BIOINFORMATICS AND NETWORK MEDICINE

Putative disease gene identification and drug repurposing for High Blood Pressure

E.Mokhtari, A.BakhshaeeBabaroud, J.Qian
GROUP 04
Delivered: 13/01/2025

ABSTRACT

High blood pressure (HBP) is a global health concern and a leading cause of cardiovascular diseases. This study integrates bioinformatics and network medicine approaches to identify novel HBP-related genes and repurpose existing drugs. By constructing the disease interactome using protein-protein interaction (PPI) and gene-disease association (GDA) data, disease modules were characterized, and computational algorithms were evaluated for gene prediction. The top-performing method identified 100 putative genes, which were analyzed for functional enrichment. Drug repurposing efforts revealed potential therapeutic candidates, supported by clinical trial evidence. This work provides insights into the molecular basis of HBP and offers pathways for targeted treatments.

INTRODUCTION

High blood pressure (HBP) is a significant global health challenge and a leading risk factor for cardiovascular diseases, stroke, and kidney disorders. Despite its prevalence, the genetic mechanisms underlying HBP remain poorly understood, limiting the development of effective treatments.

Advances in bioinformatics and network medicine enable the systematic exploration of disease mechanisms through protein-protein interaction (PPI) networks and gene-disease associations (GDAs). Computational methods, such as DIAMOnD and diffusion-based algorithms, facilitate the identification of novel disease genes, while enrichment analysis reveals their biological significance.

This study aims to integrate PPI and GDA data to construct a disease-specific interactome for HBP. Using state-of-the-art algorithms and enrichment analysis, we identify potential disease genes and evaluate their therapeutic implications. Additionally, drug repurposing analysis is conducted to highlight candidate drugs for HBP treatment.

MATERIALS AND METHODS

Data Collection

PPI Data: Sourced from BioGRID (organism ID 9606, physical interactions only).

GDA Data: Collected from DisGeNET for high blood pressure and validated using HGNC.

Interactome Construction and Analysis

Integration: Combined PPI and GDA data to construct a disease-specific interactome.

LCC Analysis: Extracted the largest connected component (LCC) and calculated key network metrics, including node degree, betweenness centrality, and eigenvector centrality.

Disease Gene Prediction

Algorithms: DIAMOnD, DiaBLE, and a diffusion-based method.

Evaluation: Conducted 5-fold cross-validation and assessed performance using precision, recall, and F1-score.

Prediction: Applied the best-performing algorithm to identify 100 novel high blood pressure-related genes.

Functional Enrichment

Tools: Used EnrichR to perform GO (BP, MF, CC), KEGG, and Reactome pathway enrichment on predicted genes.

Analysis: Compared enriched functions of predicted genes with original disease genes.

Drug Repurposing

Identification: Queried DGIdb for drugs associated with the top 20 predicted genes.

Validation: Verified top three drugs using ClinicalTrials.gov for relevance to high blood pressure.

Tools and Software

Python (pandas, NetworkX, matplotlib, seaborn) for analysis and visualization.

RESULTS AND DISCUSSION

Table 1 Summary of GDAs and basic network data

disease name	UMLS disease ID	MeSH disease class	number of associated genes	number of genes present in the interactome	LCC size of the disease interactome
High Blood Pressure	C0020538	C14907489	300	286	559

We get the first 50 disease genes in the disease LCC ordered for node degree from higher to lower. We have listed the Top 10 in the following Table 2. The complete Top 50 can be viewed in the file 'top_50_disease_genes_metrics.csv'.

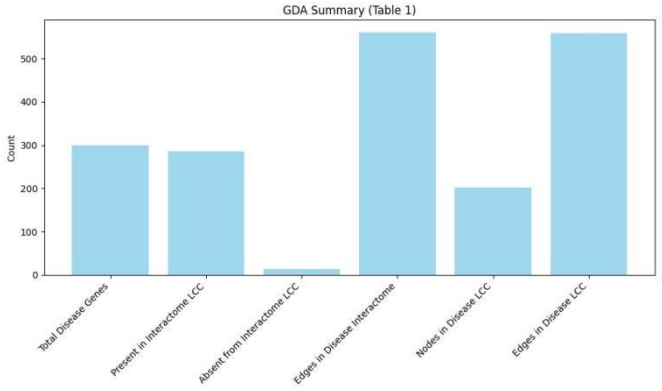
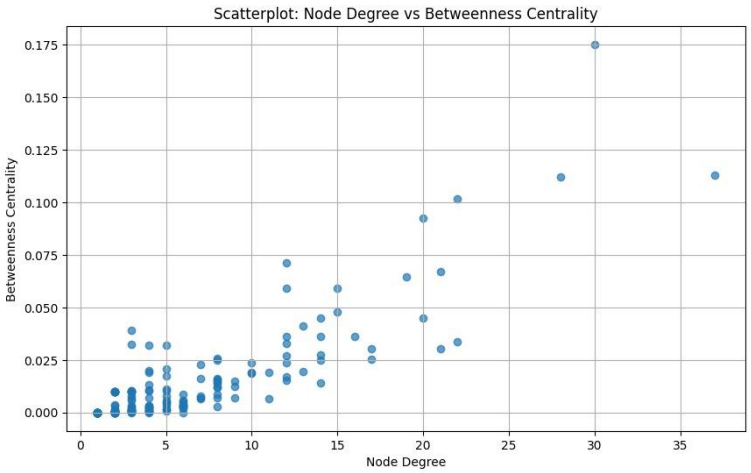


Table 2 Main network metrics of disease LCC genes

Ranking	Gene name	Degree	Betweenness	Eigenvector centrality	Closeness centrality	ratio Betw./Degree
1	TP53	37	0.112787	0.356004	0.436957	0.003048
2	CAV1	30	0.175039	0.160273	0.425847	0.005835
3	ESR2	28	0.112125	0.186855	0.424947	0.004004
4	FN1	22	0.101800	0.153719	0.399602	0.004627
5	BRCA1	22	0.033648	0.236521	0.401198	0.001529
6	HIF1A	21	0.066906	0.203476	0.396450	0.003186
7	RELA	21	0.030302	0.220274	0.385797	0.001443
8	GSK3B	20	0.044837	0.207553	0.406061	0.002242
9	VHL	20	0.092464	0.163358	0.390291	0.004623
10	ACE2	19	0.064562	0.136626	0.402806	0.003398

We plotted the node degree and node betweenness in a scatterplot which you can see below:



PERFORMANCE COMPARISON

We applied the DIAMOnD algorithm and saved the top 100 genes.

The first 5 top genes are:

Ranking	DIAMOnD_node	p_hyper
1	PLAGL1	1.476242e-14
2	NFYA	1.538286e-14
3	PIAS1	1.849565e-14
4	PPARD	2.355213e-14
5	PIAS3	2.705558e-14

Starting from the DIAMOnD code, we change the universe size used in the hypergeometric function to perform DiaBLE algorithm then we saved the top 100 genes.

Top 5 genes are:

Ranking	DIAMOnD_node	p_hyper
1	ESRRA	1.274450e-12
2	NR1H3	1.774340e-12
3	HDAC3	2.049494e-12
4	CCND1	2.599538e-12
5	PPARGC1A	5.179300e-12

Then we use Diffusion-based algorithm (available on Cytoscape), diffusion times (arbitrary unit): t=0.002, 0.005, 0.01. You can see the results in the files "Diffusion_t0.002", "Diffusion_t0.005", "Diffusion_t0.01" and here are the Top-5s:

	Gene	DiffusionScore
0	TRIM67	0.002779
1	ZRANB1	0.002370
2	PARK2	0.001956
3	RPA1	0.001731
4	PLEKHA4	0.001686

t = 0.002

	Gene	DiffusionScore
0	TRIM67	0.002774
1	ZRANB1	0.002358
2	PARK2	0.001952
3	RPA1	0.001721
4	PLEKHA4	0.001677

t = 0.005

	Gene	DiffusionScore
0	TRIM67	0.002767
1	ZRANB1	0.002338
2	PARK2	0.001945
3	RPA1	0.001704
4	KIAA1429	0.001662

t = 0.01

For computational validation we performed a 5-fold cross validation and compute the performance metrics for the diamond and diable algo the results are:

DIAMOnD:

average Precision: 0.03 ± 0.03

average Recall: 0.04 ± 0.03

average F1 Score: 0.03 ± 0.02

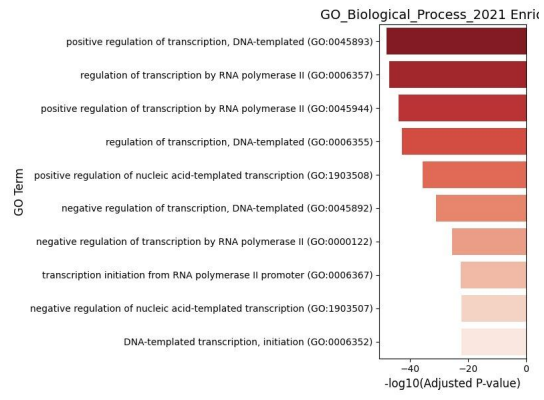
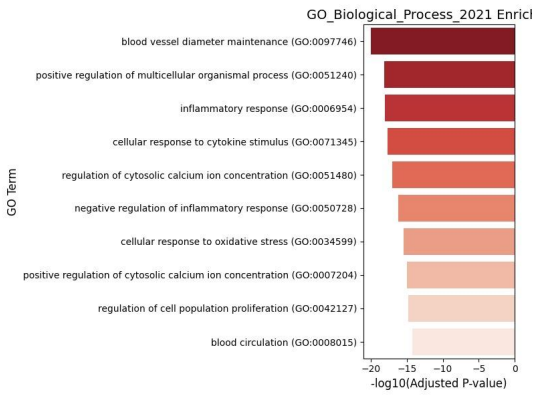
DiaBLE:

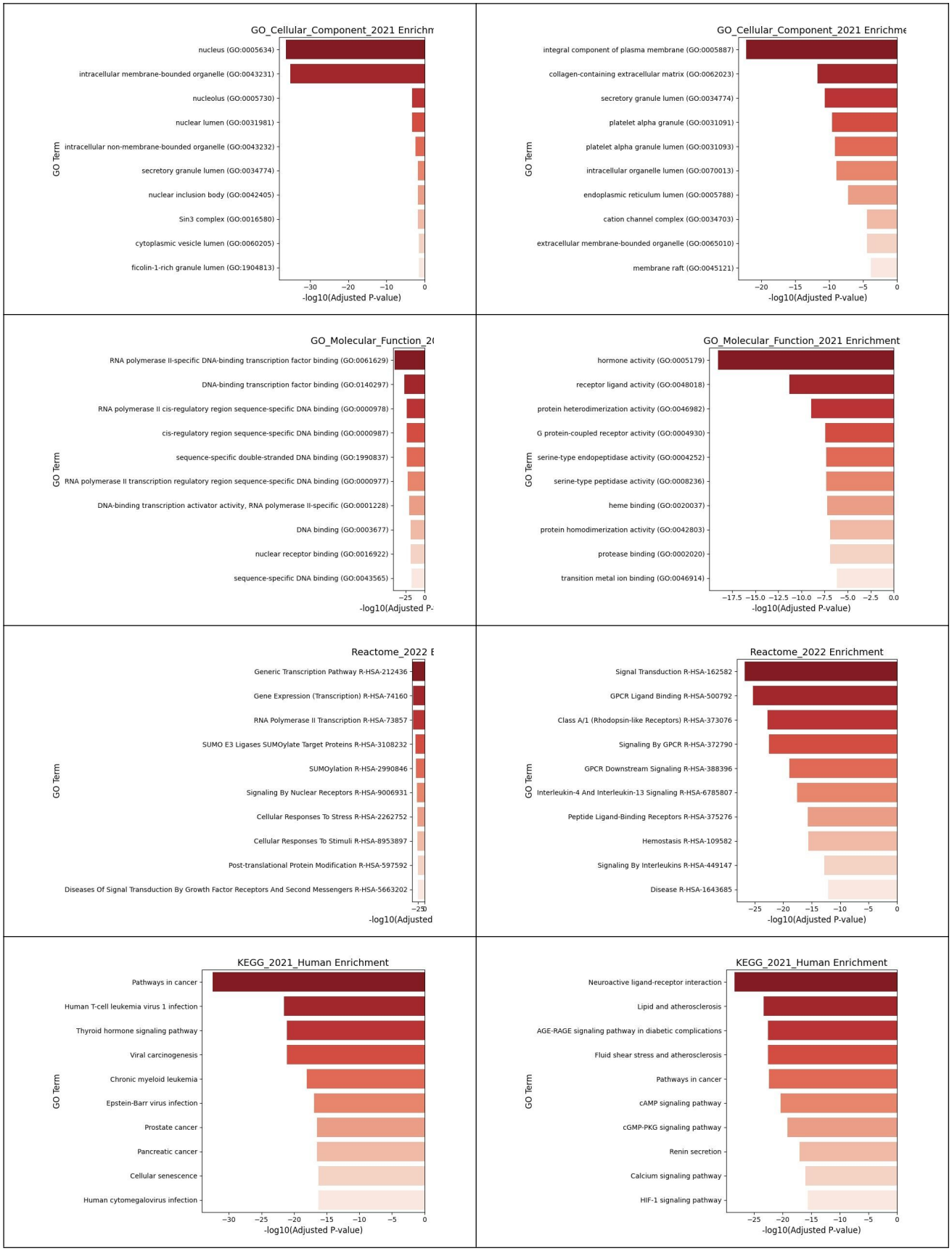
average Precision: 0.03 ± 0.03

average Recall: 0.03 ± 0.02

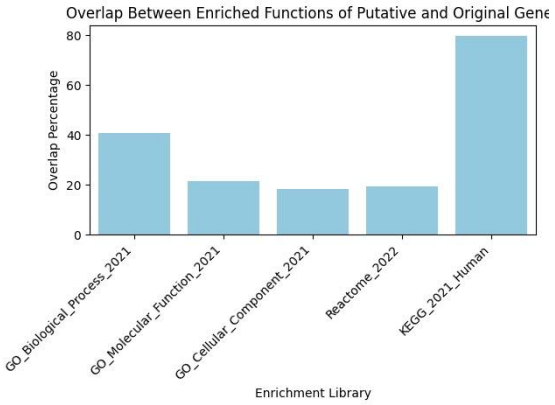
average F1 Score: 0.03 ± 0.03

According to the performance metrics obtained in the validation phase at point 2, we performed the DIAMOnD algo on the whole seed genes and saved the top 100 putative disease genes and the Enrichment analysis on the top 100 genes and also on the whole gene set you can see the results below:

On the putative disease genes	On the original disease genes
	



The overlap between enriched functions of original disease genes and putative disease genes:



We compiled a ranking of identified drugs, starting with the drug associated with the most of the above 20 genes and below is the top 3:

rank	drug_name	Gene_count
1	BORTEZOMIB	8
2	CAPIVASERTIB	7
3	CARFILZOMIB	7

Upon reviewing the top 2 drugs identified in Part 4.1 through the ClinicalTrials.gov database, we could not find any clinical trials testing the drugs for high blood pressure.

For the third one, CARFILZOMIB, we find 2 trails:

Cardiovascular Complications of Carfilzomib (NCT04407858)

Right Ventricular Function in Patients Taking Carfilzomib (NCT06568952)

AUTHOR CONTRIBUTIONS

Here is a short description of the contribution to the project of each author:

E.M, A.B, J.Q: data gathering; algorithm implementation; tasks; cross-validation; original draft preparation & review & editing.

REFERENCES

1. Wu P, Feng Q, Kerchberger VE, Nelson SD, Chen Q, Li B, Edwards TL, Cox NJ, Phillips EJ, Stein CM, Roden DM, Denny JC, Wei WQ. Integrating gene expression and clinical data to identify drug repurposing candidates for hyperlipidemia and hypertension. Nat Commun. 2022 Jan 10;13(1):46. doi: 10.1038/s41467-021-27751-1. PMID: 35013250; PMCID: PMC8748496.Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, Doig A, Guilliams T, Latimer J, McNamee C, Norris A, Sanseau P, Cavalla D, Pirmohamed M. Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov. 2019 Jan;18(1):41-58. doi: 10.1038/nrd.2018.168. Epub 2018 Oct 12. PMID: 30310233.
2. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov. 2004 Aug;3(8):673-83. doi: 10.1038/nrd1468. PMID: 15286734.
3. Scannell JW, Blanckley A, Boldon H, Warrington B. Diagnosing the decline in pharmaceutical R&D efficiency. Nat Rev Drug Discov. 2012 Mar 1;11(3):191-200. doi: 10.1038/nrd3681. PMID: 22378269.