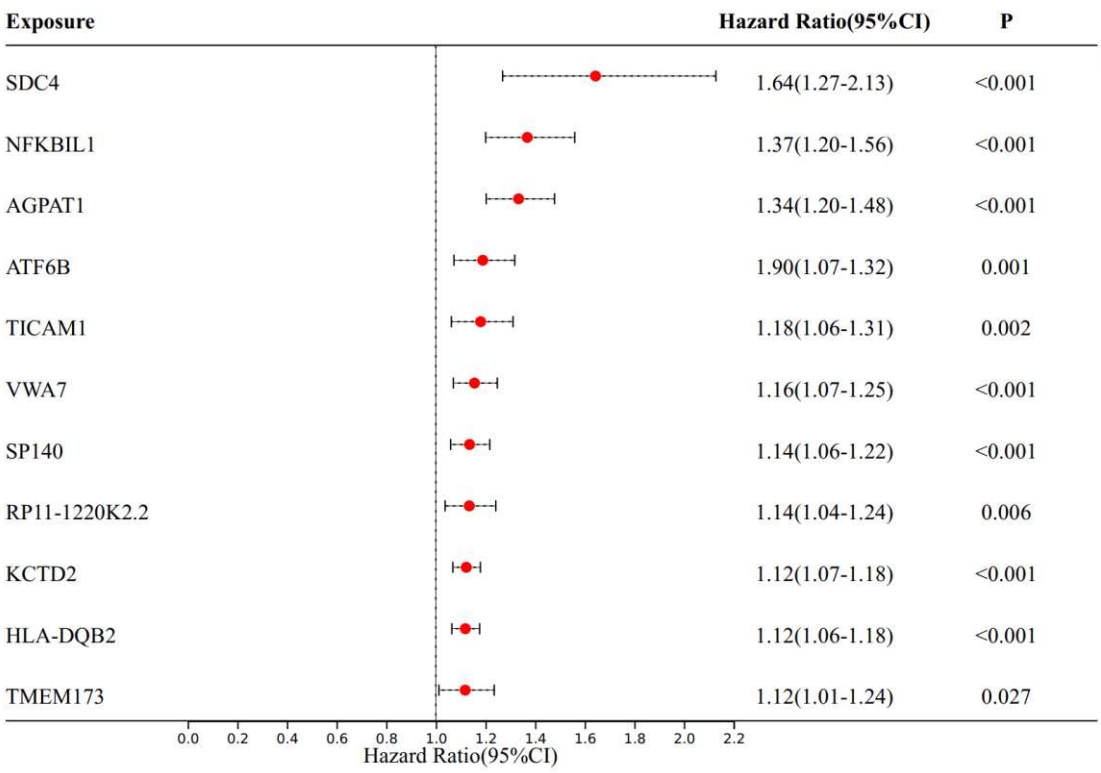


Figure 5. Forest plots displaying the findings from the replication phase for 11 significant genes.

Table 2 Mediation results of 11 significant genes.

Gene	SNP	intermediary	Mediation effect ratio
SDC4	2	cathepsin H	0.497
NFKBIL1	3	cathepsin H	0.314
AGPAT1	4	cathepsin H	0.288
ATF6B	13	cathepsin H	0.174
TICAM1	9	cathepsin H	0.166
VWA7	13	cathepsin H	0.145
SP140	30	cathepsin H	0.128
RP11-1220K2.2	6	cathepsin H	0.127
KCTD2	26	cathepsin H	0.116
HLA-DQB2	15	cathepsin H	0.113

TMEM173	12	cathepsin H	0.112

Drug Prediction

To translate our genetic findings into actionable clinical insights, we utilized the Drug Signatures Database (DSigDB) to identify potential pharmacological interventions. By mapping the high-priority genes identified in our mediation analysis, specifically SDC4, NFKBIL1, and AGPAT1, to known compound-gene interactions, we identified several existing drugs that may be repositioned for the treatment of chronic gastritis.

A selection of these high-probability candidate drugs is presented in Table 3. These candidates provide a genetically-informed foundation for future experimental validation and drug development efforts aimed at mitigating chronic mucosal inflammation.

Table 3 Candidate drug predicted using DSigDB

Gene	Source	Chemical Name
SDC4	D3(DOWN)	furazolidone
SDC4	D3(DOWN)	cimetidine
SDC4	D3(DOWN)	doxycycline
SDC4	D3(UP)	ciclopirox
SDC4	D3(UP)	podophyllotoxin
SDC4	D3(UP)	vorinostat
NFKBIL1	D3(DOWN)	0175029-0000
NFKBIL1	D4 BOSS	mycophenolic acid
NFKBIL1	D4 BOSS	Rosuvastatin
NFKBIL1	D4 BOSS	PRAVASTATIN SODIUM
NFKBIL1	D4 BOSS	bilirubin
NFKBIL1	D4 BOSS	Clopidogrel bisulfate

AGPAT1	D3(UP)	deptropine
AGPAT1	D3(UP)	scriptaid
AGPAT1	D3(UP)	trichostatin A
AGPAT1	D4 BOSS	Estrone sulfate
AGPAT1	D4 TTD	ProTec

Discussion

By integrating multiple Mendelian randomization (MR) methodologies, including Wald ratio, IVW, and MR-Egger, this study identified 11 drug targets for chronic gastritis: SDC4, NFKBIL1, AGPAT1, ATF6B, TICAM1, VWA7, SP140, RP11-1220K2.2, KCTD2, HLA-DQB2, and TMEM173. These targets are particularly robust as the MR framework effectively mitigates confounding factors inherent in observational studies, while Bayesian colocalization provides strong evidence of a shared genetic basis between these proteins and the disease. Furthermore, we utilized genome-wide phenotype association analysis to scrutinize potential pleiotropic effects, alongside enrichment and protein interaction analyses to clarify the biological significance of these candidates.

A primary finding of this study is the pivotal role of Syndecan-4 (SDC4). Heparan sulfate (HS) glycosaminoglycans are known to be recognized by the outer membrane proteins of *Helicobacter pylori* (Hp), mediating bacterial adhesion to host cells [28]. As a transmembrane heparan sulfate proteoglycan, SDC4 is the predominant source of cell-surface HS in gastric epithelial cells [29]. Our analysis supports the observation that $H. pylori$ infection—or the activation of Toll-like receptors (TLRs), upregulates SDC4 expression in an NF-κB-dependent manner[30].Given that high-pathogenicity Hp strains induce SDC4 expression via the cag pathogenicity island (cagPAI), this protein serves as a critical receptor for initial bacterial colonization [31].[32]. Interestingly, the shedding

of the SDC4 ectodomain, potentially triggered by Hp-induced HB-EGF secretion, may represent a protective host mechanism to limit bacterial binding [33] [34]. Our results also emphasize the importance of NFKBIL1 and AGPAT1. NFKBIL1 acts as a significant negative regulator of the NF-κB pathway. Since Hp exploits TLR-mediated NF-κB activation to enhance SDC4 transcription, NFKBIL1 likely participates in the pathological process by dampening inflammatory responses and maintaining the mucosal barrier [35] [36] [37]

Similarly, AGPAT1 represents a metabolic driver of mucosal health. While its specific role in chronic gastritis has not been fully elucidated, as an essential enzyme in phospholipid synthesis, AGPAT1 is critical for maintaining epidermal and mucosal permeability barriers [38] [39]. Similar to AGPAT3's role in modulating oxidative stress and ferroptosis in gastric cancer, AGPAT1 may influence the inflammatory landscape of the gastric mucosa through lipid-mediated signaling pathways [40].

This study offers several distinct advantages. It is the first to employ molecular genetics (MR) to identify therapeutic targets for chronic gastritis, leveraging the largest publicly available GWAS datasets to date. By covering the complete evaluative spectrum, from target identification and functional enrichment to protein-protein interaction (PPI) and drug-signature prediction, we have provided a genetically-supported rationale for 11 targets with significant translational potential. These findings offer a strategic foundation for reducing research costs and increasing the success rate of future clinical trials in this high-burden disease.

Limitations

Despite the rigorous methodology, this study has limitations. First, the reliance on blood-derived eQTLs may not fully capture the tissue-specific genetic regulatory mechanisms active within the gastric mucosa. Second, the study population was primarily of European ancestry, which may limit the

generalizability of these findings to other ethnic groups. Future research should integrate multi-omics data from more diverse global populations and explore alternative analytical models to deepen the understanding of chronic gastritis and its potential therapeutic landscape.

Conclusion

In summary, this study utilized a comprehensive Mendelian randomization framework to identify and validate potential therapeutic targets for chronic gastritis. By integrating large-scale genomic data across two independent cohorts, we pinpointed 11 high-priority drug targets, further substantiated by robust Bayesian colocalization evidence. Functional analysis reveals that these genes are fundamentally linked to innate immune modulation and mucosal barrier integrity, offering a genetically-supported rationale for their role in the disease's pathogenesis.

Our findings provide a strategic roadmap for the development of more efficient, targeted treatments, with the potential to significantly reduce the costs and high failure rates typically associated with drug discovery. Furthermore, these results promote the advancement of personalized medicine for a high-burden patient population that currently lacks optimal symptom control. Given the strong causal evidence presented here, the transition of these targets into drug development and clinical trial pipelines is a high-priority imperative for improving long-term outcomes in chronic gastritis.

Acknowledgements

We thank the eQTLGen consortium, the FinnGen team and other researchers

and participants for providing publicly available data for this analysis.

Author contributions

T.Q.L. and Z.Q.conceived and designed the experiments. J.W.Y. and Y.J.F. analyzed the data. D.Y. and Y.R.H. contributed materials/analysis tools. L.W. and P.F.L.wrote the paper. All authors read and approved the final manuscript.

Availability of data and materials

The datasets analysed during the current study are available in the in the following repositories: eQTLs data were obtained from eQTLGen Consortium (https://eqtlgen.org/); Cathepsin data were obtained from the open GWAS study published by Folkersen et al[13]. (https://gwas.mrcieu.ac.uk/). Cathepsins were used as exposures. Chronic gastritis data were retrieved from an open GWAS (https://gwas.mrcieu.ac.uk/).

Declarations

Ethics approval and consent to participate

This study exclusively utilized publicly available data from open access databases. As no human participants, human tissues, or individual clinical information were directly involved in the present research, ethical approval and informed consent were not required for this investigation.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding Declaration

This research received no external funding.

Clinical trial number

Not applicable.

References

[1]Zhang X, Huang Z, Hu X, Tang Y, Zhong X, Li L, Ni L, Li F, Hua H, Zhu F. Wendan Decoction ameliorates MNNG-induced chronic atrophic gastritis via the JAK2/STAT3 signaling pathway. J Ethnopharmacol. 2025 Oct 11;355(PtB):120724. doi: 10.1016/j.jep.2025.120724 . Epub ahead of print. PMID: 41077317.

[2]Elbehiry A, Marzouk E, Abalkhail A. Unraveling *Helicobacter pylori* : Insights into Pathogenesis, Immune Evasion, and Progress Toward Effective Vaccination. Vaccines (Basel). 2025 Jul 3;13(7):725. doi: 10.3390/vaccines13070725 . . . . PMID: 40733702; PMCID: PMC12299425.

[3]Morgan DR, Corral JE, Li D, Montgomery EA, Riquelme A, Kim JJ, Sauer B, Shah SC. ACG Clinical Guideline: Diagnosis and Management of Gastric Premalignant Conditions. Am J Gastroenterol. 2025 Apr 1;120(4):709-737. doi: 10.14309/ajg.0000000000003350. Epub 2025 Mar 12. PMID: 40072510.

[4]Yin Y, Liang H, Wei N, Zheng Z. Prevalence of chronic atrophic gastritis worldwide from 2010 to 2020: an updated systematic review and meta-analysis. Ann Palliat Med. 2022 Dec;11(12):3697-3703. doi: 10.21037/apm-21-1464. PMID: (1).

[5]Sampieri CL. Helicobacter pylori and gastritis: the role of extracellular matrix metalloproteases, their inhibitors, and the disintegrins and metalloproteases--a systematic literature review. Dig Dis Sci. 2013 Oct;58(10):2777-83. doi: 10.1007/s10620-013-2767-x. Epub 2013 Jul 2. PMID: 23817928.

[6]Nigam AK, Pignatelli M, Boulos PB. Current concepts in metastasis. Gut. 1994 Jul;35(7):996-1000. doi: 10.1136/gut.35.7.996. PMID: 8063231; PMCID: PMC1374851.

[7]Mason RW, Johnson DA, Barrett AJ, Chapman HA. Elastinolytic activity of human cathepsin L. Biochem J. 1986 Feb 1;233(3):925-7. doi: 10.1042/bj2330925. PMID: 3518704; PMCID: PMC1153120.

[8]Ohta T, Terada T, Nagakawa T, Tajima H, Itoh H, Fonseca L, Miyazaki I. Pancreatic trypsinogen and cathepsin B in human pancreatic carcinomas and associated metastatic lesions. Br J Cancer. 1994 Jan;69(1):152-6. doi: 10.1038/bjc.1994.25. PMID: 8286198; PMCID: PMC1968761.

[9]Shuja S, Sheahan K, Murnane MJ. Cysteine endopeptidase activity levels in normal human tissues, colorectal adenomas and carcinomas. Int J Cancer. 1991 Sep 30;49(3):341-6. doi: 10.1002/ijc.2910490305. PMID: 1917131.

[10]Katan MB. Apolipoprotein E isoforms, serum cholesterol, and cancer. The Lancet. 1986 Mar 1;1(8479):507-8. doi: 10.1016/s0140-6736(86)92972-7. PMID: 2869248.

[11]Schmidt AF, Finan C, Gordillo-Maraño M, Asselbergs FW, Freitag DF, Patel RS, Tyl B, Chopade S, Faraway R, Zwierzyna M, Hingorani AD. Genetic drug target validation using Mendelian randomisation. Nat Commun. 2020 Jun 26;11(1):3255. doi: 10.1038/s41467-020-16969-0. PMID: 32591531; PMCID: PMC7320010.

[12]Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol. 2000 Aug;29(4):722-9. doi: 10.1093/self/29.4.722. Erratum in: Int J Epidemiol. 2018 Feb 1;47(1):358. doi: 10.1093/self/dyx275. PMID: 10922351.

[13]Folkersen L, Gustafsson S, Wang Q, Hansen DH, Hedman ÅK, Schork A, Page K, Zhernakova DV, Wu Y, Peters J, Eriksson N, Bergen SE, Boutin TS, Bretherick AD, Enroth S, Kalnapenkis A, Gådin JR, Suur BE, Chen Y, Matic L, Gale JD, Lee J, Zhang W, Quazi A, Ala-Korpela M, Choi SH, Claringbould A, Danesh J, Davey Smith G, de Masi F, Elmståhl S, Engström G, Fauman E, Fernandez C, Franke L, Franks PW, Giedraitis V, Haley C, Hamsten A, Ingason A, Johansson Å, Joshi PK, Lind L, Lindgren CM, Lubitz S, Palmer T, Macdonald-Dunlop E, Magnusson M, Melander O, Michaelsson K, Morris AP, Mägi R, Nagle

MW, Nilsson PM, Nilsson J, Orho-Melander M, Polasek O, Prins B, Pålsson E, Qi T, Sjögren M, Sundström J, Surendran P, Võsa U, Werge T, Wernersson R, Westra HJ, Yang J, Zhernakova A, Ärnlöv J, Fu J, Smith JG, Esko T, Hayward C, Gyllensten U, Landen M, Siegbahn A, Wilson JF, Wallentin L, Butterworth AS, Holmes MV, Ingelsson E, Mälarstig A. Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. Nat Metab. 2020 Oct;2(10):1135-1148. doi: 10.1038/s42255-020-00287-2. Epub 2020 Oct 16. PMID: 33067605; PMCID: PMC7611474.

[14]Folkersen L, Fauman E, Sabater-Lleal M, Strawbridge RJ, Frånberg M, Sennblad B, Baldassarre D, Veglia F, Humphries SE, Rauramaa R, de Faire U, Smit AJ, Giral P, Kurl S, Mannarino E, Enroth S, Johansson Å, Enroth SB, Gustafsson S, Lind L, Lindgren C, Morris AP, Giedraitis V, Silveira A, Franco-Cereceda A, Tremoli E; IMPROVE study group; Gyllensten U, Ingelsson E, Brunak S, Eriksson P, Ziemek D, Hamsten A, Mälarstig A. Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. PLoS Genet. 2017 Apr 3;13(4):e1006706. doi: 10.1371/journal.pgen.1006706. PMID: 28369058; PMCID: PMC5393901.

[15]Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet. 2014 Sep 15;23(R1):R89-98. doi: 10.1093/hmg/ddu328. Epub 2014 Jul 4. PMID: 25064373; PMCID: PMC4170722.

[16]Deng YT, Ou YN, Wu BS, Yang YX, Jiang Y, Huang YY, Liu Y, Tan L, Dong Q, Suckling J, Li F, Yu JT. Identifying causal genes for depression via integration of the proteome and transcriptome from brain and blood. Mol Psychiatry. 2022 Jun;27(6):2849-2857. doi: 10.1038/s41380-022-01507-9. Epub 2022 Mar 16. PMID: 35296807.

[17]Kamat MA, Blackshaw JA, Young R, Surendran P, Burgess S, Danesh J, Butterworth AS, Staley JR. PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. Bioinformatics. 2019 Nov 1;35(22):4851-4853. doi: 10.1093/bioinformatics/btz469. PMID: 31233103; PMCID: PMC6853652.

[18]Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284–7.

[19]Luo W,Brouwer C.Pathview:an R/Bioconductor package for pathway-based data integration and visualization.Bioinformatics.2013;29(14):1830-1.

[20]Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R,Gable AL, Fang T, Doncheva NT, Pyysalo S, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res.2023;51(D1):D638-d646.

[21]Otasek D,Morris JH, Bouças J, Pico AR, Demchak B.Cytoscape automation: empowering workflow-based network analysis. Genome Biol.2019;20(1):185.

[22]Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P,Franz M, Grouios C, Kazi F, Lopes CT, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function.Nucleic Acids Res.2010;38(2):W214-220.

[23]Yoo M, Shin J, Kim J, Ryall KA, Lee K, Lee S, Jeon M, Kang J, Tan AC.DSigDB: drug signatures database for gene set analysis. Bioinformatics.2015;31(18):3069–71.

[24]Zhao JH, Stacey D, Eriksson N, Macdonald-Dunlop E, Hedman ÅK, Kalnapenkis A, Enroth S, Cozzetto D, Digby-Bell J, Marten J, Folkersen L, Herder C, Jonsson L, Bergen SE, Gieger C, Needham EJ, Surendran P; Estonian Biobank Research Team; Paul DS, Polasek O, Thorand B, Grallert H, Roden M, Võsa U, Esko T, Hayward C, Johansson Å, Gyllensten U, Powell N, Hansson O, Mattsson-Carlgren N, Joshi PK, Danesh J, Padyukov L, Klareskog L, Landén M, Wilson JF, Siegbahn A,

Wallentin L, Mälarstig A, Butterworth AS, Peters JE. Genetics of circulating inflammatory proteins identifies drivers of immune-mediated disease risk and therapeutic targets. Nat Immunol. 2023 Sep;24(9):1540-1551. doi: 10.1038/s41590-023-01588-w. Epub 2023 Aug 10. Erratum in: Nat Immunol. 2023 Nov;24(11):1960. doi: 10.1038/s41590-023-01635-6. PMID: 37563310; PMCID: PMC10457199.

[25]Lopera-Maya EA, Kurilshikov A, van der Graaf A, Hu S, Andreu-Sánchez S, Chen L, Vila AV, Gacesa R, Sinha T, Collij V, Klaassen MAY, Bolte LA, Gois MFB, Neerincx PBT, Swertz MA; LifeLines Cohort Study; Harmsen HJM, Wijmenga C, Fu J, Weersma RK, Zhernakova A, Sanna S. Effect of host genetics on the gut microbiome in 7,738 participants of the Dutch Microbiome Project. Nat Genet. 2022 Feb;54(2):143-151. doi: 10.1038/s41588-021-00992-y. Epub 2022 Feb 3. Erratum in: Nat Genet. 2022 Sep;54(9):1448. doi: 10.1038/s41588-022-01164-2. PMID: 35115690.

[26]Orrù V, Steri M, Sidore C, Marongiu M, Serra V, Olla S, Sole G, Lai S, Dei M, Mulas A, Virdis F, Piras MG, Lobina M, Marongiu M, Pitzalis M, Deidda F, Loizedda A, Onano S, Zoledziewska M, Sawcer S, Devoto M, Goro M, Floris M, Floris MGR Schlessinger D, Fiorillo E, Cucca F. Complex genetic signatures in immune cells underlie autoimmunity and inform therapy. Nat Genet. 2020 Oct;52(10):1036-1045. doi: 10.1038/s41588-020-0684-4. Epub 2020 Sep 14. Erratum in: Nat Genet. 2020 Nov;52(11):1266. doi: 10.1038/s41588-020-00718-6. PMID: 32929287; PMCID: PMC8517961.

[27]Bovijn J, Krebs K, Chen CY, Boxall R, Censin JC, Ferreira T, Pulit SL, Glastonbury CA, Laber S, Millwood IY, Lin K, Li L, Chen Z, Milani L, Smith GD, Walters RG, Mägi R, Neale BM, Lindgren CM, Holmes MV. Evaluating the cardiovascular safety of sclerostin inhibition using evidence from meta-analysis of clinical trials and human genetics. Sci Transl Med. 2020 Jun 24;12(549):eaay6570. doi: 10.1126/scitranslmed.aay6570. PMID: 32581134; PMCID: PMC7116615.

[28]Guzman-Murillo MA, Ruiz-Bustos E, Ho B & Ascencio F (2001) Involvement of the heparan sulphate-binding proteins of Helicobacter pylori in its adherence to HeLa S3 and Kato III cell lines. J Med Microbiol 50: 320–329.

[29]Zimmermann P, David G. The syndecans, tuners of transmembrane signaling. FASEB J. 1999;13 Suppl:S91-S100. doi: 10.1096/fasebj.13.9001.s91. PMID: 10352150.

[30]Smith MF Jr, Novotny J, Carl VS, Comeau LD. Helicobacter pylori and toll-like receptor agonists induce syndecan-4 expression in an NF-kappaB-dependent manner. Glycobiology. 2006 Mar;16(3):221-9. doi: 10.1093/glycob/cwj061. Epub 2005 Nov 29. PMID: 16319082; PMCID: PMC1370916.

[31]Magalhães A, Marcos NT, Carvalho AS, David L, Figueiredo C, Bastos J, David G, Reis CA. Helicobacter pylori cag pathogenicity island-positive strains induce syndecan-4 expression in gastric epithelial cells. FEMS Immunol Med Microbiol. 2009 Aug;56(3):223-32. doi: 10.1111/j.1574-695X.2009.00569.x. Epub 2009 May 16. PMID: 19519784.

[32]Utt M, Danielsson B, Wadström T. Helicobacter pylori vacuolating cytotoxin binding to a putative cell surface receptor, heparan sulfate, studied by surface plasmon resonance. FEMS Immunol Med Microbiol. 2001 Mar;30(2):109-13. doi: 10.1111/j.1574-695X.2001.tb01557.x. PMID: 11267842.

[33]Götte M. Syndecans in inflammation. FASEB J. 2003 Apr;17(6):575-91. doi: 10.1096/fj.02-0739rev. PMID: 12665470

[34]Fitzgerald ML, Wang Z, Park PW, Murphy G, Bernfield M. Shedding of syndecan-1 and -4 ectodomains is regulated by multiple signaling pathways and mediated by a TIMP-3-sensitive metalloproteinase. J Cell Biol. 2000 Feb 21;148(4):811-24. doi: 10.1083/jcb.148.4.811. PMID: 10684261; PMCID: PMC2169376.

[35]Zhang Y, Pasparakis M, Kollias G, Simons M. Myocyte-dependent regulation of endothelial cell syndecan-4 expression. Role of TNF-alpha. J Biol Chem. 1999 May 21;274(21):14786-90. doi: 10.1074/jbc.274.21.14786. PMID: 10329676.

[36]Zhou A, Scoggin S, Gaynor RB, Williams NS. Identification of NF-kappa B-regulated genes induced by TNFalpha

utilizing expression profiling and RNA interference. Oncogene. 2003 Apr 3;22(13):2054-64. doi: 10.1038/sj.onc.1206262. Erratum in: Oncogene. 2004 Dec 16;23(58):9447. PMID: 12673210.

[37]Guzman-Murillo MA, Ruiz-Bustos E, Ho B & Ascencio F (2001) Involvement of the heparan sulphate-binding proteins of Helicobacter pylori in its adherence to HeLa S3 and Kato III cell lines. J Med Microbiol 50: 320–329.

[38]Agarwal AK, Sukumaran S, Cortes VA, Tunison K, Mizrachi D, Sankella S, Gerard RD, Horton JD, Garg A. Human 1-acylglycerol-3-phosphate O-acyltransferase isoforms 1 and 2: biochemical characterization and inability to rescue hepatic steatosis in Agpat2(-/-) gene lipodystrophic mice. J Biol Chem. 2011 Oct 28;286(43):37676-91. doi: 10.1074/jbc.M111.250449. Epub 2011 Aug 27. PMID: 21873652; PMCID: PMC3199511.

[39]Lu B, Jiang YJ, Man MQ, Brown B, Elias PM, Feingold KR. Expression and regulation of 1-acyl-sn-glycerol- 3-phosphate acyltransferases in the epidermis. J Lipid Res. 2005 Nov;46(11):2448-57. doi: 10.1194/jlr.M500258-JLR200. Epub 2005 Sep 8. PMID: 16150824.

[40]Peng Y, Lei X, Yang Q, Zhang G, He S, Wang M, Ling R, Zheng B, He J, Chen X, Li F, Zhou Q, Zhao L, Ye G, Li G. Helicobacter pylori CagA-mediated ether lipid biosynthesis promotes ferroptosis susceptibility in gastric cancer. Exp Mol Med. 2024 Feb;56(2):441-452. doi: 10.1038/s12276-024-01167-5. Epub 2024 Feb 21. PMID: 38383581; PMCID: PMC10907675.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- tables.zip