Journal of Science Research International (JSRI)
ISSN: 2456-6365 (Online)

OPEN ACCESS

# IDENTIFICATION OF DISEASE-ASSOCIATED GENES AND FUNCTIONAL PATHWAYS IN CORONARY HEART DISEASE: A BIOINFORMATICS APPROACH

**Suprita Rao* and Shruti Salve**

*Department of Biotechnology,*

*Pillai College of Arts, Commerce & Science (Empowered Autonomous),*

*New Panvel, Navi Mumbai, Maharashtra*

**Corresponding author E-mail: supritarao@mes.ac.in*

| Received: 20 December 2025 | Revised: 22 January 2026 | Accepted: 20 February 2026 | Published: 28 February 2026 |
|---|---|---|---|

***DOI:*** **https://doi.org/10.5281/zenodo.18899485**

**Abstract:**

*Coronary Heart Disease (CHD) is one of the leading causes of morbidity and mortality worldwide and results from the narrowing or blockage of coronary arteries due to atherosclerosis. This condition reduces blood flow and oxygen supply to the myocardium, leading to clinical manifestations such as angina pectoris, myocardial infarction, heart failure, and sudden cardiac death. Major risk factors associated with CHD include hypertension, hyperlipidemia, smoking, diabetes mellitus, obesity, sedentary lifestyle, and genetic predisposition. Pathophysiologically, CHD involves endothelial dysfunction, lipid accumulation, inflammatory responses, and plaque formation within arterial walls. Early diagnosis through electrocardiography, cardiac biomarkers, imaging techniques, and stress testing plays a crucial role in disease management. Treatment strategies focus on lifestyle modification, pharmacological therapy such as antiplatelet agents and statins, and interventional procedures including angioplasty and coronary artery bypass grafting. Preventive measures emphasizing healthy diet, regular physical activity, and risk factor control are essential in reducing the burden of coronary heart disease and improving patient outcomes.*

**Keywords:** *Coronary Heart Disease (CHD), Genes, Functional Pathway, Bioinformatics.*

## Introduction

Coronary Heart Disease (CHD), also known as coronary artery disease (CAD), is a chronic and progressive cardiovascular disorder resulting from impaired blood flow to the myocardium due to structural and functional abnormalities of the coronary arteries. The predominant pathological process underlying CHD is atherosclerosis,

which involves the deposition of lipid-rich plaques within the arterial walls, leading to luminal narrowing, reduced elasticity, and compromised coronary perfusion (1).

***Pathophysiology***

The pathogenesis of CHD begins with endothelial dysfunction, often triggered by risk factors such as hypertension, smoking, hyperglycemia, and dyslipidemia. Endothelial injury increases vascular permeability, allowing low-density lipoprotein (LDL) cholesterol to infiltrate the intima. Oxidized LDL initiates an inflammatory response, recruiting monocytes that differentiate into macrophages and form foam cells, resulting in fatty streaks—the earliest visible lesion of atherosclerosis (2). Over time, smooth muscle cell proliferation and extracellular matrix deposition contribute to plaque maturation. Plaque rupture and subsequent thrombus formation can acutely obstruct blood flow, leading to myocardial infarction.

***Epidemiology and global burden***

CHD is the leading cause of death globally and represents a significant public health challenge. According to the World Health Organization (3), cardiovascular diseases account for approximately 32% of all global deaths, with CHD being the major contributor. The prevalence of CHD has increased markedly in low- and middle-income countries due to urbanization, dietary transitions, reduced physical activity, and increased incidence of diabetes and obesity (4).

***Risk factors***

Risk factors for CHD are broadly classified into modifiable and non-modifiable categories. Modifiable risk factors include hypertension, dyslipidemia, smoking, diabetes mellitus, obesity, sedentary lifestyle, unhealthy diet, and excessive alcohol consumption. Non-modifiable factors include age, sex, ethnicity, and genetic predisposition. Psychosocial stress and socioeconomic factors have also been recognized as important contributors to disease development (5).

***Clinical manifestations***

The clinical presentation of CHD varies depending on the severity and extent of arterial involvement. Common manifestations include stable angina, unstable angina, myocardial infarction, heart failure, and sudden cardiac death. Some individuals may remain asymptomatic for years, highlighting the importance of early screening and risk assessment.

***Diagnosis***

Diagnosis of CHD involves a combination of clinical evaluation, laboratory investigations, and imaging techniques. Electrocardiography (ECG), cardiac biomarkers such as troponins, stress testing, echocardiography, coronary computed tomography angiography (CCTA), and invasive coronary angiography are commonly used to assess myocardial ischemia and coronary anatomy (6).

***Management and prevention***

Management strategies for CHD aim to relieve symptoms, prevent disease progression, and reduce mortality. Therapeutic approaches include lifestyle modifications such as smoking cessation, dietary changes, regular physical activity, and weight management. Pharmacological therapy includes antiplatelet agents, statins, antihypertensive drugs, beta-blockers, and nitrates. In advanced cases, revascularization procedures such as percutaneous coronary intervention (PCI) and coronary artery bypass grafting (CABG) are employed. Preventive strategies focusing on population-based risk reduction are crucial in controlling the growing burden of CHD (7).

**Materials and Methods**

***A .DisGeNET***

DisGeNET is a comprehensive discovery platform that integrates information on gene–disease associations (GDAs) from multiple sources, including curated databases, genome-wide association studies (GWAS), animal models, and scientific literature. It provides standardized data on the genetic basis of human diseases and supports translational research by linking genes to clinical phenotypes. DisGeNET allows users to explore disease-associated genes, variants, and their supporting evidence through scores based on the number and quality of publications. It is widely used to identify candidate genes involved in complex and inherited disorders.

***B. STRING (Search Tool for the Retrieval of Interacting Genes/Proteins)***

STRING is an online database designed to predict and visualize protein–protein interaction (PPI) networks. It integrates interaction data from experimental studies, curated pathway databases, co-expression analyses, text mining, and computational predictions. STRING enables users to construct interaction networks for specific organisms and perform functional enrichment analyses, including Gene Ontology (GO) terms and pathway associations. The database helps in understanding the functional relationships between proteins and identifying biologically meaningful interaction clusters.

***C. GeneMANIA***

GeneMANIA is a flexible gene function prediction tool that identifies functionally similar genes based on a given gene list. It uses a wide range of data sources such as co-expression, physical interactions, genetic interactions, pathways, co-localization, and protein domain similarity. GeneMANIA expands an input gene list by adding related genes and assigns weights to different data networks to optimize prediction accuracy. This tool is particularly useful for hypothesis generation, gene prioritization, and exploring gene function in disease-related studies.

***D. Enrichr***

Enrichr is an interactive web-based tool used for gene set enrichment analysis. It allows users to upload gene lists and analyze them across multiple libraries, including Gene Ontology, KEGG, Reactome, WikiPathways, and disease databases. Enrichr provides enrichment scores, p-values, and visualizations such as bar plots and clustergrams to interpret biological significance. It is widely used to identify overrepresented biological processes, molecular functions, cellular components, and signaling pathways associated with a gene set.

***Reactome pathway database***

Reactome is a curated, peer-reviewed pathway database that provides detailed information on biological pathways and molecular reactions in humans. It covers key processes such as metabolism, signal transduction, DNA repair, cell cycle, and apoptosis. Reactome allows pathway enrichment analysis and interactive pathway visualization, enabling researchers to map genes or proteins onto well-defined biological pathways. The database is regularly updated and serves as a reliable reference for pathway-based interpretation of genomic and proteomic data.

**Results**

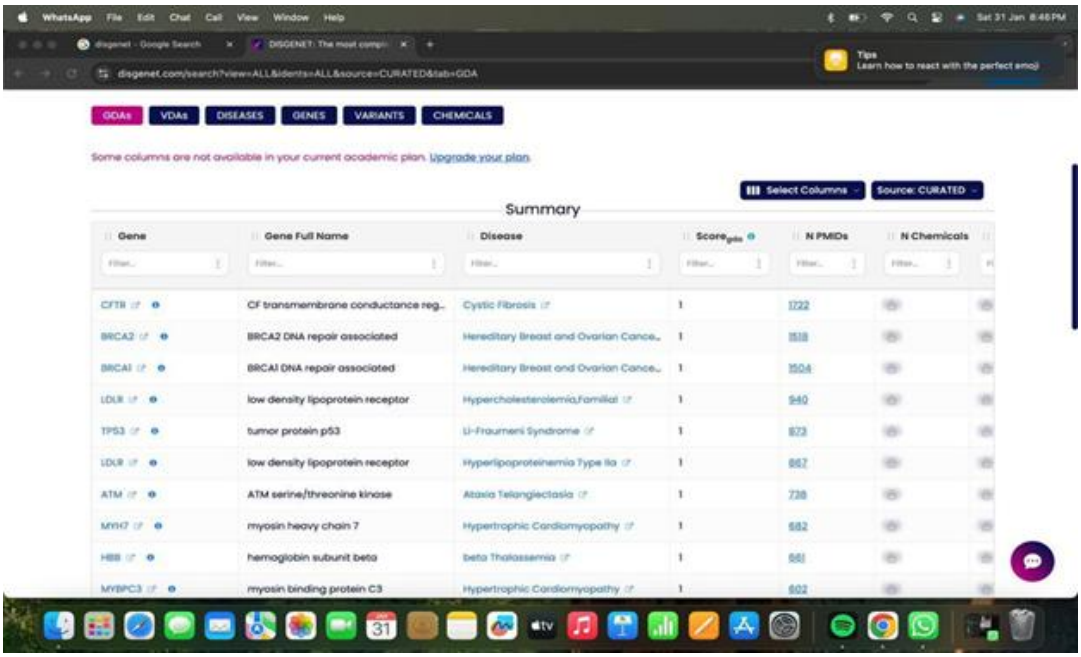
***A. DisGeNET***

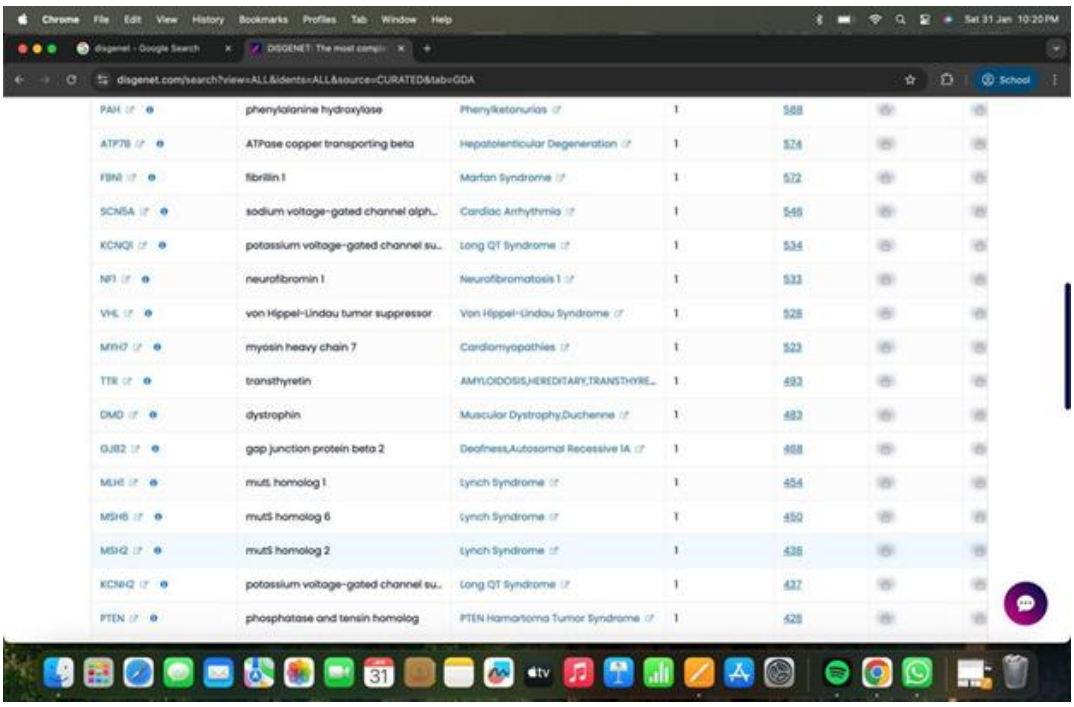
DisGeNET analysis was used to identify gene–disease associations from curated and literature-supported sources. The analysis revealed several high-confidence disease-associated genes, including CFTR, BRCA1, BRCA2, TP53,

LDLR, ATM, MYH7, MYBPC3, and HBB. These genes showed strong association scores supported by multiple PubMed references.

Genes such as LDLR were linked to familial hypercholesterolemia and lipid metabolism disorders, which are major risk factors for coronary heart disease. MYH7 and MYBPC3 were associated with hypertrophic cardiomyopathy, indicating their importance in cardiac muscle structure and function. DNA repair genes like BRCA1, BRCA2, TP53, and ATM were mainly associated with inherited cancer syndromes, reflecting their role in maintaining genomic integrity.

***Outcomes***



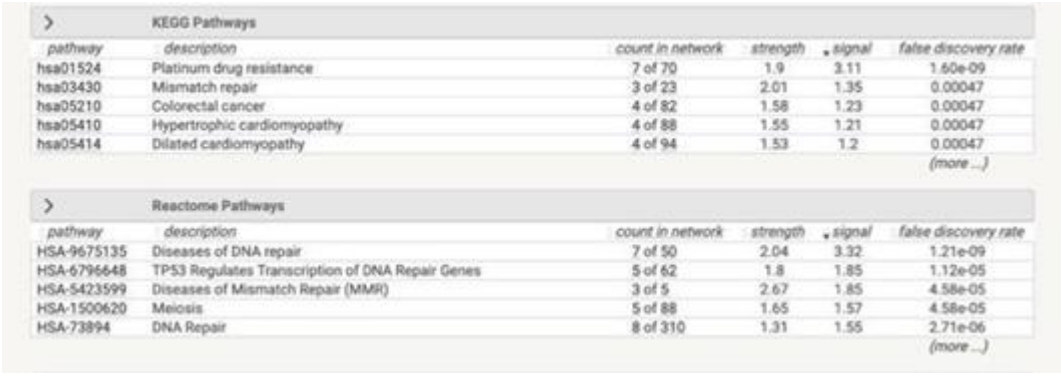



***B. STRING***

STRING analysis generated a protein–protein interaction (PPI) network to study functional relationships among the selected genes. The network consisted of 25 nodes and 86 edges, with an average node degree of 6.88 and a local clustering coefficient of 0.643. The PPI enrichment p-value ($< 1.0 \times 10^{-16}$) indicated that the interactions were highly significant and biologically meaningful.

The network revealed several hub proteins, including BRCA1, BRCA2, TP53, ATM, MLH1, MSH2, and MSH6, forming a core cluster related to DNA damage response and repair. Peripheral clusters involved genes related to cardiac muscle contraction and metabolic regulation, showing cross-talk between genomic maintenance and cardiovascular pathways.

***Outcomes***



KEGG Pathways

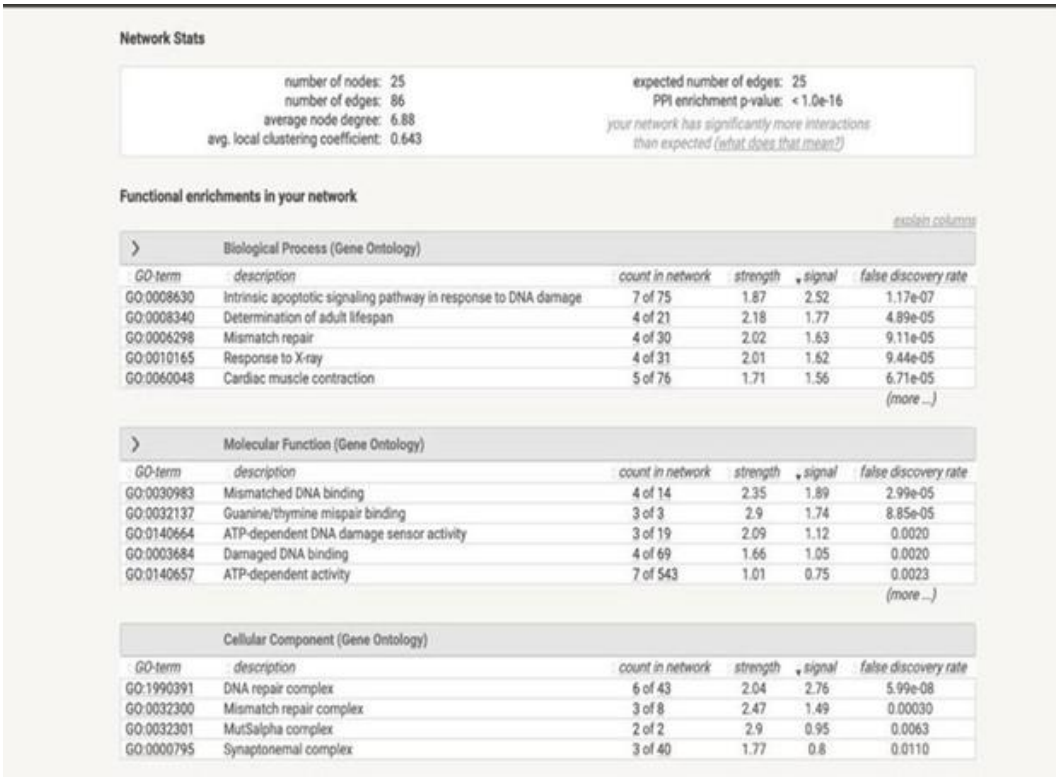
| pathway | description | count in network | strength | signal | false discovery rate |
|---|---|---|---|---|---|
| hsa01524 | Platinum drug resistance | 7 of 70 | 1.9 | 3.11 | 1.60e-09 |
| hsa03430 | Mismatch repair | 3 of 23 | 2.01 | 1.35 | 0.00047 |
| hsa05210 | Colorectal cancer | 4 of 82 | 1.58 | 1.23 | 0.00047 |
| hsa05410 | Hypertrophic cardiomyopathy | 4 of 88 | 1.55 | 1.21 | 0.00047 |
| hsa05414 | Dilated cardiomyopathy | 4 of 94 | 1.53 | 1.2 | 0.00047 |

(more ...)

Reactome Pathways

| pathway | description | count in network | strength | signal | false discovery rate |
|---|---|---|---|---|---|
| HSA-9675135 | Diseases of DNA repair | 7 of 50 | 2.04 | 3.32 | 1.21e-09 |
| HSA-6796648 | TP53 Regulates Transcription of DNA Repair Genes | 5 of 62 | 1.8 | 1.85 | 1.12e-05 |
| HSA-5423599 | Diseases of Mismatch Repair (MMR) | 3 of 5 | 2.67 | 1.85 | 4.58e-05 |
| HSA-1500620 | Meiosis | 5 of 88 | 1.65 | 1.57 | 4.58e-05 |
| HSA-73894 | DNA Repair | 8 of 310 | 1.31 | 1.55 | 2.71e-06 |

(more ...)



Network Stats

number of nodes: 25
number of edges: 86
average node degree: 6.88
avg. local clustering coefficient: 0.643

expected number of edges: 25
PPI enrichment p-value: < 1.0e-16
your network has significantly more interactions than expected (what does that mean?)

Functional enrichments in your network

explain columns

Biological Process (Gene Ontology)

| GO-term | description | count in network | strength | signal | false discovery rate |
|---|---|---|---|---|---|
| GO:0008630 | Intrinsic apoptotic signaling pathway in response to DNA damage | 7 of 75 | 1.87 | 2.52 | 1.17e-07 |
| GO:0008340 | Determination of adult lifespan | 4 of 21 | 2.18 | 1.77 | 4.89e-05 |
| GO:0006298 | Mismatch repair | 4 of 30 | 2.02 | 1.63 | 9.11e-05 |
| GO:0010165 | Response to X-ray | 4 of 31 | 2.01 | 1.62 | 9.44e-05 |
| GO:0060048 | Cardiac muscle contraction | 5 of 76 | 1.71 | 1.56 | 6.71e-05 |

(more ...)

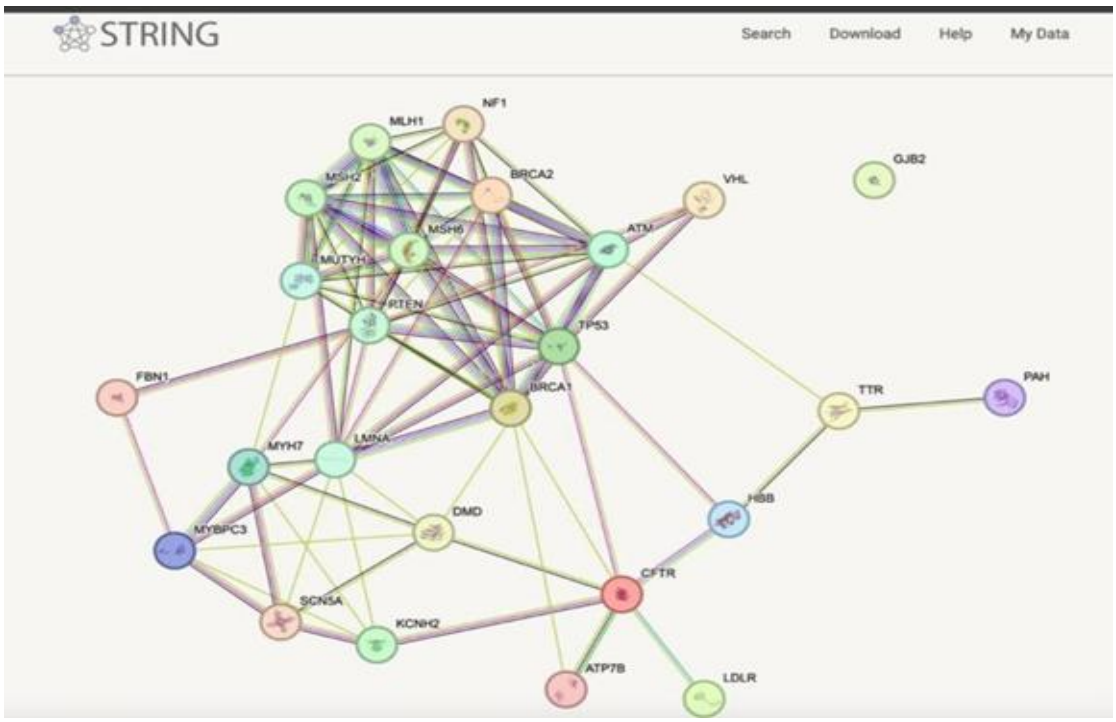
Molecular Function (Gene Ontology)

| GO-term | description | count in network | strength | signal | false discovery rate |
|---|---|---|---|---|---|
| GO:0030983 | Mismatched DNA binding | 4 of 14 | 2.35 | 1.89 | 2.99e-05 |
| GO:0032137 | Guanine/thymine mispair binding | 3 of 3 | 2.9 | 1.74 | 8.85e-05 |
| GO:0140664 | ATP-dependent DNA damage sensor activity | 3 of 19 | 2.09 | 1.12 | 0.0020 |
| GO:0003684 | Damaged DNA binding | 4 of 69 | 1.66 | 1.05 | 0.0020 |
| GO:0140657 | ATP-dependent activity | 7 of 543 | 1.01 | 0.75 | 0.0023 |

(more ...)

Cellular Component (Gene Ontology)

| GO-term | description | count in network | strength | signal | false discovery rate |
|---|---|---|---|---|---|
| GO:1990391 | DNA repair complex | 6 of 43 | 2.04 | 2.76 | 5.99e-08 |
| GO:0032300 | Mismatch repair complex | 3 of 8 | 2.47 | 1.49 | 0.00030 |
| GO:0032301 | MutSalpha complex | 2 of 2 | 2.9 | 0.95 | 0.0063 |
| GO:0000795 | Synaptonemal complex | 3 of 40 | 1.77 | 0.8 | 0.0110 |

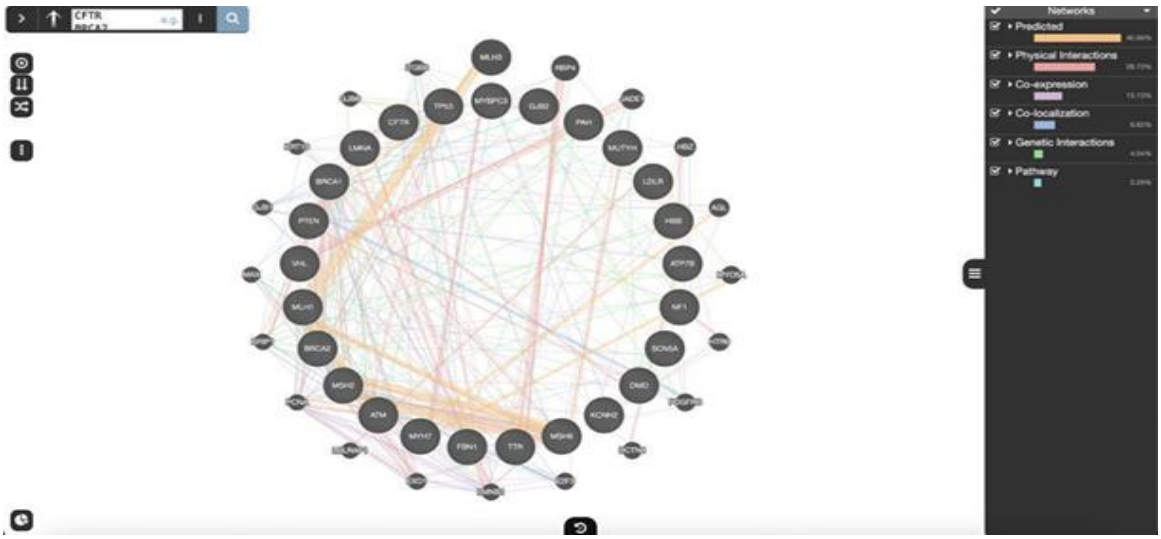


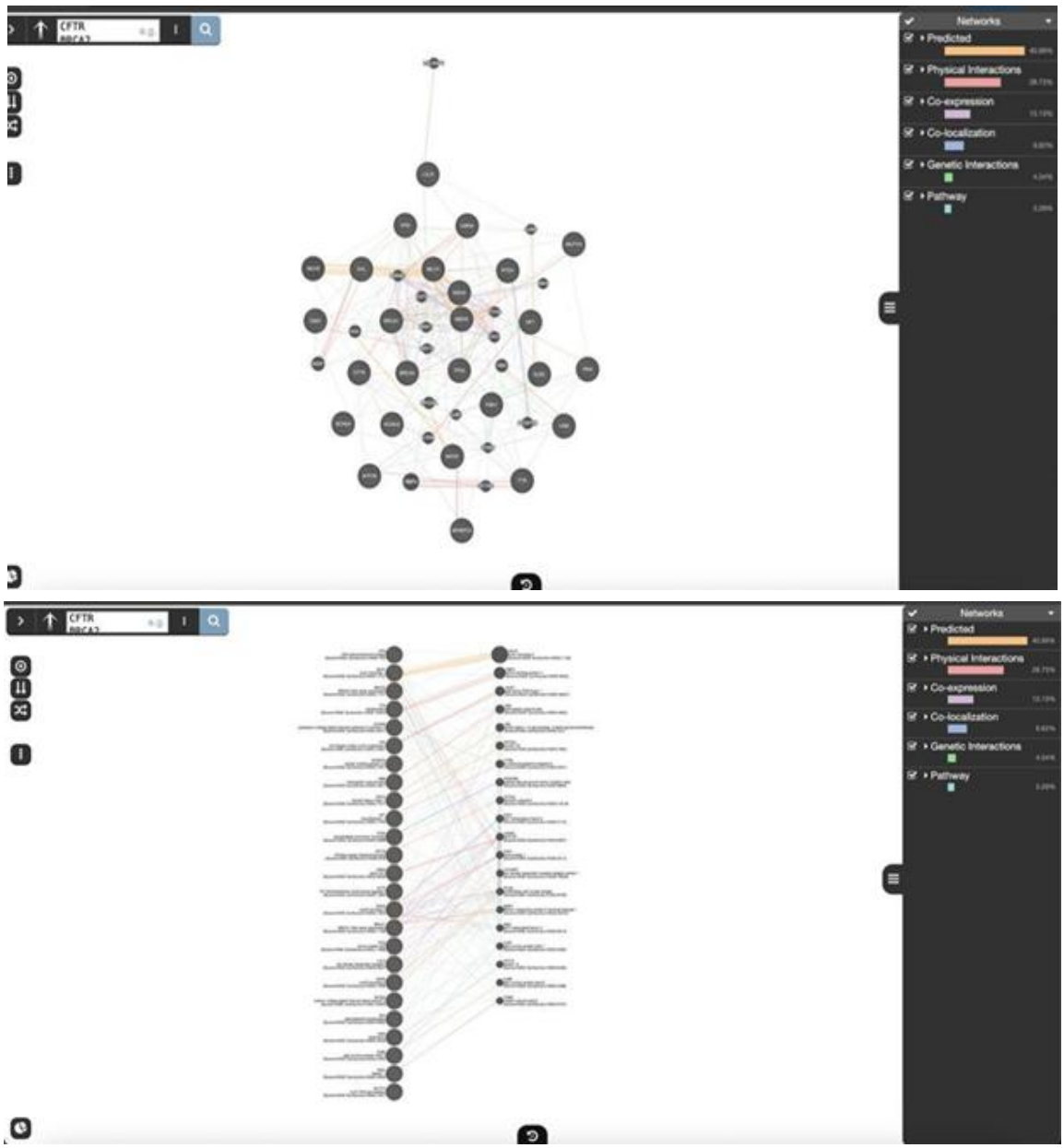

***C. GeneMANIA***

GeneMANIA analysis was performed to predict functionally related genes and expand the input gene list. The tool integrated multiple data sources such as co-expression, physical interactions, pathways, and shared protein domains. GeneMANIA revealed strong co-expression and interaction networks linking DNA repair genes with genes involved in cell cycle control, apoptosis, and cardiac function.

The analysis helped identify additional candidate genes that may play supportive or regulatory roles in disease mechanisms. This functional expansion strengthened the biological relevance of the gene set and aided in gene prioritization for further studies.

***Outcomes***



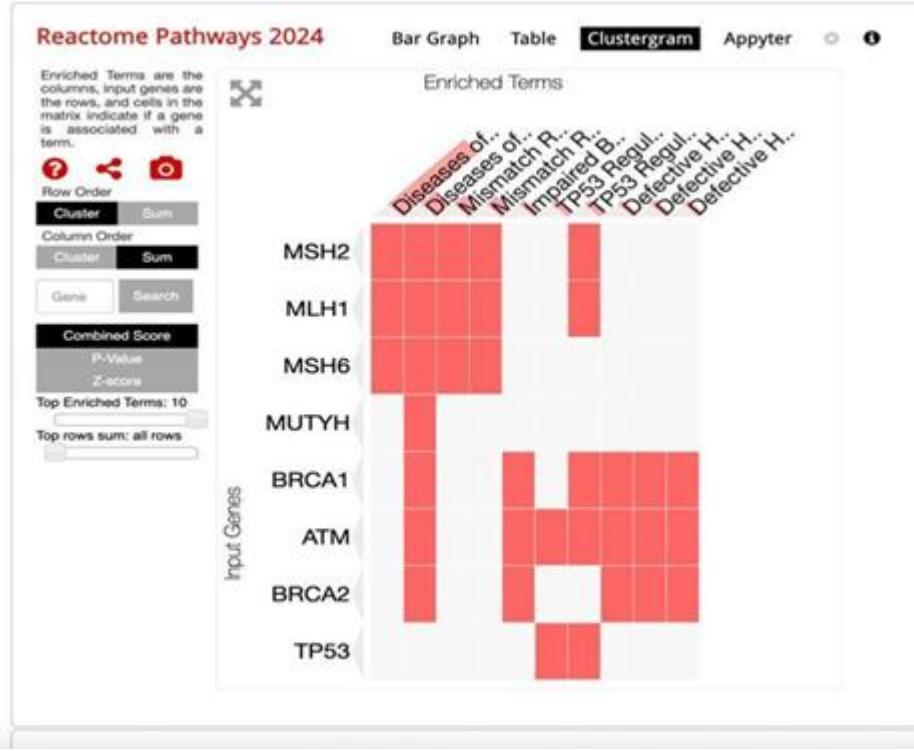



***D. Enrichr***

Enrichr gene set enrichment analysis was used to identify overrepresented biological terms and pathways associated with the selected genes. Significant enrichment was observed across multiple libraries, including Gene Ontology, KEGG, Reactome, and WikiPathways.

Enriched biological processes included DNA damage response, mismatch repair, apoptosis, and cardiac muscle contraction. Molecular functions such as mismatched DNA binding and damaged DNA binding were also prominent. These results indicate that the gene set is functionally involved in both genomic stability and cardiovascular physiology.

***Clutogram output***




**E. Reactome pathway analysis**

Reactome pathway analysis mapped the genes onto curated human biological pathways. The analysis showed significant involvement in pathways related to DNA repair, p53 signaling, cell cycle regulation, and muscle contraction. Reactome visualization highlighted how multiple genes converge within the same pathways, indicating coordinated regulation of cellular processes.

These pathways provide mechanistic insight into how genetic alterations may contribute to disease development, including cardiovascular and systemic disorders.

**Discussion**

The present study integrated multiple bioinformatics tools to explore the genetic and molecular basis of disease, with a focus on coronary heart disease (CHD). CHD is a complex cardiovascular disorder caused by the narrowing or blockage of coronary arteries, usually due to atherosclerosis, and is influenced by both genetic and environmental factors (7). Understanding the molecular networks and functional pathways involved in CHD is essential for identifying potential biomarkers and therapeutic targets.

***1. DisGeNET: Gene–disease associations***

DisGeNET allowed the identification of disease-associated genes with strong literature support. Genes such as LDLR, MYH7, MYBPC3, TP53, and ATM were highlighted for their relevance to cardiovascular health. LDLR, associated with familial hypercholesterolemia, is directly implicated in lipid metabolism, a key risk factor for CHD. Similarly, MYH7 and MYBPC3 play critical roles in cardiac muscle structure, and mutations in these genes can lead to cardiomyopathies, which increase susceptibility to heart failure and CHD complications.

The inclusion of DNA repair genes such as TP53, ATM, BRCA1, and BRCA2 suggests that genomic instability may indirectly influence CHD by affecting vascular cell survival, apoptosis, and response to oxidative stress, which are known contributors to atherosclerosis (1). DisGeNET thus provided a genetically validated list of genes to be further analyzed for functional interactions and pathway involvement.

***2. STRING: Protein–Protein interaction networks***

STRING analysis revealed that these genes are highly interconnected at the protein level, forming dense clusters with significant PPI enrichment ($p < 1.0 \times 10^{-16}$). Core hubs such as TP53, BRCA1, BRCA2, ATM, MLH1, MSH2, and MSH6 are involved in DNA damage response and repair, while peripheral clusters involve cardiac structural proteins (MYH7, MYBPC3) and metabolic regulators (LDLR).

This network demonstrates that CHD is not driven by isolated genes but rather by complex interactions between DNA repair, apoptosis, lipid metabolism, and cardiac muscle function. Proteins involved in genome maintenance may influence endothelial cell survival, oxidative stress response, and plaque stability, indirectly contributing to CHD pathogenesis.

***3. GeneMANIA: Functional gene prediction***

GeneMANIA extended the gene list by identifying functionally related genes based on co-expression, physical interactions, and shared pathways. The predicted network emphasized cell cycle regulation, apoptosis, and cardiac function, providing additional candidate genes that may participate in CHD-related molecular pathways. By expanding the network beyond the initial gene list, GeneMANIA highlighted novel functional connections that could explain the multi-factorial nature of CHD and suggest new avenues for experimental validation.

***4. Enrichr: Functional enrichment analysis***

Enrichr performed gene set enrichment analysis across multiple databases (GO, KEGG, Reactome). The results showed significant enrichment of the gene set in:

Biological Processes: DNA damage response, mismatch repair, apoptosis, cardiac muscle contraction

Molecular Functions: Mismatched DNA binding, damaged DNA binding, ATP-dependent repair activity

Pathways: p53 signaling, homologous recombination, lipid metabolism, and cardiac muscle contraction pathways

These findings provide a functional context for how the identified genes may contribute to CHD. For example, impaired DNA repair and apoptosis regulation can lead to endothelial dysfunction, while mutations in cardiac structural proteins directly affect contractility and cardiac output.

***5. Reactome: pathway analysis***

Reactome analysis mapped genes to well-characterized pathways, confirming their involvement in DNA repair, apoptosis, cell cycle control, and cardiovascular processes. Pathways such as p53-mediated cell cycle checkpoints, homologous recombination repair, and cardiac muscle contraction were prominently enriched. This reinforces the hypothesis that CHD arises from the interplay of genomic stability, cellular stress response, and cardiac structural integrity.

**Integration of tools and disease implications**

By combining DisGeNET, STRING, GeneMANIA, Enrichr, and Reactome, the study provides a comprehensive multi-level view of CHD-associated genes:

- DisGeNET identifies disease relevance and literature support.
- STRING shows functional connectivity and protein interactions.
- GeneMANIA predicts additional functionally related genes.
- Enrichr highlights enriched processes, molecular functions, and pathways.
- Reactome visualizes detailed molecular pathways.

Together, these tools reveal that CHD is influenced not only by lipid metabolism and cardiac structural proteins but also by DNA repair and apoptosis pathways. The combination of these pathways may determine the onset, severity, and progression of CHD, suggesting potential targets for early diagnosis, genetic screening, and therapeutic intervention.

***Conclusion of discussion***

The bioinformatics analysis emphasizes the multi-factorial nature of CHD, where genomic instability, impaired DNA repair, cardiac muscle dysfunction, and lipid metabolism dysregulation intersect. Using these complementary tools provides both molecular-level insight and predictive power to understand disease mechanisms, prioritize candidate genes, and guide future research into genetic and therapeutic strategies for CHD.

**Conclusion**

This study provides a comprehensive bioinformatics analysis of coronary heart disease (CHD) by integrating multiple publicly available tools and databases. Using DisGeNET, disease-associated genes were identified, highlighting key genes involved in lipid metabolism (LDLR), cardiac muscle function (MYH7, MYBPC3), and DNA repair (TP53, ATM, BRCA1, BRCA2). These genes represent clinically relevant candidates for understanding the genetic basis of CHD and associated disorders.

STRING analysis revealed a highly interconnected protein–protein interaction network, demonstrating functional relationships between DNA repair proteins and cardiac-related proteins, confirming the multi-factorial nature of CHD. GeneMANIA expanded the gene set by identifying functionally related genes, emphasizing co-expression, shared pathways, and interactions that may contribute to disease susceptibility.

Functional enrichment using Enrichr and pathway mapping with Reactome highlighted that the gene set is significantly involved in DNA damage response, apoptosis, cell cycle regulation, lipid metabolism, and cardiac muscle contraction. These results suggest that CHD arises not only from traditional risk factors like lipid

dysregulation and cardiac structural defects but also from genomic instability and impaired cellular stress responses.

The integrated use of these tools provides a holistic understanding of CHD at the molecular, functional, and pathway levels, offering potential targets for diagnosis,therapeutic intervention, and genetic screening. Overall, this study demonstrates the power of bioinformatics approaches in unraveling complex disease mechanisms and guiding future research in cardiovascular genetics.

**References**

1. Libby, P. (2021). The changing landscape of atherosclerosis. *Nature*, 592(7855), 524–533. https://doi.org/10.1038/s41586-021-03392-8
2. Ross, R. (1999). Atherosclerosis—An inflammatory disease. *New England Journal of Medicine*, 340(2), 115–126. https://doi.org/10.1056/NEJM199901143400207
3. World Health Organization. (2023). *Cardiovascular diseases (CVDs)*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
4. Mensah, G. A., Roth, G. A., & Fuster, V. (2019). The global burden of cardiovascular diseases and risk factors: 2020 and beyond. *Journal of the American College of Cardiology*, 74(20), 2529–2532. https://doi.org/10.1016/j.jacc.2019.10.009
5. Yusuf, S., Joseph, P., Rangarajan, S., Islam, S., Mente, A., Hystad, P., Brauer, M., Kutty, V. R., Gupta, R., Wielgosz, A., AlHabib, K. F., Dans, A., Lopez-Jaramillo, P., Avezum, A., Lanas, F., Oguz, A., Kruger, I. M., Diaz, R., Yusoff, K., … Teo, K. (2020). Modifiable risk factors, cardiovascular disease, and mortality in 155,722 individuals from 21 high-income, middle-income, and low-income countries (PURE): A prospective cohort study. *The Lancet*, 395(10226), 795–808. https://doi.org/10.1016/S0140-6736(19)32008-2
6. American Heart Association. (2022). Heart disease and stroke statistics—2022 update: A report from the American Heart Association. *Circulation*, 145(8), e153–e639. https://doi.org/10.1161/CIR.0000000000001052
7. Benjamin, E. J., Muntner, P., Alonso, A., Bittencourt, M. S., Callaway, C. W., Carson, A. P., Chamberlain, A. M., Chang, A. R., Cheng, S., Das, S. R., Delling, F. N., Djousse, L., Elkind, M. S. V., Ferguson, J. F., Fornage, M., Jordan, L. C., Khan, S. S., Kissela, B. M., Knutson, K. L., … Virani, S. S. (2019). Heart disease and stroke statistics—2019 update: A report from the American Heart Association. *Circulation*, 139(10), e56–e528. https://doi.org/10.1161/CIR.0000000000000659