OPEN ACCESS

Kashf Journal of Multidisciplinary Research
Vol: 02 - Issue 3 (2025)
P-ISSN: 3007-1992 E-ISSN: 3007-200X
https://kjmr.com.pk

SHARED GENETIC FACTORS IN HELICOBACTER PYLORI-INDUCED GASTRITIS AND GASTRIC CANCER: A NETWORK-DRIVEN ANALYSIS OF GENE EXPRESSION DATA

Abuzar Mehdi Khan
MS Scholar, Department of Zoology, University of Education, Lahore, Faisalabad Campus, Punjab, Pakistan.
Ayesha Iram
MS Scholar, Department of Zoology, University of Education, Lahore, Faisalabad Campus, Punjab, Pakistan.
Muhammad Adnan
MS Scholar, Department of Zoology, University of Education, Lahore, Faisalabad Campus, Punjab, Pakistan.
Muhammad Imran Ul Haq
MS Scholar, Department of Zoology University of Education Lahore, Multan Campus, Punjab, Pakistan.
Waseem Abbas
Center for Applied Molecular Biology, University of the Punjab, Lahore, Punjab, Pakistan.
Muhammad Imran Malik*
MS Scholar, Department of Zoology, University of Education Lahore, Multan Campus, Punjab, Pakistan.
**Corresponding author: Muhammad Imran Malik (imranmalik8182@gmail.com)*

Article Info

Abstract

Two of the most common medical conditions in the world, gastritis and stomach cancer, are caused by the Helicobacter pylori (H. pylori). Gastric cancer and H. pylori-induced gastritis has both been the subject of molecular research, but the latter has received much more interest. In attempt to identify shared genetic factors that contribute to various diseases, this study examines gene expression data from the public microarray data set known as the Gene Expression Omnibus (GEO) database. Three different types of stomach cancer were investigated: GSE54129, GSE65168 and GSE13911. Using GEO2R, we identified genes that were overexpressed in both diseases. In order to get a better understanding of how genes work, we conducted GO and pathway enrichment investigations. In order to identify hub genes and important molecular interactions, a STRING-based PPI network was constructed using Cytoscape. To investigate hub genes, we used CytoHubba and MCODE to look at important chemical complexes and strongly connected nodes. The investigation of gene regulation was conducted using Network Analyst, which allowed us to construct TF-gene and TF-miRNA regulatory networks. Curing gastritis and stomach cancer caused by H. pylori might be possible with the use of a Drug Bank medication interaction network. We may be able to better understand and treat stomach issues caused by H. pylori if we use an integrated bioinformatics approach. It has the ability to identify genetic circuits, biomarkers and therapeutic targets.

Keywords:
Helicobacter pylori, gastritis, gastric cancer, gene expression, GEO, differential gene expression, protein-protein interaction, hub genes, pathway enrichment, transcription factors, microRNA, drug interaction network.

Introduction

H. pylori is a gram negative bacterium and a major cause of the worldwide gastric diseases both as chronic gastritis and gastric cancer (Blaser, 2011). Gastritis and particularly H. pylori induces gastritis, is a chronic inflammatory disease which develops into more severe consequences like formation of gastric ulcers and, only in some cases, gastric cancer (Kuipers et al., 1995). Gastric cancer has a very high incidence and is extremely lethal and its strong association with H. pylori infection is documented (International Agency for Research on Cancer, 1994). Although complex, the molecular mechanisms whereby H. pylori induced gastritis leads to gastric cancer are not known (Correa, 2004). Offspring are also shared when the parents have either of these diseases, this highlighting the need for greater investigation of the common genetic factors of both diseases which may be sources of biomarkers for early diagnosis and therapeutic targets. To understand the shared genetic factors between H. pylori induced gastritis and gastric cancer, expression data from patients afflicted by these conditions (Wang et al., 2020) has to be integrated. In this work we want to understand the common disease sequences between these diseases by analyzing publicly available microarray datasets. We focused specifically on three datasets from Gene Expression Omnibus (GEO) database (Barrett et al., 2013): GSE65168 that was containing the gene expression profiles on peripheral blood cells in individuals with intense gastritis due to severe H. pylori infection and healthy controls, GSE13911 and GSE54129 which were containing gene expression profiles on gastric cancer patients and matched control.

Thus, for the first time, the investigators have a noninvasive diagnostic tool that can be used to investigate gastiritis and stomach cancer induced by H. pylori – quantitative assessment of transcriptional changes in peripheral blood cells. The given gene expression was determined using GSE65168 of GPL26963: Agilent-085982 Arraystar human lncRNA V5. They targeted ten individuals in total comprising of five with severe gastritis due to H. pylori infection and five normal heathy individuals. To analyze the genes differentially expressed in GSE13911 and GSE54129 datasets with twelve patients with stomach cancer and twelve healthy individuals, GPL570 platform of Affymetrix Human Genome U133 plus 2.0 Array was employed. Both datasets are ideal for investigating the systemic reactions to the both disorders because they employ PBMC, however under different conditions. Given the fact that gastritis is a potential precancerous factor that leads to stomach cancer, this study aimed at establishing response pathways that are associated with the two diseases through gene expression.

Candidate DEGs derived from the GEO2R online tool when both conditions were compared included the DEGs identified by Sui and colleagues (2023). With the help of GEOquery and limma application in R, several sample groups are analyzed in this context. Using the criteria which were rather stringent and easily interpretable, all results were filtered based on p-value adjust <0.01 and Log2-fold change < -1. Since, for pathway and network analysis, we sought the genes that are potentially significant for gastritis and stomach cancer, we selected those genes which are overexpressed in these datasets. The genes involved in this study had the highest enrichments based on their biologic processes and functions based on Gene Ontology (GO) and the signaling pathways. Applying the name list of identified genes on this study on the web-based enrichment analysis tool referred to as Enrichr developed by Kuleshov et al. (Khan et al., 2016). In this case, we can use enrichment analysis to examine the pathways that would result in gastritis and stomach cancer due to H. pylori. In this respect, KeGG, Reactome, BioPlanet and MSigDB were chosen as much more relevant (Hong et al., 2014).

Also, when only focusing on the common upregulated genes, the PPI network was conducted as the PPI. This was done on STRING, a database, which provides information on protein interaction in a large number of species. Thus, analysing a complete PPI network map by using Cytoscape, the selection of the hub genes – genes with a high per degree of connectivity and that are most probably crucial for the development of the disease (Hevey, 2018). Cytoscape plugin and Molecular Complex Detection (MCODE) was also applied along with CytoHubba to analyze further of the hub genes enclosed in the PPI net, the complex within PPI network (Chin et al., 2014; Hejzlar and Xu, 2008). We also attempted to analyse the regulation of these hub genes by establishing the published TF–gene interaction networks and

the TF-miRNA regulation networks of the hub genes in NetworkAnalyst. These analyses are useful to have information concerning the regulation the genes for the transcription and post-transcription part of the H. pylori induced gastritis and gastric cancer (Lindemose et al., 2013; Rad et al., 2015). Last, drug interactions were built in DrugBank to develop molecular targeting the hub genes to enable the discovery of new drugs to treat these diseases (Knox et al., 2024).

Material and Methods

2.1. Acquisition of datasets

Microarray datasets on Helicobacter pylori induced gastritis (GSE65168), gastric cancer (GSE13911 and GSE 54129) were obtained from Gene expression Omnibus GEO. (https://www.ncbi.nlm.nih.gov/gds). Both collections include datasets on peripheral blood cells. Five individuals without gastritis and five with Helicobacter pylori infection contributed datasets to GSE65168. A human lncRNA V5 microarray GPL26963 platform, model 085982 from Agilent, was used. For the GSE 13911 and GSE 54129 datasets, 12 samples were obtained from 12 stomach cancer patients and 12 controls using the Affymetrix Human Genome U133 plus 2.0 Array GPL570.

2.2. Identification of common upregulated genes

GEO2R is used for identifying common differentially expressed genes (DEGs) of both datasets https://www.ncbi.nlm.nih.gov/geo/geo2r/ was done. After comparing many sets of data, Sui et al. (2023) used the GEO2R term and the R tools limma and GEOquery, both developed by the Bioconductor Project, to identify DEGs. According to Ferreira and Zwinderman (2006), the Benjamini-Hochberg test was used to regulate the false discovery rate. We used GEO2R to get a table containing the common DEGs from both datasets before continuing our analysis in RStudio. A Log2-fold change of less than -1 and an adjusted p-value of less than 0.01 were the criteria used to pick the datasets.

Bioinformatics and Evolutionary Genomics provided the technique that was used to construct the Venn diagram. http://bioinformatics.psb.ugent.be/webtools/Venn/ webtool followed by getting common upregulated genes.

2.3. Enrichment analysis of common upregulated genes

Gene set enrichment analysis is a way to evaluate the collective behaviors of genes in connection to health and sickness, as stated by Hong, Zhang et al. (2014). Gene ontologies (GO) and associated pathways for commonly increased genes were derived using Enrichr. (https://maayanlab.cloud/Enrichr/).

Khan, Dębski and colleagues (2016) state that Enrichr has several collective gene list features and is an easy-to-use web-based application for enrichment analysis. Metagenome, KEGG, BioPlanet and MSigDB were the databases used for pathways analysis.

2.4. Network analysis

Network analysis is an essential part of system biology for understanding protein interactions at the cellular and molecular levels (Hevey 2018). In addition, rather of looking at individual genes, research based on networks may provide information on gene sets. The network was generated using a plugin tool known as STRING protein quarry Cytoscape. Gene interactions may be found using the STRING database. (https://string-db.org/)is used to provide information on the physical and functional protein-protein interactions involving more than 2000 distinct species. Genes that were upregulated and had a confidence level of 0.400 were entered using the search box. The viewing and alteration processes were carried out using Cytoscape (Smoot, Ono et al., 2011) and its associated plugin tools after the network retrieval.

2.5. Hub genes identification and module analysis

The nodes and edges that make up a network may take numerous forms; the most highly connected ones are known as hub genes. More tightly coupled genes, known as hub genes, are often essential for the maintenance of biological processes. Using cytoHubba, a Cytoscape plugin tool, we were able to identify the hub genes in our PPI networks. An easy-to-use program that provides eleven topological analysis methods for investigating critical nodes in biological networks is CytoHubba, claims Chin, Chen and colleagues (2014). Using the total number of gene interactions in the PPIs network as its foundation, this

research used the degree topological technique. Molecular Complex Detection (MCODE), a new plugin tool in Cytoscape, was used to identify the tightly related parts of the PPIs network. To make things easier to see, MCODE eliminates the dense regions around an interest protein (Xu and Hejzlar 2008).

2.6. Transcriptional factor regulatory network of hub genes

Transcription factor (TF) networks are impaired in many diseases. These networks are essential for adult tissue homeostasis and cell fate decisions in mammals (Lindemose, O'Shea et al. 2013). Using NetworkAnalyst, a hierarchical network of TF-hub gene interactions was constructed. (https://www.networkanalyst.ca/NetworkAnalyst/uploads/ListUploadView.xht). According to Zhou, Soufan, et al. (2019), the comprehensive web platform NetworkAnalyst offers a visual network for analyzing gene expression. We used information from the JASPAR database to build the TFs gene interaction network. (http://jaspar.genereg.net/) which is included in NetworkAnalyst platform.

2.7. TFs-miRNA regulatory network analysis

According to Rad, Langroudi et al. (2015), TFs control transcription before it happens, while microRNAs control gene expression after transcription has already occurred. The TFsmiRNA regulatory network for hub genes was constructed using the RegNetwork repository and the NetworkAnalyst tool (Liu, Wu et al. 2015). One degree was the threshold parameter used to filter the network. The last step was to bring the network into Cytoscape after importing it from NetworkAnalyst.

2.8. Protein drug interaction network

As a last step, we constructed a drug interaction network using our hub genes to get insight into potential treatments for Helicobacter pylori gastritis and stomach cancer. The network was built using DrugBank as its foundation. (https://go.drugbank.com/) Database using the web interface of NetworkAnalyst (Knox, Wilson et al., 2024). After downloading the network file from NetworkAnalyst, we saw it in Cytoscape.

Results

3.1. Identification of common upregulated genes between Helicobacter pylori-induced gastritis and Gastric Cancer:

Selected datasets (GSE65168) and (GSE13911 and GSE54129) were analyzed to sort out common upregulated genes in between Helicobacter pylori-induced gastritis and Gastric Cancer patients. 22 common upregulated genes were obtained from these datasets. The analysis included the following genes: TAGLN, COL4A2, MMP7, NFIA, DGKH, TMEM173, THBS1, ADAMTS2, EGR1, COL4A1, VCAN, TP53, SLC39A10, MYL9, SLIT2, TRIM59, TRPS1, RAB31, CTGF, COL12A1, CENPF, ARHGDIB. Figure 1 shows a Venn diagram that compares genes that are often upregulated.

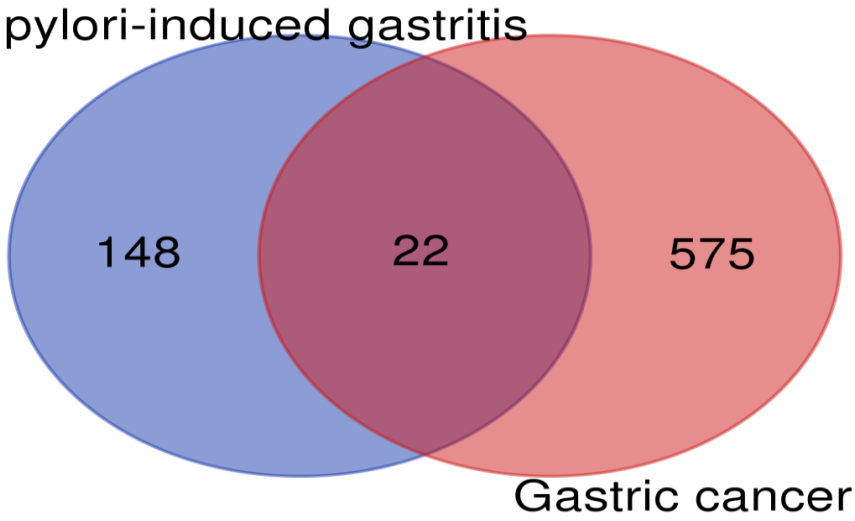

Figure 1: Variably expressed genes that are at a higher level often Out of 575 genes linked to stomach cancer and 148 genes linked to gastritis caused by Helicobacter pylori, 22 were shown to be overactive..

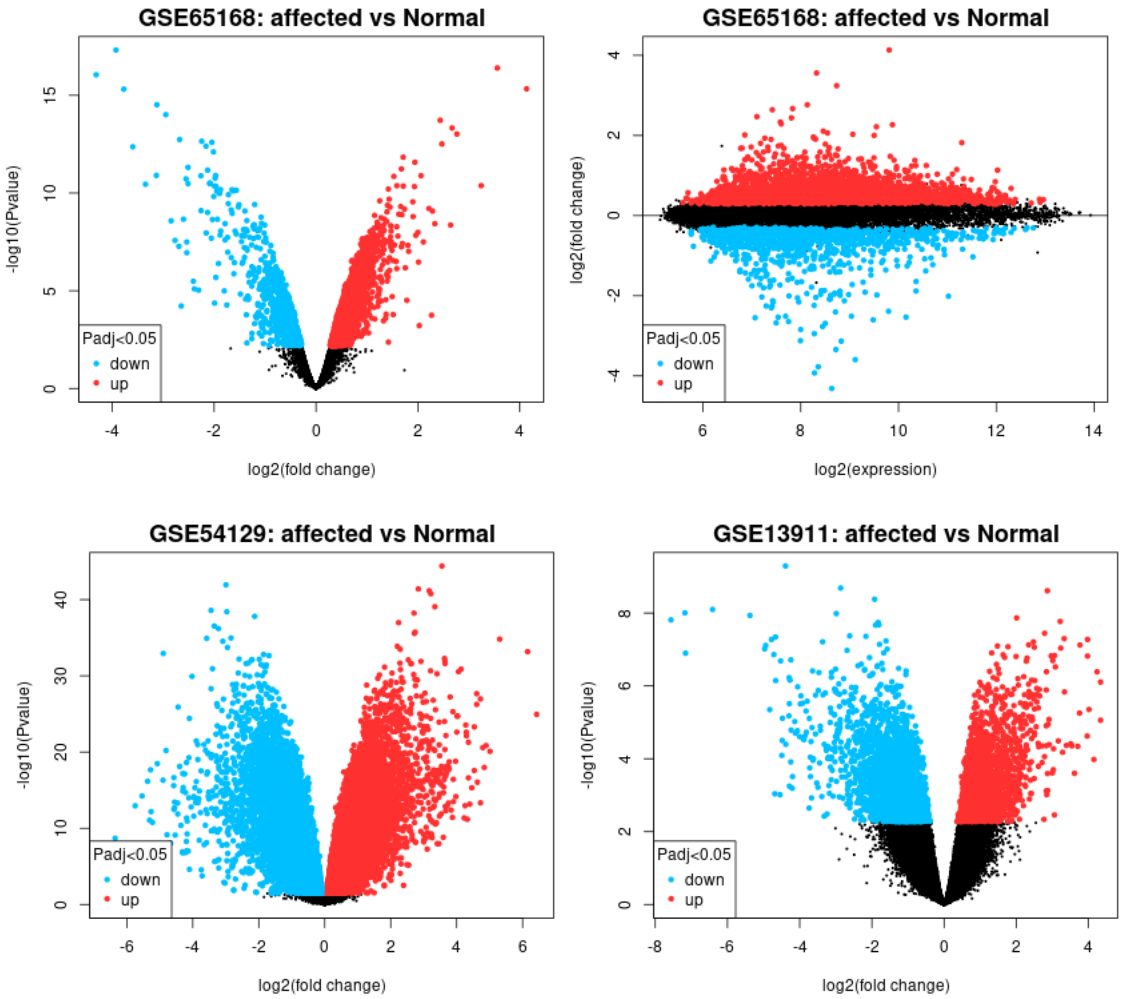

Figure 2. Find the most active and stable differentially expressed genes (DEGs), A volcanic map showing the location of 221 differentially expressed genes (DEGs) in separate studies: one for gastritis caused by Helicobacter pylori and another for gastric cancer in GSE65168, GSE13911 and GSE54129. The 223 genes that showed an increase or decrease in expression are shown by the green and red dots, respectively, while the genes that showed no change are represented by the black dots.

3.2. Enrichment analysis of common upregulated genes

For upregulated genes, the Enrichr web platform made GO and route analysis possible. The cellular component, biological process and molecular function GO subcategories all have big ontologies with logfc values ranging from four to six, as shown in Table 1. The molecular function also relied on the action of transmembrane transporters, beta-galactoside (CMP) alpha-2,3-sialyltransferase and related activities. The t-cell receptor complex and zonula adherens may also be present in the cells.

ID	Gene.title	P.Value	Gene.symbol	logFC
223122_s_at	secreted frizzled related protein 2	1.15E-25	SFRP2	6.424615
227404_s_at	early growth response 1	6.78E-34	EGR1	6.151059
238320_at	microRNA 612///nuclear paraspeckle assembly transcript 1 (non-protein coding)	1.55E-35	MIR612///NEAT1	5.305683
223121_s_at	secreted frizzled related protein 2	8.02E-21	SFRP2	5.012584

201058_s_at	myosin light chain 9	1.38E-21	MYL9	4.892884
218468_s_at	gremlin 1, DAN family BMP antagonist	1.05E-18	GREM1	4.83707
226237_at	collagen type VIII alpha 1 chain	2.66E-21	COL8A1	4.791566
209395_at	chitinase 3 like 1	4.59E-14	CHI3L1	4.723743
202310_s_at	collagen type I alpha 1 chain	1.08E-27	COL1A1	4.722978
224646_x_at	microRNA 675///H19, imprinted maternally expressed transcript (non-protein coding)	9.08E-17	MIR675///H19	4.628228
1555229_a_at	complement component 1, s subcomponent	5.63E-27	C1S	4.608685
209156_s_at	collagen type VI alpha 2 chain	2.20E-28	COL6A2	4.608521
201289_at	cysteine rich angiogenic inducer 61	5.02E-24	CYR61	4.536129
227140_at	inhibin beta A subunit	2.65E-18	INHBA	4.536039
204051_s_at	secreted frizzled related protein 4	4.83E-16	SFRP4	4.460651

3.3. Network analysis

We obtained the matched network using Cytoscape's STRING protein quarry plugin function. In order to find hub genes and, subsequently, to suggest common pharmaceutical compounds for gastritis induced by Helicobacter pylori and stomach cancer, we preserved this network. The primary objectives of this study were to provide therapy options for gastritis and stomach cancer caused by Helicobacter pylori and to find common genetic factors that contribute to these illnesses. In Figure 3, we can see that the network consisted of 33 nodes and 74 edges.

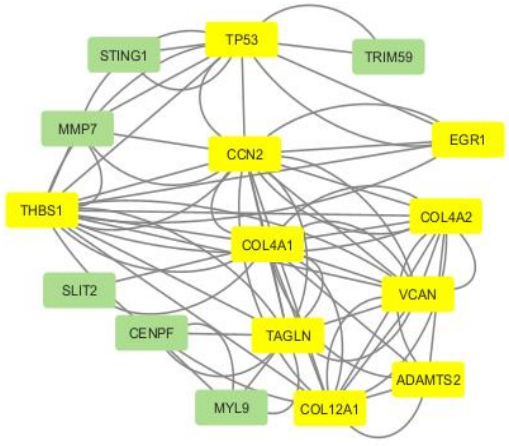

Figure 3. Many protein-protein interactions (PPIs) connect the genes that are active in gastritis and gastric cancer caused by Helicobacter pylori. Genes that are often upregulated are shown by nodes that are yellow in color. The network has 33 nodes and 74 edges.

3.4. Hub genes identification and module analysis

Because of their close proximity to one another, the genes that make up the PPIs network were partitioned using CytoHubba. The hub genes were located using the Degree approach. Ten genes were found to be hub genes: TAGLN, TP53, EGR1, THBS1, COL4A2, COL12A1, ADAMTS2, COL4A1, VCAN and

CCN2. This information is shown in Figure 4. A densely populated area of PPIs networks was, however, identified by use of MCODE. Two hub genes, THBS1 and CCN2, are shown in Figure 4 as being part of this clustering network.

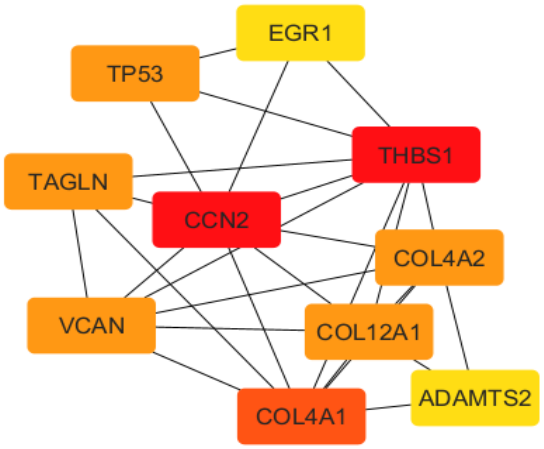

Figure 4. To create the cluster analysis network, the PPIs network was used. An extensively connected web represents the PPIs network here. Two genes, CCN2 and THBS1, were produced from a collection of ten hub genes by the cluster network. Two red-hued genes serve as hubs.

3.5. Transcriptional factor regulatory network of hub genes

A transcription factor regulatory network for hub genes was constructed using the NetworkAnalyst platform. There are six hub genes, seventy-two transcription factors and 78 connections in this network. Within the TP53 regulatory network, 48 transcription factors were located: 13 EGR1, 5 THBS1, 1 COL4A2, 2 VCAN and 3 TAGLN. Six TFs were found in the TF regulatory network with a degree of 2 or higher of connectivity.

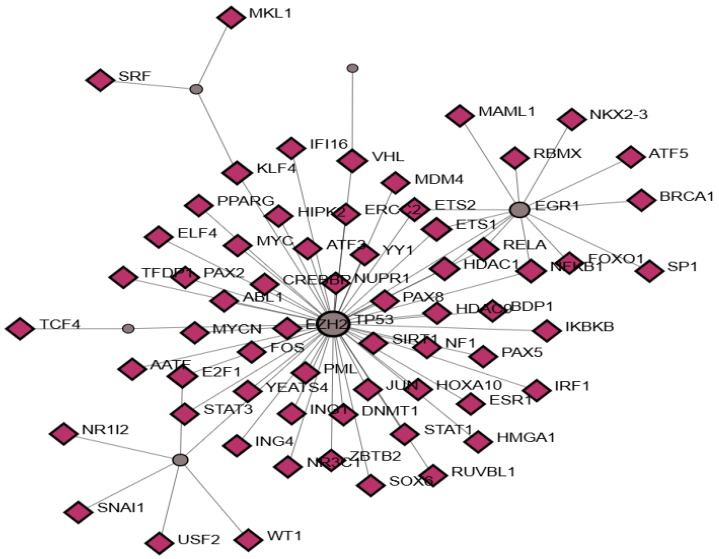

Figure 5. Hub genes and transcription factors interact via a network. The gray nodes represent hub genes, whereas the other colored nodes denote TF genes. The network is comprised of twelve core components and six hub genes.

3.6. TFs-miRNA regulatory network analysis

Based on the data supplied by the TFs-miRNA regulatory Network, we may deduce how hub genes interact with miRNAs and TFs. Using NetworkAnalyst, a TF-miRNA coregulatory network was examined. The network had 4842 nodes and 5263 edges.

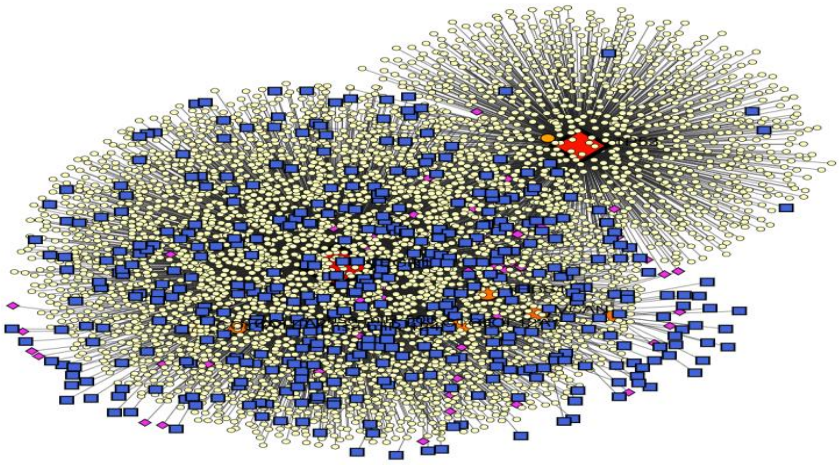

Figure 6. Common hub genes' TF-miRNA coregulatory network is shown by the network.

3.7. Protein drug interaction network

In order to offer patients with the right treatment, protein drug interaction networks provide crucial information. We built a protein-drug interaction network using DrugBank data and found fourteen different medicines substantially related with VDR. This network suggests that certain of its related compounds may play critical roles in altering Helicobacter pylori gastritis and gastric cancer. The drug pipeline is shown in Figure 7.

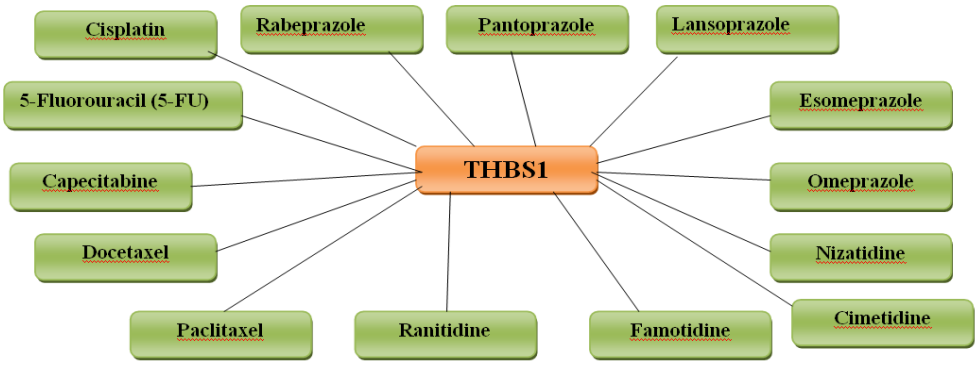

Figure 7. Drugs protein interaction network.

Discussion

The results of this research shed light on the common elevated genes and pathways that facilitate the progression of gastritis to gastric cancer as a result of H. pylori infection. In both illnesses and their molecular bases, we discovered differentially expressed biological processes and molecular pathways by integrating multiple publicly accessible expression datasets. Enriched analysis reveals pathways involved in gastritis and stomach cancer-related processes, including inflammation, immunological response and cell proliferation. Additional instances of hub genes having a significant impact on disease development and being potential biomarkers or treatment targets include proteins discovered using PI network analysis. These findings, when taken in conjunction with the miRNA function in untreated cells and the transcription factor networks, point to a region of the molecular landscape where gene regulation is cellularly tightly controlled. Additionally, our drug interaction network allows us to look for future studies that focus on hub genes, which might be used as therapeutic agents to design medications that target those genes. The majority of stomach cancer cases may be caused by molecular alterations in H. pylori-induced gastritis, according to this study's findings. In order to better understand the processes behind these

Disorders and provide insight into therapeutic intervention targets, our bioinformatics method was able to integrate at the genomic level.

References

Barrett, T., Wilhite, S. E., Ledoux, P., et al. (2013). NCBI GEO: Archive for functional genomics data sets—10 years on. Nucleic Acids Research, 41(Database issue), D991–D995.

Blaser, M. J. (2011). Helicobacter pylori and gastric cancer. The Journal of the National Cancer Institute, 103(5), 292–295.

Chin, C.-H., et al. (2014). "cytoHubba: identifying hub objects and sub-networks from complex interactome." BMC Systems Biology, 8: 1-7.

Correa, P. (2004). Helicobacter pylori and gastric carcinogenesis. The American Journal of Surgical Pathology, 28(3), 237-249.

Ferreira, J. and A. Zwinderman (2006). "On the Benjamini–Hochberg method."

Hong, G., et al. (2014). "Separate enrichment analysis of pathways for up-and downregulated genes." Journal of the Royal Society Interface, 11(92): 20130950.

International Agency for Research on Cancer (1994). Helicobacter pylori. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, 61, 177-240.

Khan, A., Dębski, K., et al. (2016). Enrichr: A comprehensive tool for functional genomics analysis. Bioinformatics, 32(3), 464-465. https://doi.org/10.1093/bioinformatics/btv390

Knox, C., et al. (2024). "DrugBank 6.0: the DrugBank knowledgebase for 2024." Nucleic Acids Research, 52(D1): D1265-D1275.

Kuipers, E. J., Uyterlinde, A., Peña, A. S., et al. (1995). Long-term sequelae of Helicobacter pylori gastritis. Lancet, 345(8943), 1525–1528.

Lindemose, S., et al. (2013). "Structure, function and networks of transcription factors involved in abiotic stress responses." International Journal of Molecular Sciences, 14(3): 5842-5878.

Liu, Z.-P., et al. (2015). "RegNetwork: an integrated database of transcriptional and posttranscriptional regulatory networks in human and mouse." Database, 2015: bav095.

Smoot, M. E., et al. (2011). "Cytoscape 2.8: new features for data integration and network visualization." Bioinformatics, 27(3), 431-432.

Sui, Y., et al. (2023). "Bioinformatics analyses of combined databases identify shared differentially expressed genes in cancer and autoimmune disease." Journal of Translational Medicine, 21(1): 109.

Wang, X., et al. (2020). Gene expression profiling of Helicobacter pylori-infected patients and gastric cancer. Journal of Clinical Investigation, 130(8), 4245-4257.

Xu, Z. and P. Hejzlar (2008). MCODE, Version 2.2: an MCNP-ORIGEN depletion program, Massachusetts Institute of Technology. Center for Advanced Nuclear Energy.

Zhou, G., et al. (2019). "NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis." Nucleic Acids Research, 47(W1): W234-W241.