**Name of Journal:** *World Journal of Diabetes*

**Manuscript NO:** 88359

**Manuscript Type:** ORIGINAL ARTICLE

***Clinical and Translational Research***

**Identification of hub genes associated with *Helicobacter pylori* infection and type 2 diabetes mellitus: A pilot bioinformatics study**

ChenH *et al. H. pylori* infection and T2DM

Han Chen, Guo-Xin Zhang, Xiao-Ying Zhou

**Han Chen,** Department of Gastroenterology, The First Affiliated Hospital of Nanjing Medical University, Nanjing 210003, Jiangsu Province, China

**Guo-Xin Zhang, Xiao-Ying Zhou,** Department of Gastroenterology, Jiangsu Province Hospital, Nanjing 210029, Jiangsu Province, China

**Co-corresponding authors:** Guo-Xin Zhang and Xiao-Ying Zhou.

**Author contributions:** Zhou XY and Zhang GX concepted and designed the research study; Chen H and Zhou X developed methodology; Chen H acquired the data; Zhou XY analyzed and interpretated the data; Chen H wrote the first version of the manuscript; Zhang GX and Zhou XY revised the manuscript; all authors were involved in the critical review of the results and have contributed to, read, and approved the final manuscript. Zhou XY and Zhang GX contributed equally to this work as co-corresponding authors. The reasons for designating Zhou XY and Zhang GX as co-corresponding authors are threefold. First, the research was performed as a collaborative effort, and the designation of co-corresponding authorship accurately reflects the distribution of responsibilities and burdens associated with the time and effort required to complete the study and the resultant paper. This also ensures effective communication and management of post-submission matters, ultimately enhancing the paper's quality and reliability. Second, the overall research team encompassed authors with a variety of expertise and skills from different fields, and the designation of co-corresponding authors best reflects this diversity. This also promotes the most comprehensive and in-depth examination of the research topic, ultimately enriching readers' understanding by offering various expert perspectives. Third, Zhou XY and Zhang GX contributed to almost the same funding on this research. The choice of these researchers as co-corresponding authors acknowledges and respects this equal contribution, while recognizing the spirit of teamwork and collaboration of this study. In summary, we believe that designating Zhou XY and Zhang GX as co-corresponding authors of is fitting for our manuscript as it accurately reflects our team's collaborative spirit, equal contributions, and diversity.

**Supported by** National Natural Science Foundation of China, No. 82100594.

**Corresponding author: Xiao-Ying Zhou, MD, PhD, Associate Chief Physician,** Department of Gastroenterology, Jiangsu Province Hospital, No. 300 Guangzhou Road, Nanjing 210029, Jiangsu Province, China. zhouxiaoying0926@njmu.edu.cn

**Received:** September 21, 2023

**Revised:** November 21, 2023

**Accepted:** December 27, 2023

**Published online:**

**Abstract**

BACKGROUND

*Helicobacter pylori* (*H. pylori*) infection is related to various extragastric diseases including type 2 diabetes mellitus (T2DM). However, the possible mechanisms connecting *H. pylori* infection and T2DM remain unknown.

AIM

To explore potential molecular connections between *H. pylori* infection and T2DM.

METHODS

We extracted gene expression arrays from three online datasets (GSE60427, GSE27411 and GSE115601). Differentially expressed genes (DEGs) commonly present in patients with *H. pylori* infection and T2DM were identified. Hub genes were validated using human gastric biopsy samples. Correlations between hub genes and immune cell infiltration, miRNAs, and transcription factors (TFs) were further analyzed.

RESULTS

A total of 67 DEGs were commonly presented in patients with *H. pylori* infection and T2DM. Five significantly upregulated hub genes, including *TLR4*, *ITGAM*, *C5AR1*, *FCER1G*, and *FCGR2A*, were finally identified, all of which are closely related to immune cell infiltration. The gene-miRNA analysis detected 13 miRNAs with at least two gene cross-links. TF-gene interaction networks showed that *TLR4* was coregulated by 26 TFs, the largest number of TFs among the 5 hub genes.

CONCLUSION

We identified five hub genes that may have molecular connections between *H. pylori* infection and T2DM. This study provides new insights into the pathogenesis of *H. pylori*-induced onset of T2DM.

**Key Words:** *Helicobacter pylori*; Type 2 diabetes mellitus; Bioinformatics analysis; Differentially expressed genes; Hub genes

Chen H, Zhang GX, Zhou XY. Identification of hub genes associated with *Helicobacter pylori* infection and type 2 diabetes mellitus: A pilot bioinformatics study. *World J Diabetes* 2023; In press

**Core Tip:** This bioinformatic research is the one of the first studies to identify the key genes and pathways associated with both *Helicobacter pylori* (*H. pylori*) infection and type 2 diabetes mellitus (T2DM), using integrated bioinformatics analyses. Five hub genes were identified, including *TLR4*, *C5AR1*, *ITGAM*, *FCGR2A*, *FCER1G*, and all of which were closely related to immune cell infiltration. We also verified their expression in clinical specimens. Hopefully, this study will shed some light on the pathogenesis of *H. pylori*-induced T2DM in the future. This study is of great clinical importance.

**INTRODUCTION**

The infection rate of *Helicobacter pylori* (*H. pylori*) is still increasing recently and it infects almost 50% of the world’ population. The prevalence rate is even higher in less developed countries[1]. It not only affects gastric disease but also affects extragastric diseases such as non-alcoholic fatty liver disease[2], cardiovascular disease[3], autoimmune disease[4], and endocrine disorders, such as diabetes[5]. In recent years, the prevalence rate of type 2 diabetes mellitus (T2DM) and its complications have also increased significantly[6]. The consequences of poor glycemic control in the long and short term can be significant on social and economic levels[7,8]. Patients with T2DM are more susceptible to *H. pylori* infection, according to our previous meta-analysis[9,10]. There is a significant decrease in the eradication rate of *H. pylori* infection in T2DM patients with *H. pylori* infection compared to T2DM patients without infection[11]. Additionally, *H. pylori*-infected T2DM patients have worse glycemic control capability[12]. All these clinical studies strongly suggest that there is an association between *H. pylori* infection and T2DM.

However, the detailed mechanisms underlying *H. pylori* infection and T2DM remain unclear. According to previous studies, both innate and adaptive immune reactions may be activated in the mucosa of the stomach as a result of *H. pylori* infection[13]. This local inflammation in the stomach may spread systematically as a result of proinflammatory cytokines released by the stomach[14]. Chronic low-grade inflammation, which is a feature of *H. pylori*-associated T2DM, would be more likely to develop as a result[15]. Our previous mechanistic study suggested that *H. pylori* infection induces hepatic insulin resistance by the c-Jun/miR-203/SOCS3 signaling pathway[16]. The gut microbiota may also play a role in the immune and metabolic homeostasis of the host, and the infection of *H. pylori* not only disrupts the balance of commensal bacterial species in the gastric mucosa but also causes alterations in the microbial composition of the human gut[17]. However, these hypotheses have not been formally confirmed and validated.

This study aimed to investigate the potential molecular connections between *H. pylori* infection and T2DM. We identified differentially expressed genes (DEGs) by analyzing gene expression datasets through comprehensive bioinformatics analysis. DEGs were screened by combining the results from GEO datasets. Protein-protein interaction (PPI) construction, Gene Ontology (GO) term analysis, and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed to identify the hub genes linked to the two diseases. A miRNA-hub gene network and transcription factor (TF)‒gene mRNA interaction network were also constructed. We sought to provide new insights into the pathogenesis of *H. pylori*-induced onset of T2DM.

**MATERIALS AND METHODS**

***Data sources***

The NCBI-GEO database is a publicly available database containing gene expression datasets[18,19]. Three datasets were retrieved from the GEO database (https://www.ncbi.nlm.nih.gov/geo/), including two gene expression profiles related to *H. pylori* (GSE60427 and GSE27411) and one dataset related to T2DM (GSE115601). Detailed information on the microarray datasets is provided in Supplementary Table 1. Gene expression profiles were set accordingly, including (1) tissue samples collected from diseased and normal gastric tissues and (2) datasets with more than three samples.

***Identification of DEGs***

The NCBI-GEO2R interactive tool was utilized to analyze and compare data under similar experimental conditions from two or more sample groups to identify genes significantly differentially expressed for both diseases (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?)[20]. Genes that satisfied the criteria of log fold change > 0.4 with adjusted *P* value less than 0.05 were identified as DEGs. Genes presenting upregulation or downregulation in both *H. pylori* and T2DM were selected using the Venn diagram web tool (http://bioinfogp.cnb.csic.es/tools/venny/).

***Functional enrichment analysis of DEGs***

DAVID (Database for Annotation, Visualization, and Integrated Discovery), as an online tool, was used to predict the functions of hub genes based on GO enrichment analysis and KEGG pathway analysis (https://david.ncifcrf.gov/)[21] at three levels: Biological process (BP), molecular function (MF), and cellular component (CC). Bubble maps were used for representing BP, MF, CC, and KEGG pathways, using R package of ggPlot2. A statistically significant *P* value was defined as *P* value less than 0.05.

***Construction of PPI network and identification of hub genes***

A public online database, named STRING (https://string-db.org/), can be used to search for and predict PPIs. This inclusive resource facilitates the investigation of direct physical associations between proteins, as well as the detection of indirect functional connections unveiled through correlation analyses[22]. When common DEGs between different groups were identified, they were uploaded to STRING’s official website (https://cn.string-db.org/) and the interactions between DEGs and STRING database proteins were then assigned (with a minimum needed interaction score of 0.40). We followed the method of Liu *et al*[23], in which PPI interaction networks were visualized using Cytoscape (Version 3.6.1). Cytoscape is from National Institute of General Medical Sciences, United States. We used CytoHubba (Version 0.1) to identify hub genes using a maximal clique centrality algorithm.

***Evaluation of infiltrated immune cells***

To explore the association between infiltrating immune cells and *H. pylori* infection, data on proportions of the 22 immune cell types were obtained using the “cell-type identification by estimating relative subsets of RNA transcripts” (CIBERSORT) algorithm (https://cibersort.stanford.edu/). As a result, only samples with a *P* value of < 0.05 were included in the immune cell infiltration matrix. Boxplots and violin plots were utilized to visualize the proportions of infiltrated immune cells in each sample and each group. The correlation between expression of the five hub genes and the abundance of six immune cell subsets [B cells, CD4+ T cells, CD8+ T cells, macrophages, dendritic cells (DCs), and neutrophils] was analyzed in the gene module of TIMER (http://timer.cistrome.org/)[24].

***MiRNAs prediction and gene–miRNA interaction network construction***

In order to predict their targeted miRNAs, hub genes were selected and analyzed using the miRWalk database (http://mirwalk.umm.uni-heidelberg.de/). The filter setting with a score of > 0.90 was implemented. The target gene binding region was the 3ʹ-UTR, and the intersection with other databases was set to miRDB. Further data processing was carried out by Cytoscape.

***TF-gene interaction network***

The Network Analyst database (https://www.networkanalyst.ca/) was applied to identify human TFs of the related hub genes[25]. The database includes all three data sources named JASPAR, ENCODE and ChIP Enrichment Analysis. ChIP Enrichment Analysis was used to identify target TFs of hub genes in our current study. Moreover, the Cytoscape tool was used to visualize the TF-gene interaction network among TFs and hub genes.

***Single‑gene gene set enrichment analysis***

Gene set enrichment analysis (GSEA) of each hub gene was performed using the “clusterProfiler” R package to identify regulatory pathways and biological functions associated with each hub gene. An adjusted *P* < 0.05 was used to indicate significant thresholds for GSEA.

***Hub genes validated in clinical specimens***

The results of our bioinformatics-based analysis were further verified by RT-qPCR assays. Gastric antrum tissues from patients and controls were collected (control: *n* = 30; T2DM: *n* = 30; *H. pylori*: *n* = 30; T2DM + *H. pylori*: *n* = 30).

*H. pylori* infection was diagnosed by the 13C-urea breath test (Headway Bio-Sci Co., Ltd, Shenzhen, China) according to the manufacturer’s instructions. A delta over baseline of > 4% indicates a positive *H. pylori* infection status. Patients with T2DM were diagnosed based on one of the following American Diabetes Association diagnostic criteria: fasting blood glucose level ≥ 7.0 mmol/L, 2-hour postload glucose level ≥ 11.1 mmol/L during an oral glucose tolerance test, glycated hemoglobin level ≥ 6.5%, or a random plasma glucose level ≥ 11.1 mmol/L in a patient with classic symptoms of hyperglycemia or hyperglycemic crisis. This study was approved by the ethics committee of the First Affiliated Hospital of Nanjing Medical University (2021-SRFA-034). Total RNA was extracted from each tissue sample using TRIzol (Invitrogen, F10488, Waltham, MA, United States), following the manufacturer’s instructions. The kit, EasyScript All-in-One First-Strand cDNA Synthesis SuperMix for RT-qPCR Kit (TransGen Biotech, Beijing, China), was utilized for reverse transcription, with incubations performed at a tempertature 42°C for 15 min and then at 85°C for 15 s. Subsequently, StarLighter SYBR Green RT-qPCR Mix (Universal) (Forever Star, Beijing, China) kit was utilized for RT-qPCR analysis, with an ABI 7500 system (Applied Biosystems, United States). The primers used are listed in Supplementary Table 2. The reaction conditions were as follows: Predenaturation (95°C for 5 min), 40 cycles of denaturation (94°C for 20 s), annealing and extension (60°C for 34 s). β-actin was served as an internal control for RT-qPCR. The 2^-ΔΔCt method was utilized to determine relative the expression levels of genes. Statistical analysis was performed using GraphPad Prism (Version 9.0, Boston, MA, United States). Expression differences of hub genes were compared using one-way ANOVA in four groups (control, *H. pylori* infection, T2DM, and T2DM with *H. pylori* infection), and pairwise comparisons within the two groups were performed using Student’s *t* test. Statistically significant was defined as *P* < 0.05.

**RESULTS**

***Identification of DEGs***

Figure 1 illustrated the overall study design. In brief, a total of 3541, 2186 and 1364 DEGs were identified from the GSE60427, GSE217411 and GSE115601 datasets, respectively. In the GEO datasets, volcano plots (Figure 2A-C) and heatmaps (Supplementary Figure 1) were used to illustrate the dysregulated genes (including upregulated and downregulated). Among these datasets, 67 common DEGs were extracted, including 48 upregulated and 19 downregulated genes (Supplementary Table 3; Figure 2D).

***Functional annotation of DEGs***

After DEGs were selected, GO and KEGG pathway enrichment analyses were performed to explore the biological functions of these genes involving three functional categories: BP, MF, and CC. Major BP terms associated with DEGs included regulation of the immune effector process, neutrophil activation and neutrophil mediated immunity (Figure 3A). Major CC terms associated with these DEGs included the secretory granule membrane, blood microparticle, and tertiary granule (Figure 3B). Finally, MF-associated GO terms were mainly associated with sulfur compound binding, heparin binding, glycosaminoglycan binding, *etc.* (Figure 3C). According to KEGG pathway analysis results, the DEGs were mainly enriched for pathways related to complement and coagulation cascades, *Staphylococcus aureus* infection, and neutrophil extracellular trap formation (Figure 3D).

***PPI network construction and hub gene selection***

The PPI network of DEGs obtained from STRING was subjected to the MCODE plugin of Cytoscape to analyze significant modules. A total of 38 nodes and 84 edges were mapped in the PPI network (Figure 4A). From these modules, the top functional cluster of modules was selected based on the cutoff criteria of node > 3 and score > 3 (Figure 4B).

Then, the key genes with degree connectivity were ranked by the CytoHubba plugin of Cytoscape. Finally, five intersecting genes (*TLR4*, *ITGAM*, *C5AR1*, *FCER1G* and *FCGR2A*) with the highest degree were considered hub genes for further analyses (Figure 4C and D).

***Validation of hub genes in human gastric tissues***

Expression levels of the five hub genes in the three datasets are shown in Supplementary Figure 2; and were significantly upregulated in patients with either *H. pylori* infection or T2DM alone compared to negative controls. Human gastric tissues from four groups were collected (control group, *H. pylori* infection alone group, T2DM alone group and T2DM with *H. pylori* infection group). All included patients underwent upper gastrointestinal endoscopy and were pathologically diagnosed with chronic superficial gastritis without acute inflammation or atrophy according to the Sydney System[26]. The baseline characteristics of the groups are shown in Supplementary Table 4. Through RT-qPCR analysis, we found that *TLR4*, *ITGAM*, *C5AR1*, *FCER1G* and *FCGR2*A were expressed at significantly higher levels in the T2DM with *H. pylori* infection group (*P* < 0.05) than in the T2DM group or the *H. pylori* infection group alone (Figure 4E).

***Immune infiltration analysis***

Using the CIBERSORT algorithm, we explored differences in immune infiltration between *H. pylori-*infected versus normal gastric tissues. Compared with normal tissues, *H. pylori-*infected gastric tissues generally contained a higher proportion of regulatory T cells, activated NK cells, eosinophils and neutrophils, whereas the proportions of plasma cells, activated mast cells and M2 macrophages were lower in *H. pylori*-infected gastric tissues (Figure 5A and B).

The results obtained using TIMER showed that *TLR4* and *ITGAM* expression correlated positively with CD8+ T cells, CD4+ T cells, macrophages, neutrophils, and DCs. *C5AR1*, *FCER1G* and *FCGR2A* expression was significantly associated with infiltration of B cells, CD8+ T cells, macrophages, neutrophils, and DCs, among which their mRNA expression levels all correlated negatively with B cells (Figure 5C).

***Prediction of further miRNA and analysis of gene-miRNA network***

A total of 225 miRNAs was predicted after we uploading the 5 identified hub genes to the miRWalk database. The gene–miRNA interaction network is shown in Figure 6A. We detected 13 miRNAs (miR-6848-5p, miR-6796-5p, miR-6740-5p, miR-8060, miR-6730-5p, miR-5698, miR-12119, miR-6881-5p, miR-6846-5p, miR-7703, miR-6728-5p, miR-7107-5p and miR-1914-3p) associated with at least two gene cross-links, as shown in Supplementary Table 5.

***TF-gene interaction network***

The top ranked TFs were SPI1, MECOM, GATA2, TP63, SALL4, GATA1, MITF, RUNX1 and FLI1 (Figure 6B). Based on the results, we found that *TLR4* was coregulated by 26 TFs, the highest among the identified hub genes.

***Functional analysis of hub genes by single-gene GSEA***

We performed GSEA on *TLR4*, *ITGAM*, *C5AR1*, *FCER1G* and *FCGR2A* to explore the role of these genes in the course of *H. pylori* infection and T2DM and found the top 10 significant items (Figure 7). According to GSEA results, it suggested that all these five genes play a direct or indirect role in the pathogenesis of *H. pylori* infection and T2DM. For example, *FCG2A* is involved in the signaling pathway of “type 1 diabetes mellitus” and the “insulin signaling pathway”, *C5AR1* and *FCER1G* are involved in the signaling pathway of “type 1 diabetes mellitus”, and *ITGAM* is involved in the signaling pathway of “glycosaminoglycan biosynthesis chondroitin sulfate”.

**DISCUSSION**

Approximately 50% of the world’s population is infected with *H. pylori*, and the infection rate is even higher in patients with T2DM. Infected patients with T2DM have worse blood glucose control abilities, with great social and economic burdens[7,8]. However, the detailed mechanism of the interaction between T2DM and *H. pylori* infection remains unknown. Therefore, it is necessary to increase our understanding of the underlying mechanisms leading to the risk of *H. pylori* infection and T2DM to develop effective treatment approaches.

In this study, we investigated the biological functions, expression levels, and correlations with immune infiltrates of common genes with significantly altered expression in both *H. pylori-*infected individuals and T2DM patients through integrated bioinformatics analyses. Our results showed that expression of 67 overlapping genes was altered in gastric samples from both *H. pylori*-infected individuals and T2DM patients. Among these genes, 48 were upregulated and 19 downregulated. Five hub genes were further identified through PPI analysis. However, regardless of the statistical probability, the causality between a candidate genotype and the phenotype of the host remains uncertain[27]. To further identify the relationship between genotype (the 5 hub genes) and phenotype (*H. pylori*-associated T2DM), rigorous validation of mechanisms at the molecular, cellular, tissue, and whole-organism levels is needed.

Chronic low-grade inflammation has been definitively shown to correspond with obesity[28] and diabetes[29]. However, whether obesity and diabetes drive the inflammation or vice versa remains to be elucidated. Gut microbiota play a critical role in the development of the host immune system, making it an important immune organ[30]. Disturbance of the gut microbiota promotes inflammation within the lining of the intestines[31]. The dysbiosis of the gut results in bacterial infiltration, allowing microbes to contact the epithelium and causing inflammation[32]. Toll-like receptors (TLR) play a key role in host recognition of microbes[33]. *TLR4* has been implicated in recognition of bacterial lipopolysaccharides, a key element of the cell walls of gram-negative bacteria. This triggers the expression of proinflammatory cytokines and chemokines, including tumor necrosis factor-alpha[34]. This inflammatory response is strongly linked to insulin resistance, and both *TLR4* and its coreceptor CD14 are needed to induce insulin resistance in mice[35]. It is believed that *TLR4*, one of the TLR family members, possesses the potential to trigger nuclear factor-κB when confronted with short-chain fatty acids. Consequently, this leads to subsequent stimulation of the immune system[36]. Therefore, the inflammation caused by *TLR4* serves a crucial function in the development of T2DM related to *H. pylori*. The study conducted by Devaraj *et al*[37] exhibited a notable rise in the level of *TLR4* expression among individuals diagnosed with type 1 diabetes. This finding implies that *TLR4* actively participates in the inflammatory state associated with diabetes. Moreover, knockout of *TLR4* alleviated inflammation in rats with diabetes and *TLR4* antagonists attenuated atherogenesis in mice with diabetes[38]. Based on our results, we speculated that *TLR4* participates in the pathogenesis of H. pylori-associated T2DM *via* the TLR signaling pathway.

Other hub genes, *ITGAM*[39], *C5AR1*[40], *FCER1G*[41] and *FCGR2A*[42], are also reported to be associated with diabetes. *ITGAM*, a monocyte/macrophage marker, is upregulated in T2DM patients[39]. *FCER1G* was identified as a significant gene related to diabetic kidney disease. Gene Expression Omnibus validation using additional datasets showed that FCER1G is upregulated in diabetic glomerular lesions compared with normal tissues. This report also revealed that abnormal upregulation of *FCER1G* is related to diabetic glomerular lesions[41].

Clinical variability between individuals infected with any pathogen is enormous, ranging from silent to lethal. One of the main reasons is immunity differs among individuals[43]. Tumor-infltrating immune cells function together to defend the body against invading factors, such as bacterial infection. Therefore, they can be used as important predictors for diagnosis and treatment of diseases[44]. Based on KEGG pathway and immune cell infiltration analyses, we found that *H. pylori* infection is associated with multiple immune cell changes, especially NK cells and regulatory T cells. Through single-gene GSEA, we found that high expression of the hub genes *TLR4*, *FCGR2A*, and *FCER1G* was associated with NK cell-mediated cytotoxicity in diabetes, which suggests that *H. pylori* infection might change hub gene expression and downstream NK cells to induce T2DM. Further analysis suggested that these 5 hub genes all correlated with B cells, CD8+ T cells, macrophages, neutrophils, and DCs. It has been shown that isolated NK cells from T2DM subjects show defects in the NK cell-activating receptors NKG2D and NKp46, in association with functional defects in NK degranulation capacity[45]. Restrepo *et al*[46] demonstrated that chronic hyperglycaemia is significantly associated with defects in complement receptors and Fcγ receptors on isolated monocytes, resulting in phagocytosis impairment. An *in vitro* study using macrophages derived from mouse bone marrow and treated with high glucose showed reduced antibacterial activity and phagocytosis for the treated macrophages[47]. In the same study, reduced phagocytosis was shown in peritoneal macrophages from mice with T2DM. This might be related to the reduced glycolytic capacity and reserve of macrophages following long-term sensitization to high levels of glucose. Reactive oxygen species production was reportedly reduced in isolated neutrophils from T2DM tuberculosis patients following phorbol 12-myristate 13-acetate stimulation, and this defect in reactive oxygen species production was associated with increased levels of resistin in T2DM patient serum[48]. In a comparable study, Perner *et al*[49] documented the inhibition of superoxide production in neutrophils isolated from healthy individuals when subjected to a high-glucose environment. This hindrance was observed to be a consequence of the suppression of glucose-6-phosphate dehydrogenase, which disrupted the generation of nicotinamide adenine dinucleotide phosphate. Thus, we speculate that these 5 hub genes are involved in *H. pylori*-associated T2DM through immune infiltration. We will validate their relationship through experiments in the future.

This study provides some new insights into the pathogenesis of *H. pylori-*associated T2DM. However, several limitations should be mentioned. First of all, this study had a relatively small sample size and a larger sample size would be necessary for further investigations. Secondly, hub genes were identified using bioinformatics analysis and validated by a small clinical sample. Validation including RNA-seq from a larger clinical cohort is needed. It is necessary to investigate the potential underlying mechanisms involved in these findings in future large-scale prospective studies. Thirdly, despite statistical probability, the causality between a candidate genotype and the phenotype of the host is uncertain[27]. To identify the relationship between genotype (the 5 hub genes) and phenotype (*H. pylori*-associated T2DM), rigorous validation of mechanisms at the molecular, cellular, tissue, and whole-organism levels is needed.

**CONCLUSION**

We report 67 common DEGs and five hub genes (*TLR4*, *ITGAM*, *C5AR1*, *FCER1G* and *FCGR2A*) in *H. pylori* infection and T2DM. We validated expression of the five hub genes by RT-qPCR. All hub genes were significantly upregulated in T2DM patients with *H. pylori* infection compared with noninfected T2DM patients. Immune infiltration analysis showed that *H. pylori-*infected gastric tissues generally contained a higher proportion of regulatory T cells, activated NK cells, eosinophils and neutrophils. Our gene-miRNA analysis detected 13 miRNAs with at least two gene cross-links, and TF-gene interaction networks showed that *TLR4* to be coregulated by 26 TFs, the largest number of TFs among the 5 hub genes. This study provides a new idea for elucidating the pathogenesis of H. pylori-associated T2DM at the genetic level.

**ARTICLE HIGHLIGHTS**

***Research background***

This prevalence rate of *Helicobacter pylori* (*H. pylori*) is high, especially in less developed countries. Its infection related to not only gastric diseases but also extragastric diseases such as type 2 diabetes mellitus (T2DM). However, the underlying mechanisms connecting *H. pylori* infection and T2DM remains unclear.

***Research motivation***

The potential molecular connections between *H. pylori* infection and T2DM are needed to be identified, in order to further elucidate the pathogenesis and the new treatment strategy of *H. pylori-*infected T2DM.

***Research objectives***

We aimed to explore the potential molecular connections between *H. pylori* infection and T2DM using bioinformatics analysis. In the future research, we will investigating these identified genes and downstream signaling pathway to further understand their relationship.

***Research methods***

Differentially expressed genes from three datasets commonly present in patients with *H. pylori* infection and T2DM were identified. Hub genes were validated by RT-qPCR using human gastric biopsy samples. Correlations between hub genes and immune cell infiltration, miRNAs, and transcription factors were further analyzed.

***Research results***

This is the first study to identify the key genes and pathways associated with *H. pylori* infection and T2DM using integrated bioinformatics analysis. We identified five hub genes, all of which were closely related to immune cell infiltration.

***Research conclusions***

We were the first to find out that the 5 hub genes identified are playing important roles in the pathogenesis of *H. pylori-*infected T2DM.

***Research perspectives***

It is necessary to investigate the potential underlying mechanisms involved in these findings in future large-scale prospective studies.

**REFERENCES**

1 **Butt J**, Epplein M. How do global trends in Helicobacter pylori prevalence inform prevention planning? *Lancet Gastroenterol Hepatol* 2023; **8**: 498-499 [PMID: 37086740 DOI: 10.1016/S2468-1253(23)00101-2]

2 **Alvarez CS**, Florio AA, Butt J, Rivera-Andrade A, Kroker-Lobos MF, Waterboer T, Camargo MC, Freedman ND, Graubard BI, Lazo M, Guallar E, Groopman JD, Ramírez-Zea M, McGlynn KA. Associations between Helicobacter pylori with nonalcoholic fatty liver disease and other metabolic conditions in Guatemala. *Helicobacter* 2020; **25**: e12756 [PMID: 33006810 DOI: 10.1111/hel.12756]

3 **Zhang P**, He Q, Song D, Wang Y, Liu X, Ding G, Xing W. Association of Helicobacter pylori Infection With Carotid Atherosclerosis in a Northern Chinese Population: A Cross-Sectional Study. *Front Cardiovasc Med* 2021; **8**: 795795 [PMID: 35174222 DOI: 10.3389/fcvm.2021.795795]

4 **Wang L**, Cao ZM, Zhang LL, Dai XC, Liu ZJ, Zeng YX, Li XY, Wu QJ, Lv WL. Helicobacter Pylori and Autoimmune Diseases: Involving Multiple Systems. *Front Immunol* 2022; **13**: 833424 [PMID: 35222423 DOI: 10.3389/fimmu.2022.833424]

5 **Mansori K**, Dehghanbanadaki H, Naderpour S, Rashti R, Moghaddam AB, Moradi Y. A systematic review and meta-analysis of the prevalence of Helicobacter pylori in patients with diabetes. *Diabetes Metab Syndr* 2020; **14**: 601-607 [PMID: 32417710 DOI: 10.1016/j.dsx.2020.05.009]

6 **Zhang K**, Ma Y, Luo Y, Song Y, Xiong G, Ma Y, Sun X, Kan C. Metabolic diseases and healthy aging: identifying environmental and behavioral risk factors and promoting public health. *Front Public Health* 2023; **11**: 1253506 [PMID: 37900047 DOI: 10.3389/fpubh.2023.1253506]

7 **Pan Y**, Zhong S, Zhou K, Tian Z, Chen F, Liu Z, Geng Z, Li S, Huang R, Wang H, Zou W, Hu J. Association between Diabetes Complications and the Triglyceride-Glucose Index in Hospitalized Patients with Type 2 Diabetes. *J Diabetes Res* 2021; **2021**: 8757996 [PMID: 34671683 DOI: 10.1155/2021/8757996]

8 **Wang M**, He Y, He Q, Di F, Zou K, Wang W, Sun X. Comparison of clinical characteristics and disease burden between early- and late-onset type 2 diabetes patients: a population-based cohort study. *BMC Public Health* 2023; **23**: 2411 [PMID: 38049796 DOI: 10.1186/s12889-023-17280-5]

9 **Zhou X**, Zhang C, Wu J, Zhang G. Association between Helicobacter pylori infection and diabetes mellitus: a meta-analysis of observational studies. *Diabetes Res Clin Pract* 2013; **99**: 200-208 [PMID: 23395214 DOI: 10.1016/j.diabres.2012.11.012]

10 **Bener A**, Ağan AF, Al-Hamaq AOAA, Barisik CC, Öztürk M, Ömer A. Prevalence of Helicobacter pylori Infection among Type 2 Diabetes Mellitus. *Adv Biomed Res* 2020; **9**: 27 [PMID: 33072639 DOI: 10.4103/abr.abr\_248\_19]

11 **Yao CC**, Kuo CM, Hsu CN, Yang SC, Wu CK, Tai WC, Liang CM, Wu KL, Huang CF, Bi KW, Lee CH, Chuah SK. First-line Helicobacter pylori eradication rates are significantly lower in patients with than those without type 2 diabetes mellitus. *Infect Drug Resist* 2019; **12**: 1425-1431 [PMID: 31239721 DOI: 10.2147/IDR.S194584]

12 **Song X**, Cai C, Jin Q, Chen X, Yu C. The efficacy of Helicobacter pylori eradication in diabetics and its effect on glycemic control: A systematic review and meta-analysis. *Helicobacter* 2021; **26**: e12781 [PMID: 33465265 DOI: 10.1111/hel.12781]

13 **Thai TD**, Chuenchom C, Donsa W, Faksri K, Sripa B, Edwards SW, Salao K. Helicobacter pylori extract induces purified neutrophils to produce reactive oxygen species only in the presence of plasma. *Biomed Rep* 2023; **19**: 89 [PMID: 37901879 DOI: 10.3892/br.2023.1671]

14 **Han L**, Shu X, Wang J. Helicobacter pylori-Mediated Oxidative Stress and Gastric Diseases: A Review. *Front Microbiol* 2022; **13**: 811258 [PMID: 35211104 DOI: 10.3389/fmicb.2022.811258]

15 **Wu YY**, Hsieh CT, Tsay GJ, Kao JT, Chiu YM, Shieh DC, Lee YJ. Recruitment of CCR6(+ ) Foxp3(+) regulatory gastric infiltrating lymphocytes in Helicobacter pylori gastritis. *Helicobacter* 2019; **24**: e12550 [PMID: 30412323 DOI: 10.1111/hel.12550]

16 **Zhou X**, Liu W, Gu M, Zhou H, Zhang G. Helicobacter pylori infection causes hepatic insulin resistance by the c-Jun/miR-203/SOCS3 signaling pathway. *J Gastroenterol* 2015; **50**: 1027-1040 [PMID: 25689935 DOI: 10.1007/s00535-015-1051-6]

17 **Martin-Nuñez GM**, Cornejo-Pareja I, Clemente-Postigo M, Tinahones FJ. Gut Microbiota: The Missing Link Between Helicobacter pylori Infection and Metabolic Disorders? *Front Endocrinol (Lausanne)* 2021; **12**: 639856 [PMID: 34220702 DOI: 10.3389/fendo.2021.639856]

18 **Pfeifer SP**. From next-generation resequencing reads to a high-quality variant data set. *Heredity (Edinb)* 2017; **118**: 111-124 [PMID: 27759079 DOI: 10.1038/hdy.2016.102]

19 **Davis S**, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 2007; **23**: 1846-1847 [PMID: 17496320 DOI: 10.1093/bioinformatics/btm254]

20 **Sufyan M**, Ali Ashfaq U, Ahmad S, Noor F, Hamzah Saleem M, Farhan Aslam M, El-Serehy HA, Aslam S. Identifying key genes and screening therapeutic agents associated with diabetes mellitus and HCV-related hepatocellular carcinoma by bioinformatics analysis. *Saudi J Biol Sci* 2021; **28**: 5518-5525 [PMID: 34588861 DOI: 10.1016/j.sjbs.2021.07.068]

21 **Huang DW**, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC, Lempicki RA. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol* 2007; **8**: R183 [PMID: 17784955 DOI: 10.1186/gb-2007-8-9-r183]

22 **Rosandić M**, Paar V, Gluncić M, Basar I, Pavin N. Key-string algorithm--novel approach to computational analysis of repetitive sequences in human centromeric DNA. *Croat Med J* 2003; **44**: 386-406 [PMID: 12950141]

23 **Liu S**, Ren W, Yu J, Li C, Tang S. Identification of Hub Genes Associated with Diabetes Mellitus and Tuberculosis Using Bioinformatic Analysis. *Int J Gen Med* 2021; **14**: 4061-4072 [PMID: 34354368 DOI: 10.2147/IJGM.S318071]

24 **Wang Y**, Zhao M, Zhang Y. Identification of fibronectin 1 (FN1) and complement component 3 (C3) as immune infiltration-related biomarkers for diabetic nephropathy using integrated bioinformatic analysis. *Bioengineered* 2021; **12**: 5386-5401 [PMID: 34424825 DOI: 10.1080/21655979.2021.1960766]

25 **Stolte M**, Meining A. The updated Sydney system: classification and grading of gastritis as the basis of diagnosis and treatment. *Can J Gastroenterol* 2001; **15**: 591-598 [PMID: 11573102 DOI: 10.1155/2001/367832]

26 **Cheng KP**, Yang YJ, Hung HC, Lin CH, Wu CT, Hung MH, Sheu BS, Ou HY. Helicobacter pylori eradication improves glycemic control in type 2 diabetes patients with asymptomatic active Helicobacter pylori infection. *J Diabetes Investig* 2019; **10**: 1092-1101 [PMID: 30556347 DOI: 10.1111/jdi.12991]

27 **Casanova JL**, Su HC; COVID Human Genetic Effort. A Global Effort to Define the Human Genetics of Protective Immunity to SARS-CoV-2 Infection. *Cell* 2020; **181**: 1194-1199 [PMID: 32405102 DOI: 10.1016/j.cell.2020.05.016]

28 **She Y**, Mangat R, Tsai S, Proctor SD, Richard C. The Interplay of Obesity, Dyslipidemia and Immune Dysfunction: A Brief Overview on Pathophysiology, Animal Models, and Nutritional Modulation. *Front Nutr* 2022; **9**: 840209 [PMID: 35252310 DOI: 10.3389/fnut.2022.840209]

29 **Koh GY**, Rowling MJ, Pritchard SK. Possible role of type 1 and type 2 taste receptors on obesity-induced inflammation. *Nutr Rev* 2022; **80**: 1919-1926 [PMID: 35150265 DOI: 10.1093/nutrit/nuac007]

30 **Bonde A**, Daly S, Kirsten J, Kondapaneni S, Mellnick V, Menias CO, Katabathina VS. Human Gut Microbiota-associated Gastrointestinal Malignancies: A Comprehensive Review. *Radiographics* 2021; **41**: 1103-1122 [PMID: 33989072 DOI: 10.1148/rg.2021200168]

31 **Lock JY**, Caboni M, Strandwitz P, Morrissette M, DiBona K, Joughin BA, Lewis K, Carrier RL. An *in vitro* intestinal model captures immunomodulatory properties of the microbiota in inflammation. *Gut Microbes* 2022; **14**: 2039002 [PMID: 35316142 DOI: 10.1080/19490976.2022.2039002]

32 **Bui TI**, Gill AL, Mooney RA, Gill SR. Modulation of Gut Microbiota Metabolism in Obesity-Related Type 2 Diabetes Reduces Osteomyelitis Severity. *Microbiol Spectr* 2022; **10**: e0017022 [PMID: 35315698 DOI: 10.1128/spectrum.00170-22]

33 **Rong Z**, Huang Y, Cai H, Chen M, Wang H, Liu G, Zhang Z, Wu J. Gut Microbiota Disorders Promote Inflammation and Aggravate Spinal Cord Injury Through the TLR4/MyD88 Signaling Pathway. *Front Nutr* 2021; **8**: 702659 [PMID: 34589510 DOI: 10.3389/fnut.2021.702659]

34 **Noori MS**, Courreges MC, Bergmeier SC, McCall KD, Goetz DJ. Modulation of LPS-induced inflammatory cytokine production by a novel glycogen synthase kinase-3 inhibitor. *Eur J Pharmacol* 2020; **883**: 173340 [PMID: 32634441 DOI: 10.1016/j.ejphar.2020.173340]

35 **Lu Z**, Zhang X, Li Y, Lopes-Virella MF, Huang Y. TLR4 antagonist attenuates atherogenesis in LDL receptor-deficient mice with diet-induced type 2 diabetes. *Immunobiology* 2015; **220**: 1246-1254 [PMID: 26162692 DOI: 10.1016/j.imbio.2015.06.016]

36 **Yuan Y**, Lu L, Bo N, Chaoyue Y, Haiyang Y. Allicin Ameliorates Intestinal Barrier Damage *via* Microbiota-Regulated Short-Chain Fatty Acids-TLR4/MyD88/NF-κB Cascade Response in Acrylamide-Induced Rats. *J Agric Food Chem* 2021; **69**: 12837-12852 [PMID: 34694121 DOI: 10.1021/acs.jafc.1c05014]

37 **Devaraj S**, Tobias P, Jialal I. Knockout of toll-like receptor-4 attenuates the pro-inflammatory state of diabetes. *Cytokine* 2011; **55**: 441-445 [PMID: 21498084 DOI: 10.1016/j.cyto.2011.03.023]

38 **Ekuni D**, Yoneda T, Endo Y, Kasuyama K, Irie K, Mizutani S, Azuma T, Tomofuji T, Morita M. Occlusal disharmony accelerates the initiation of atherosclerosis in apoE knockout rats. *Lipids Health Dis* 2014; **13**: 144 [PMID: 25189624 DOI: 10.1186/1476-511X-13-144]

39 **Westerbacka J**, Cornér A, Kolak M, Makkonen J, Turpeinen U, Hamsten A, Fisher RM, Yki-Järvinen H. Insulin regulation of MCP-1 in human adipose tissue of obese and lean women. *Am J Physiol Endocrinol Metab* 2008; **294**: E841-E845 [PMID: 18270300 DOI: 10.1152/ajpendo.00653.2006]

40 **Li L**, Wei T, Liu S, Wang C, Zhao M, Feng Y, Ma L, Lu Y, Fu P, Liu J. Complement C5 activation promotes type 2 diabetic kidney disease *via* activating STAT3 pathway and disrupting the gut-kidney axis. *J Cell Mol Med* 2021; **25**: 960-974 [PMID: 33280239 DOI: 10.1111/jcmm.16157]

41 **Liu S**, Wang C, Yang H, Zhu T, Jiang H, Chen J. Weighted gene co-expression network analysis identifies FCER1G as a key gene associated with diabetic kidney disease. *Ann Transl Med* 2020; **8**: 1427 [PMID: 33313172 DOI: 10.21037/atm-20-1087]

42 **Mehrbod P**, Eybpoosh S, Farahmand B, Fotouhi F, Khanzadeh Alishahi M. Association of the host genetic factors, hypercholesterolemia and diabetes with mild influenza in an Iranian population. *Virol J* 2021; **18**: 64 [PMID: 33766078 DOI: 10.1186/s12985-021-01486-3]

43 **Casanova JL**. Human genetic basis of interindividual variability in the course of infection. *Proc Natl Acad Sci U S A* 2015; **112**: E7118-E7127 [PMID: 26621739 DOI: 10.1073/pnas.1521644112]

44 **Qi Z**, Yan F, Chen D, Xing W, Li Q, Zeng W, Bi B, Xie J. Identification of prognostic biomarkers and correlations with immune infiltrates among cGAS-STING in hepatocellular carcinoma. *Biosci Rep* 2020; **40** [PMID: 33006365 DOI: 10.1042/BSR20202603]

45 **Gajovic N**, Jurisevic M, Pantic J, Radosavljevic G, Arsenijevic N, Lukic ML, Jovanovic I. Attenuation of NK cells facilitates mammary tumor growth in streptozotocin-induced diabetes in mice. *Endocr Relat Cancer* 2018; **25**: 493-507 [PMID: 29459428 DOI: 10.1530/ERC-17-0529]

46 **Restrepo BI**, Twahirwa M, Rahbar MH, Schlesinger LS. Phagocytosis *via* complement or Fc-gamma receptors is compromised in monocytes from type 2 diabetes patients with chronic hyperglycemia. *PLoS One* 2014; **9**: e92977 [PMID: 24671137 DOI: 10.1371/journal.pone.0092977]

47 **Srinontong P**, Wandee J, Aengwanich W. Paraquat modulates immunological function in bone marrow-derived macrophages. *Acta Vet Hung* 2022 [PMID: 35262507 DOI: 10.1556/004.2022.00003]

48 **Chao WC**, Yen CL, Wu YH, Chen SY, Hsieh CY, Chang TC, Ou HY, Shieh CC. Increased resistin may suppress reactive oxygen species production and inflammasome activation in type 2 diabetic patients with pulmonary tuberculosis infection. *Microbes Infect* 2015; **17**: 195-204 [PMID: 25528597 DOI: 10.1016/j.micinf.2014.11.009]

49 **Perner A**, Nielsen SE, Rask-Madsen J. High glucose impairs superoxide production from isolated blood neutrophils. *Intensive Care Med* 2003; **29**: 642-645 [PMID: 12552364 DOI: 10.1007/s00134-002-1628-4]

**Footnotes**

**Institutional review board statement:** The original data in this study were retrieved from the public GEO database with an open license for data use. This study was approved by the ethic committee of the First Affiliated Hospital of Nanjing Medical University (Approval No. 2022-SR-406).

**Informed consent statement:** All study participants or their legal guardian provided informed written consent about personal and medical data collection prior to study enrollment.

**Conflict-of-interest statement:** The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Data sharing statement:** The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

**Open-Access:** This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: https://creativecommons.org/Licenses/by-nc/4.0/

**Provenance and peer review:** Unsolicited article; Externally peer reviewed.

**Peer-review model:** Single blind

**Peer-review started:** September 21, 2023

**First decision:** November 9, 2023

**Article in press:**

**Specialty type:** Endocrinology and metabolism

**Country/Territory of origin:** China

**Peer-review report’s scientific quality classification**

Grade A (Excellent): 0

Grade B (Very good): B, B

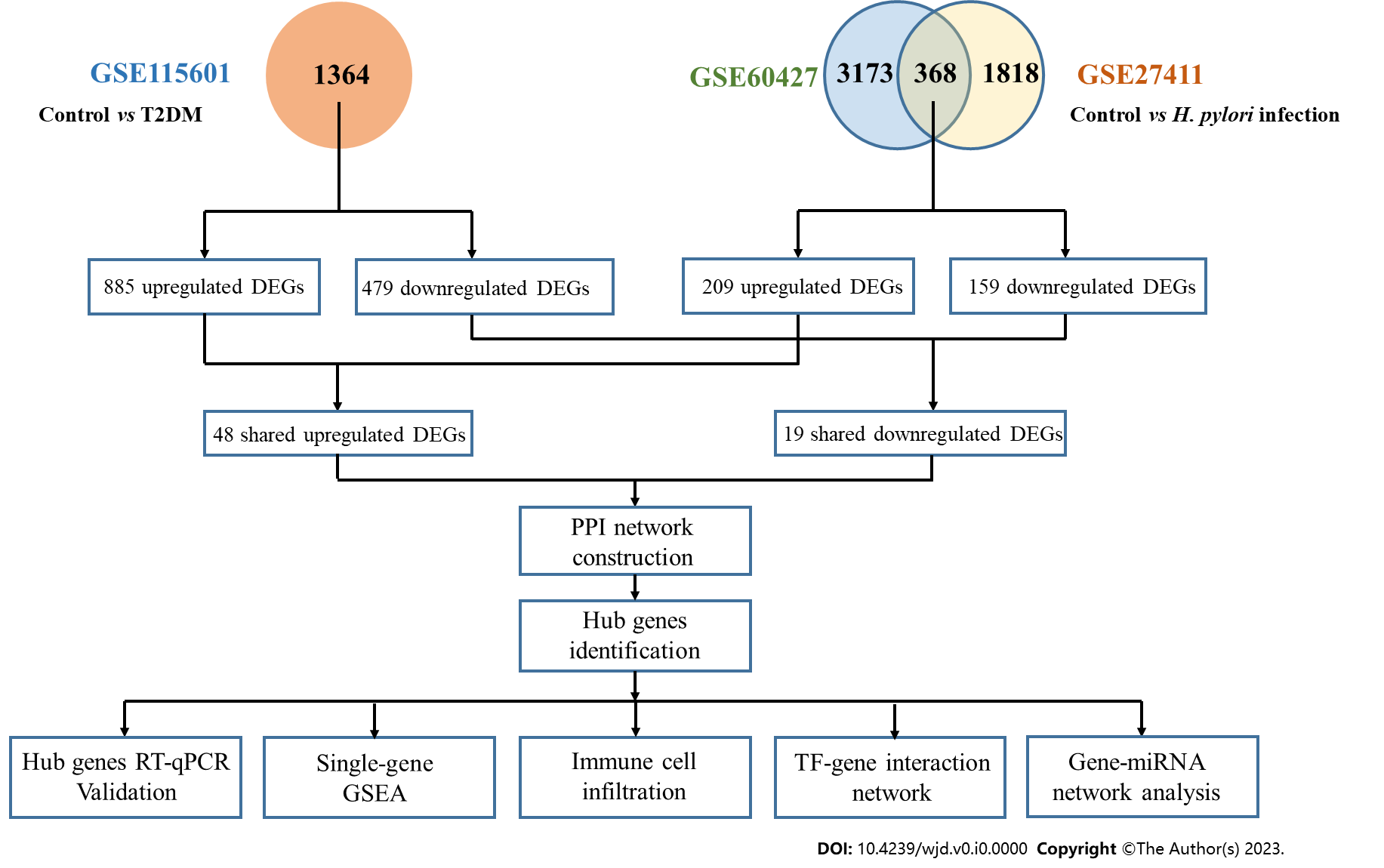
Grade C (Good): C, C, C

Grade D (Fair): 0

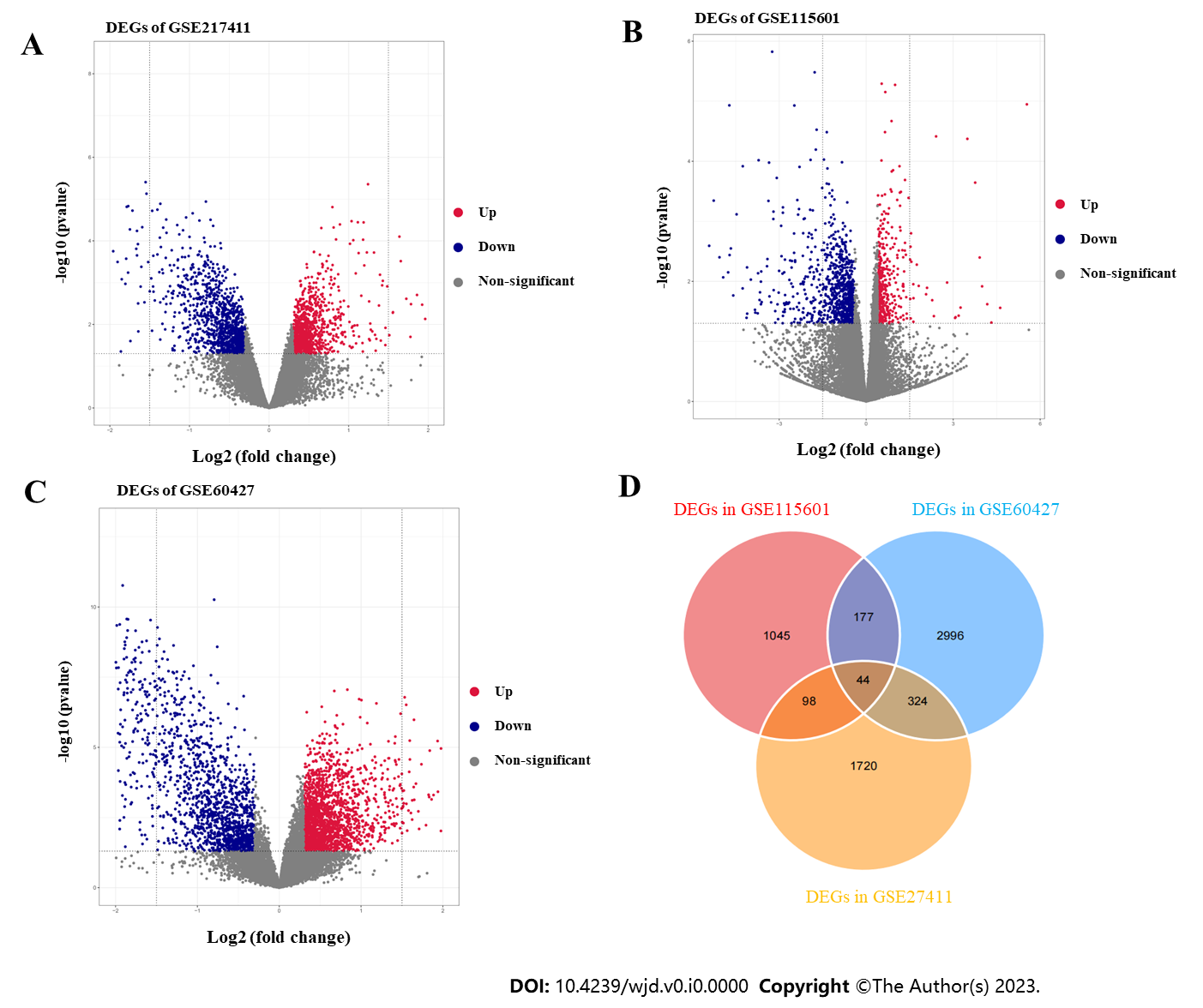
Grade E (Poor): 0

**P-Reviewer:** Arumugam VA, India; Horowitz M, Australia; Suzuki H, Japan **S-Editor:** Lin C **L-Editor:** A **P-Editor:**

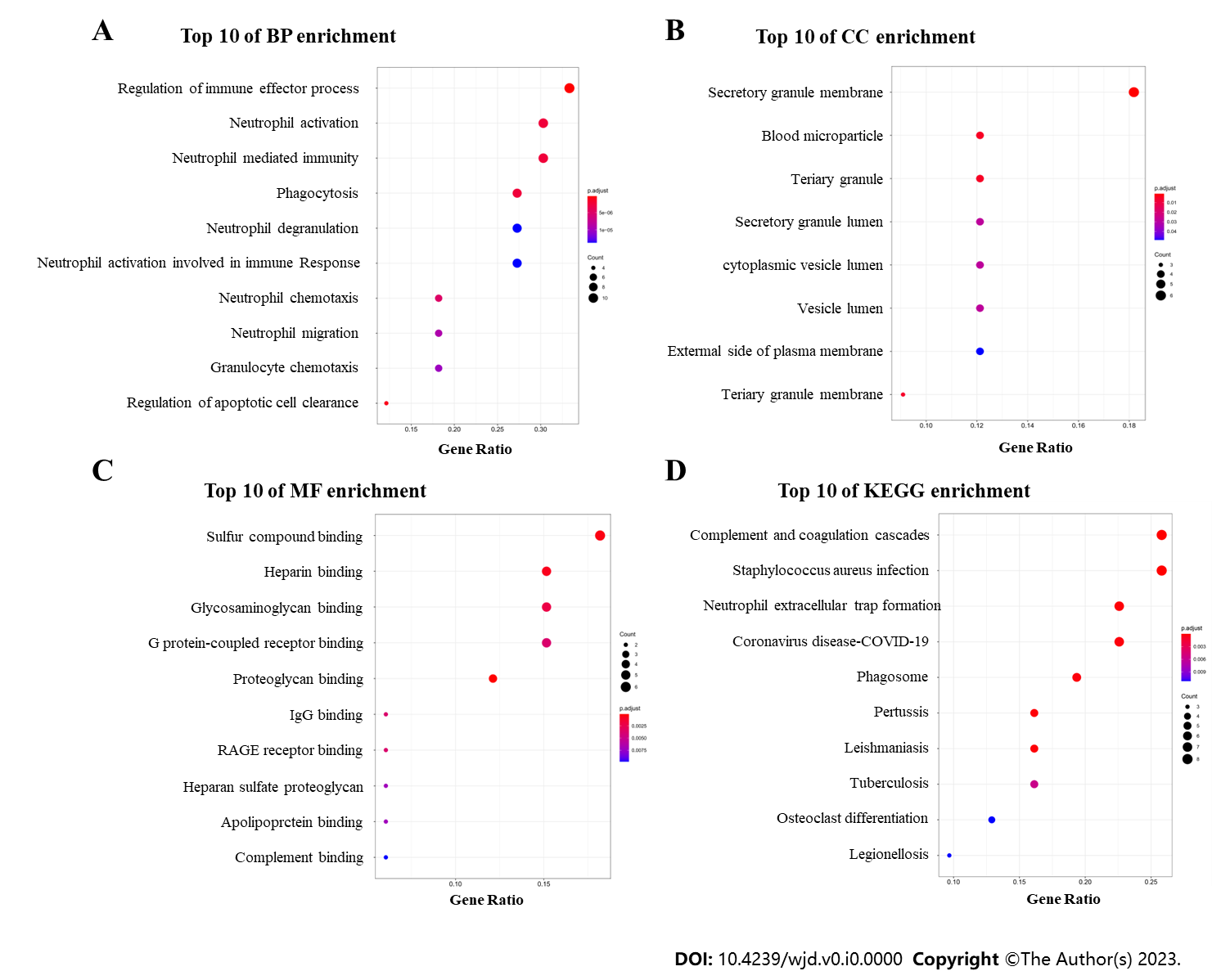
**Figure Legends**



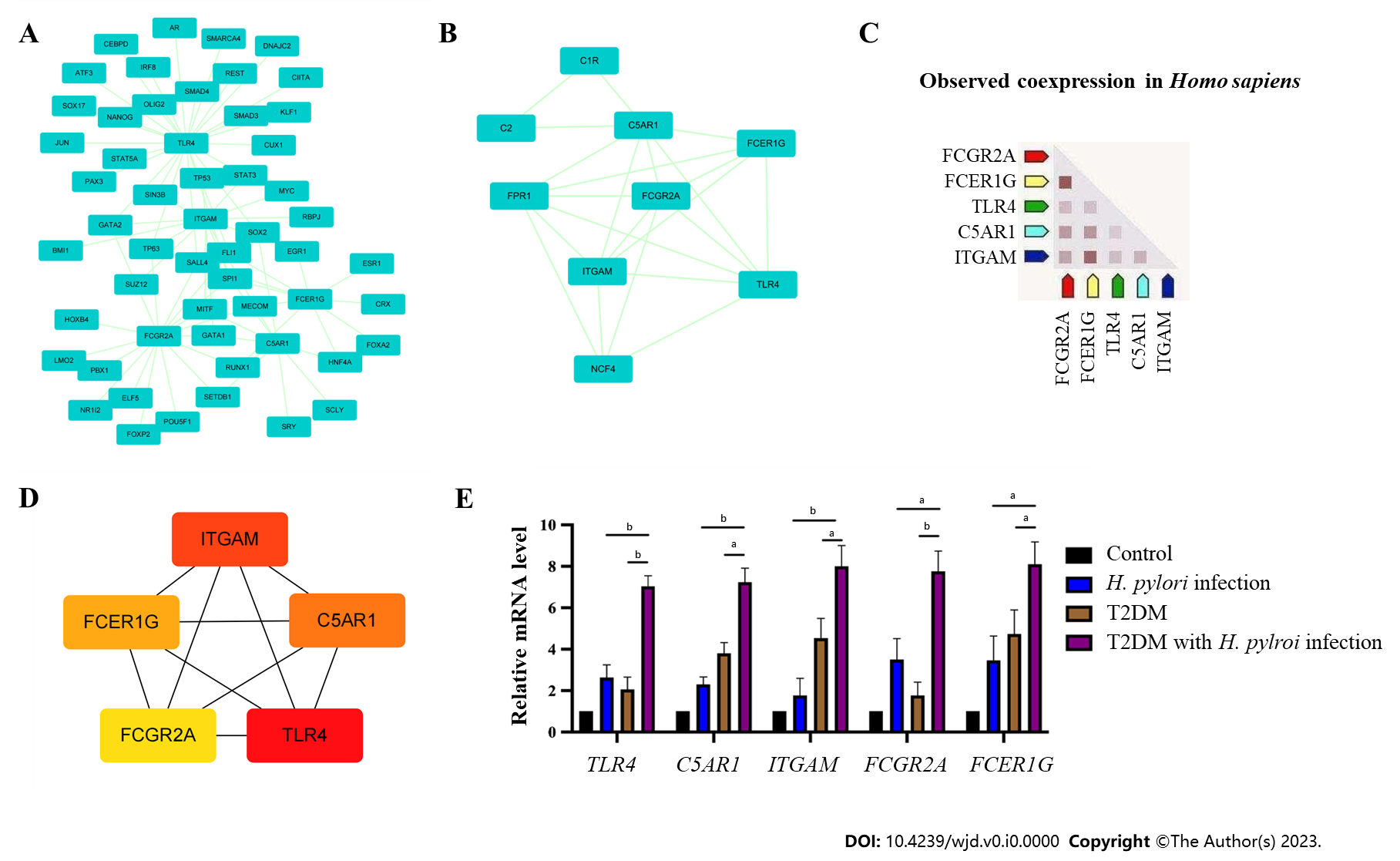
**Figure 1 Overall workflow of the study.** T2DM: Type 2 diabetes mellitus; *H. pylori*: *Helicobacter pylori*; DEGs: Differentially expressed genes; TF: Transcriptional factor; PPI: Protein-protein interaction; GSEA: Gene set enrichment analysis.



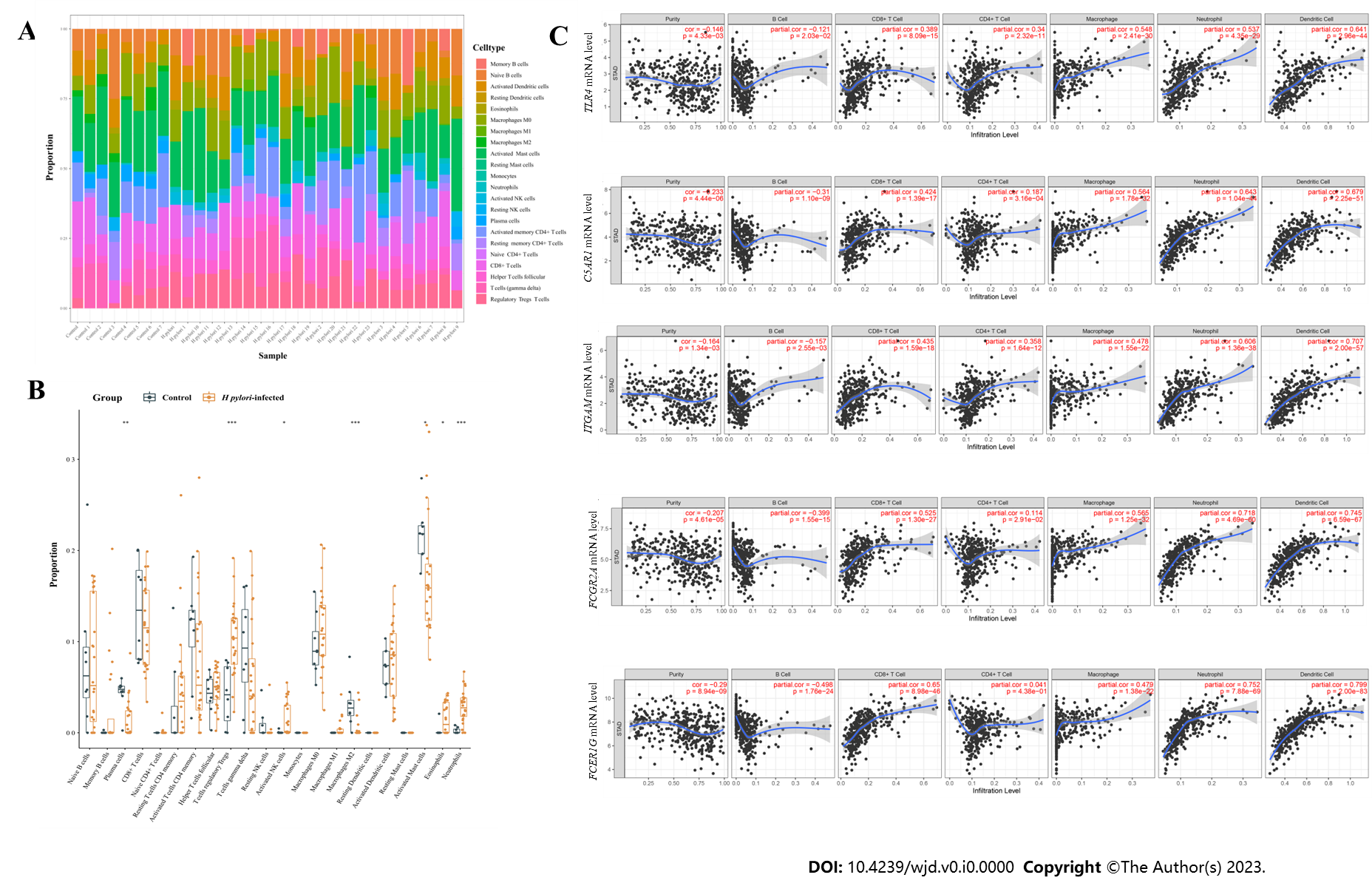
**Figure 2 The expression levels of differentially expressed genes in three datasets.** A-C: The volcano plot distribution of differentially expressed genes (DEGs) of GSE60427 (A), GSE27411 (B) and GSE115601 (C). The blue dots indicate the screened downregulated DEGs, red dots indicate the screened upregulated DEGs, and the grey dots indicate genes with no significant differences; D: The Venn diagram of DEGs based on the three datasets. DEGs: Differentially expressed genes.



**Figure 3 Functional enrichment analysis of common differentially expressed genes.** A: Biological process analysis of differentially expressed genes (DEGs); B: Cellular component analysis of DEGs; C: Molecular function analysis of DEGs; D: Kyoto Encyclopedia of Genes and Genomes pathway analysis of DEGs. BP: Biological process; CC: Cellular component; MF: Molecular function; KEGG: Kyoto Encyclopedia of Genes and Genomes.

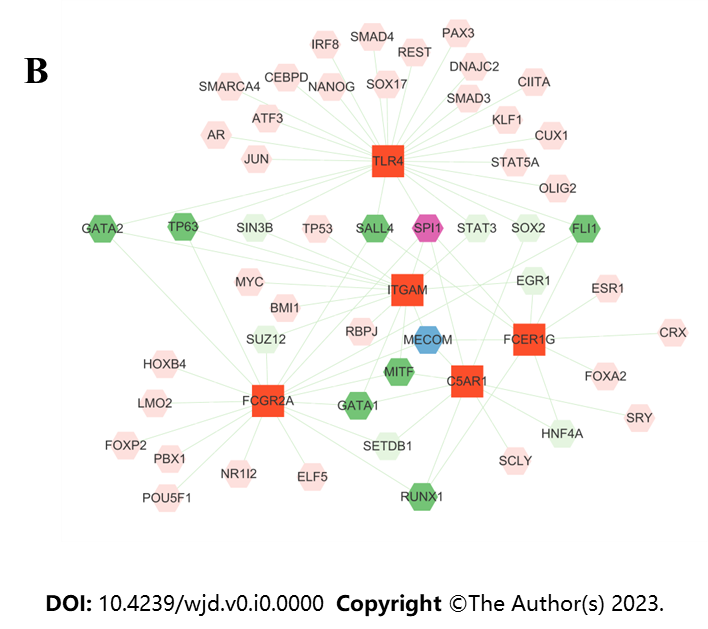


**Figure 4 Protein-protein interaction network showing interactions between common genes and identification of differentially expressed genes from this network.** A: The protein-protein interaction (PPI) network of differentially expressed genes was constructed by Cytoscape software. The criteria of the PPI network were as follows: Confidence score ≥ 0.4 and a maximum number of interactions ≤ 5; B: The top module of the PPI network. MCODE score ≥ 3, 9 nodes and 21 edges; C: Construction of the PPI network among the 5 hub genes; D: Coexpression analysis of the 5 hub genes using STRING; E: The expression of 5 hub genes in clinical specimens by RT-qPCR analysis. a*P* < 0.05; b*P* < 0.01). T2DM: Type 2 diabetes mellitus; *H. pylori*: *Helicobacter pylori*.

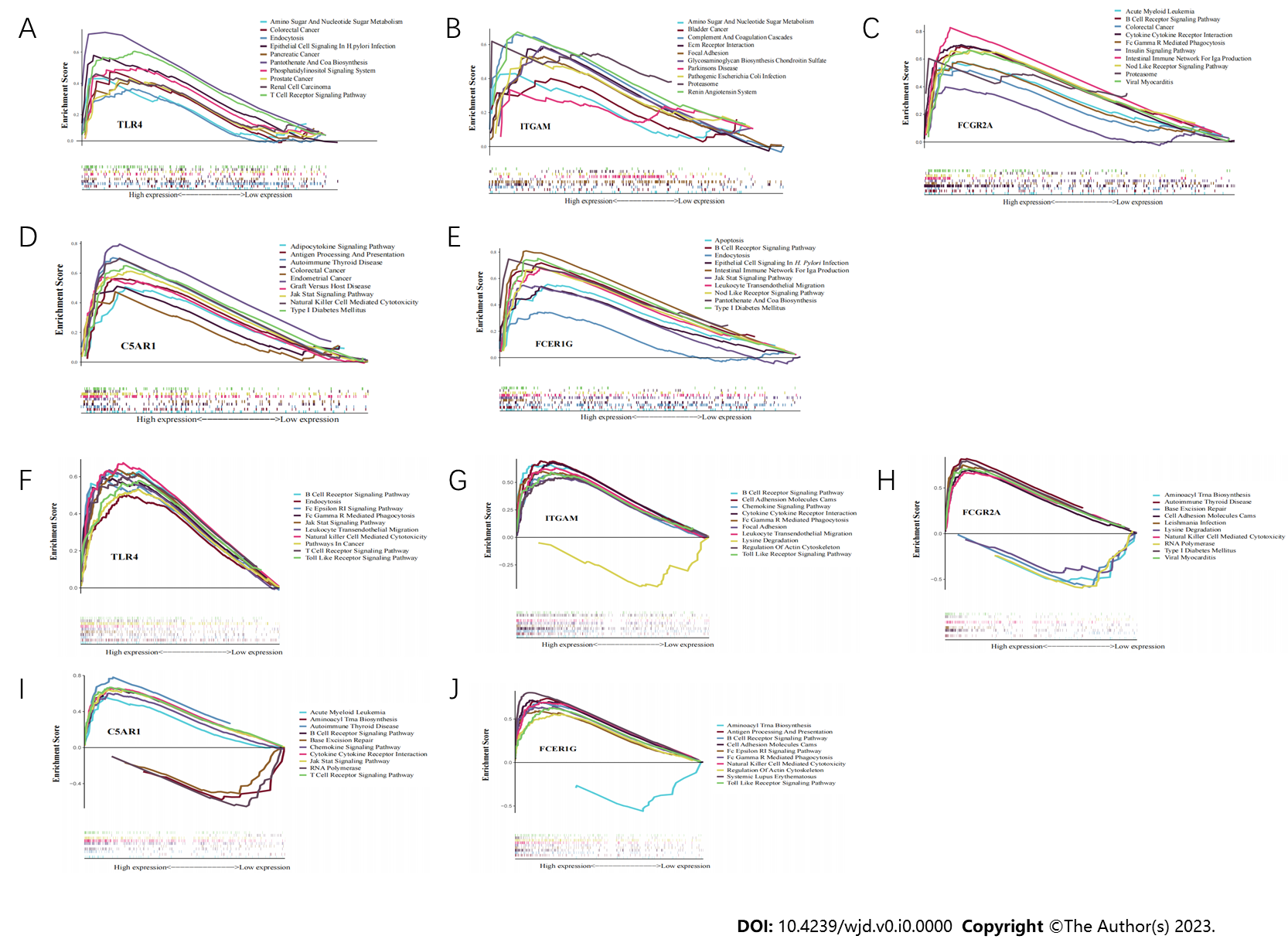


**Figure 5 The relationship between hub genes and immune infiltration.** A and B The differences in immune infiltration between *Helicobacter pylori* (*H. pylori*)-infected gastric tissues and normal gastric tissues; C: Correlation analysis between hub gene expression and immune cell infiltration levels in *H. pylori* infection. *H. pylori*: *Helicobacter pylori*.





**Figure 6 The interaction of hub genes with miRNA/transcriptional factors.** A: Interaction network between the hub genes and their targeted miRNAs. Hub genes are presented in red squares, whereas miRNAs are shown in green circles. Orange circles represent miRNAs targeting two or more genes simultaneously; B: Construction of the transcriptional factor-gene interaction network from Cytoscape.



**Figure 7 Results of single-gene gene set enrichment analysis.** A: *Helicobacter pylori* infection; B: Type 2 diabetes mellitus.