Identification of the shared gene signatures and biological mechanism in Atrial Fibrillation and Ischemic cardiomyopathy

Tuersunjiang Naman
First Affiliated Hospital of Xinjiang Medical University
Ailiman Mahemuti
First Affiliated Hospital of Xinjiang Medical University
Aihaidan Abudouwayiti
First Affiliated Hospital of Xinjiang Medical University
Juan Sun
First Affiliated Hospital of Xinjiang Medical University
Refukaiti Abuduhalike (refukaiti623@163.com)
First Affiliated Hospital of Xinjiang Medical University

Article

Keywords: Atrial fibrillation, Ischemic cardiomyopathy, gene signatures

Posted Date: May 2nd, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2855761/v1

Abstract

Objective

the relationship between Atrial fibrillation(AF) and Ischemic cardiomyopathy(ICM) has long been widely recognized, but no previous study explore the interaction mechanisms and method to confirm the casual relation ship of them. this study was aimed to investigate the shared gene signatures and molecular process between AF and ICM.

Methods

The Gene expression Omnibus(GEO) database was used to retrieve the data of AF and ICM used in this syudy. Weighted gene co-expression network analysis(WGCNA) was performed to discover associated genes with AF and ICM. Than intersecting the correlated genes by venn and filtered out the co-expression gene both AF and ICM, than performed GO enrichmen analysis and PPI network analysis, than usig the cytoscape software screened out the most significant hub genes correlated with AF and ICM, finally validated by other dataset.

Result

using WGCNA the most correlated module with AF and ICM respectively were identified. By venn the most correlated genes with AF and ICM filtered out 188 genes as the co-expressed genes with AF and ICM. than performed GO enrichmen analysis and PPI network analysis, than usig the cytoscape software screened out the three most significant hub genes(CHD1,MSH2, NIBPL) correlated with AF and ICM, finally validated using other dataset by ROC, all of thre genes has the good discrimination ability both in AF and ICM.

Conclusion

This work firstly revealed the three co-expression genes correlated with AF and ICM, maybe the expression level of these genes have diagnostic value of AF and ICM and have diagnostic value to confirm the cause of AF is ICM.

1. Introduction

AF is the most prevalent cardiac tachyarrhythmia encountered in clinical practice[1]. In 2010, it was estimated that there were already 33 million people with AF worldwide, and projection suggest this may exceed 70 million in Asia alone by 2050[2], Complications of AF include an elevated risk of stroke, heart failure, dementia, and premature death [3]. Additionally, associated symptoms include chest pain, palpitations, fatigue, dizziness, syncope, dyspnea, and a decrease in exercise capacity[4]. Furthermore, recent data suggest that there is an increasing burden of AF-related symptoms and complications on healthcare systems. For example, hospitalizations for AF in Australia have demonstrated a 295% increase

in recent years and now represent the most common cause for cardiovascular hospitalization[5, 6]. Tese trends are concerning and suggest that there is a need to evaluate current management approaches and investigate new treatments.

The population suffering from AF complicated with coronary heart disease (CHD) is rising rapidly, A strong correlation between the two diseases has been reported. In addition, CHD can directly promote the progression of AF by affecting reentry formation, focal ectopic activity and neural remodeling. At the same time, AF also affects CHD through three aspects atherosclerosis, the mismatch of blood supply and oxygen consumption, and thrombosis. In conclusion, CHD and AF can aggravate each other and seem to form a vicious cycle Many studies have shown that patients with CHD tend to develop atrial fibrillation (AF)[7] On the other hand, AF was an independent predictor and led a 2.2 times higher probability of new coronary events[8]. According to the REGARDS study, the rate of myocardial infarction (MI) was approximately 2-fold higher in patients with AF[9].This association is particularly notable in women and black people[10]. To be precise, the incidence of CHD in patients with AF was as high as 34% [11]. To make things worse, patients with CHD and AF together tend to have worse outcomes, including higher complication rates and mortality[12], regardless of the type of CHD and AF.

Although it is reported that: CHD is correlated with AF, The lesions and severity of CHD are the significant facrors which related to the occurrence of AF. Some studies have found a significant correlation between infarction or lesions in the right coronary artery (RCA) and NOAF[11, 13]. Controversially, several studies have indicated that patients with NOAF are more likely to have left main coronary artery disease (CAD)[14]. According to a study in Israel, severe CAD in the RCA and left circumflex were significant predictors of early AF, while left main and left anterior descending CAD were linked with late AF[15]. The blockage in atrial branches, which usually originate from the RCA and supply blood to the atria, provides an explanation for the association between RCA lesions and AF.

In summary, NOAF is more common in severe CHD, and the incidence of NOAF is proportional to the degree of myocardial ischemia. Long-period of myocardial ischemia resulted in ischemic cardiomyopathy, the Ischemic cardiomyopathy is the end stage of coronary heart disease. Although previous studies reported the correlation between AF and coronary artery disease and myocardial ischemia, but no study reported the correlation between AF and ischemic cardiomyopathy, and let alone the co-expressive biomarker for diagnosis of AF and ICM. Thus this study tend to investigate the coexpressive genes in AF and ICM, and give clinicians to a an effective toll to diagnosis of AF complicated with ICM, and help to confirm the cause of AF is ICM.

2. Materials and Methods

2.1 Data acquisition and preprocessing. In this study, two microarray expression datasets of ICM(GSE1869, GSE5406) and other two microarray expression datasets of AF(GSE41177, GSE14975) were acquired from the Nationnal Center for Biotechnology Information Gene Expression Omnibus

database(NCBI-GEO; https://www.ncbi.nlm.nih.gov/geo/) . in order to obtain a larger sample size, two GEO datasets of ICM (GSE1869 and GSE5406) were merged, samely merged the two GEO datasets of AF(GSE41177 and GSE14975). And ICM datasets and AF datasets are standardized and normalized using the “limma” package seperately, Used the "sva" package to conduct the batch effect. And download the another ICM datasets(GSE116250 and GSE9128) and AF dataset(GSE79768) to use for the validation cohort.

2.2 Weighted gene Co-expression network analysis, The Weighted gene Co-expression network analysis(WGCNA) is an algorithm that can find the co-expression gene modules with high biological significance and explore the relationship between gene networks and disease[16]. Therefore, we used the WGCNA to obtain the ICM and AF associated modules. The “WGCNA” package in R 4.0.3 software were used to perorm the WGCNA analysis. Before analysis, the hierarchial clustering analysis was performed using the Hclust function in Rlanguage to exclude the outlier samples. Then the appropriate softpowers β(ranged from 1 to 20) was selected using the function of “picksoft ” in the WGCNA package according to the standard of scale free network. Next, the soft power value β and gene correlation matrix among all gene pairs calculated by pearson analysis were used to build adjacency matrix. Then the topological overlap matrix(TOM) and the corresponding dissimilarity(1 - TOM) was transformed from the adjacency matrix. A hierarchical clustering dendogram was further built and similar gene expression were devided into different modules. Finally, The expression profiles of each module were summarized by the module enginee(ME), and correlation between the ME and clinical features was calculated. Therefore, the modules with high correlation coefficient with clinical features were focused and the genes in these modules were selected subsequent analysis. In this study, the soft thresholdβwas 6 in the WGCNA analysis of ICM and AF.

2.3 Identification of Shared and Unique Gene signatures in ICM and AF and performed pathway enrichment analysis. We selected the modules that were highly relevant to ICM and AF. The shared genes in modules positively associated with ICM and AF were overlapped using venn[17]. Then performed the protein-protein interaction(PPI) network analysis on these overlapped gene in venn and make it visualize by Cytoscape. then the GO(BP, CC,MF) analysis and KEGG analysis of these genes were performed. the *P*-value< 0.05 was considered significant.

we performed Subnetwork extraction, on overlapped genes by venn, By using the “cytohubba” algorithm in cytoscape in cytoscape software version 3.7.2 Four topological analysis methods are used, and extract the top 10 genes in the every subnetwork. The Four topological analysis methods are EPC Edge Percolated component , DMNC Density of Maximum Neighborhood Component MCC Maximal Clique Centrality and MNC Maximum neighborhood component . These topology analysis methods rank and select the top 10 according to the attributes of nodes in the network. Then the every ten genes retained by Four topological analysis methods were overlapped using venn.

2.4 then we performed Lasso regression analysis on 7 hub genes in ICM merged dataset and AF merged dataset, then performed intersection by venn.

2.5 Lollipop chart of correlation between single gene and immune cell and immune process. Calculate the enrichment fraction of 29 immune cells and immune processes in the ICM combined gene set and AF combined gene set separately by using ssGSEA method. Then performed correlation analysis between enrichment fraction and expression level of hub genes, and displayed the Lollipop chart.

2.6 Validation, drawed ROC curve and calculated the AUC value of hub genes in the ICM combined gene set and AF combined gene set separately. And developed ROC and calculated AUC both in the validation dataset of ICM(GSE116250 and GSE9128) and validation dataset of AF(GSE79768).

2.7 predict and developed the PPI map of the genes which express the Protein-protein interaction with three hub genes. and performed functional enrichment analysis of these genes then displayed the top five genes. And Histogram of the potential transcription factors of the three genes jointly predicted by six databases, and according to the Mean rank displayed top ten transcription factors.

3. Results

Our study collected and merged two ICM expression profiling datasets (GSE1869, GSE5406) and two AF expression profiling datasets(GSE41177, GSE14975), All datasets are standardized and normalized using the “limma” package, and by using the “sva” package batch effects were performed both on the ICM dadasets and AF datasets for subsequent analyses **(Figure1 A, B)**

3.1 Co-expression modules in ICM and AF

In the ICM combined dataset(GSE1869 and GSE5406), When 0.8 was used as the correlation coefficient threshold, the soft-thresholding power was selected as seven (Figure 1A). Through WGCNA analysis, 7 co-expression modules were identified. And It clearly indicated that the Red module was most significantly associated with ICM(P=0.004), and it include 1307 genes.

In the AF combined dataset(GSE41177 and GSE14975), When 0.8 was used as the correlation coefficient threshold, the soft-thresholding power was selected as five (Figure 1A). Through WGCNA analysis, 7 co-expression modules were identified. And It clearly indicated that the Blue module was most significantly associated with ICM(P= 4*10-4), and it include 3154 genes(Figure2).

3.2 Interaction of genes interacted correlated genes screened by WGCNA from ICM combined dataset and AF combined dataset by venn. And filtered out 188 correlated Comorbidity genes(**Figure3,A**), and performed protein-protein interaction network(PPI) analysis and visualized it by Cytoscape(**Figure3,B**).

3.3 Enrichment analysis of Comorbidity genes. Performed GO(Figure4 A.BP B.CC C.MF) and KEGG(Figure4 D.) enrichment analysis on these 188 Comorbidity genes. Enrichment analysis suggested that they were mosly involved in pathways of regulation of histone H3-K9methylation(BP), vacuolar membrane(CC), single-stranged DNA binding(MF) and Steroid biosynthesis(KEGG).

than performed Subnetwork extraction, on overlapped genes by venn, By using the “cytohubba” algorithm in cytoscape in cytoscape software version 3.7.2 Four topological analysis methods are used, and extract the top 10 genes in the every subnetwork. The Four topological analysis methods are EPC Edge Percolated component , DMNC Density of Maximum Neighborhood Component MCC Maximal Clique Centrality and MNC Maximum neighborhood component . These topology analysis methods rank and select the top 10 according to the attributes of nodes in the network(Figure5 A,B,C,D). Then the every ten genes retained by Four topological analysis methods were overlapped using venn Figure5 E and filtered out 7 hub genes.

3.4 Performing Lasso regression analysis, we filtered out the expression matrix of 7 hub genes both form ICM combined datasets and AF combined datasets, and performed Lasso regression analysis with 7 hub genes on ICM combined datasets and filtered out 5 hub genes(Figure6,A,B), samely performed Lasso regression analysis with 7 hub genes on AF combined datasets and filtered out 4 hub genes(Figure6,C,D), than intersected by venn, finally screened ou 3 hub genes(CHD1, MSH2, NIPBL),(Figure6,E).

3.5 Lollipop chart of correlation between single gene and immune cell and immune process. Calculate the enrichment fraction of 29 immune cells and immune processes in the ICM combined gene set and AF combined gene set separately by using ssGSEA method. Then performed correlation analysis between enrichment fraction and expression level of three hub genes, and displayed the Lollipop chart Figure7 .

3.6 Validation, we drawed ROC curve and calculated the AUC value of hub genes in the ICM combined gene set, the AUC value of three hub genes are 0.846 (CHD1), 0.722 (MSH2) and 0.885(NIPBL) separately(Figure8,A,B,C), The result indicated that these genes has the good diagnostic value in diagnosing ICM.

And in AF combined dataset we also drawed ROC curve and calculated the AUC value of three hub genes, the AUC value of three hub genes are 0.83 (CHD1), 0.869 (MSH2) and 0.882(NIPBL) separately(Figure8,D,E,F), The result indicated that these genes has the perfect diagnostic value in diagnosing AF.

And we validated the discermination ability of these genes by other dataset. We darawed the ROC curve and calculated the AUC value of three hub genes in the ICM genesets, GSE116250 dataset and GSE9128 dataset(Figure8,G,H). and darawed the ROC curve and calculated the AUC value of three hub genes in the AF dataset ,GSE116250 dataset,(Figure8,I), these results indicated that these three hub genes has the good discrimination ability both in ICM and AF.

3.7 we predict and developed the PPI map of the genes which have the Protein-protein interaction with three hub genes and performed functional enrichment analysis of these genes then displayed the top five genes(Figure8,A). And Histogram of the potential transcription factors of the three genes jointly predicted by six databases, and according to the Mean rank displayed top ten transcription factors(Figure9,B).

Discussion

A strong correlation between the AF and CHD has been reported, and can aggravate each other,and the many common risk factors they share may play prominent roles in their development[18], after long period of CHD result in ICM, namely ICM is the end satge of CHD, thus ICM make AF heavier then CHD.Although the objective relationship between ICM and AF has been proven by a series of studies, the underlying mechanisms are unclea[18].AF is correlated to several factors that either individually or in combination promote the initial development of the arrhythmia and the episodes that characterize the disease[19,20],Including:arterial hypertension, obesity, diabetes mellitus and genetic factors.

Just because the many factors result in AF, Thus, we can't confirm whether the ICM is the cause of AF or not in AF patients complicated with ICM. Thus clinitians unable to confirm the treatmen stategy of AF, about have to ablation therapy or treating ICM by percutaneous coronary intervention and improve cardiac function. And no previous study reported the efficient diagnostic method to fonfirm this.

In this study we screned the three co-expression genes which correlated with AF and ICM.these three genes include:CHD1, MSH2, NIPBL.

CHD1,located Chr: 98,853,985 - 98,929,007,, is the founding member of the CHD family and is conserved across all eukaryotes[21] . CHD1 is capable of assembling nucleosomes, remodeling chromatin structure, modulating histone turnover, and regulating gene transcription[22,23]. And it plays multifaceted roles in prostate cancer, the CHD1 protein is stabilized and contributes to cancer progression, tumor microenvironment remodeling, and drug resistance[24-26] .

Carriers of an MSH2 gene mutation have the highest cancer risks across the spectrum, especially for the development of urinary tract cancer[27]. Accordingly, the patient and her relatives should be followed-up closely to discover these new-onset cancers early, especially urinary tract cancer[27].

NIPBL gene plays an important role in cancer cell proliferation, migration, and infiltration in the G0/ G1 phase of cell cycle, preventing from apoptosis or autophagy and generating cell resistance to drugs[28]

By searching we found that no previous study reported the correlation between the three genes and cardiovascular disease. And foud these three genes correlated with tumor, this is the first study pupose the notion of these three genes are the co-expressed genes of AF and ICM. By the result of this clinitians can diagnose AF complicated with ICM, and can confirm the cause of AF is ICM.

However, this study still has some limitations: we screened out the three co-expression genes of AF and ICM by bioinformatic analysis, but we did not validated mechanism and function of these genes about how these genes correlated with AF and ICM, Thus further in vitro and in vivo research need to perform to identify these mechanism.

To sum up, three genetic biomarkers closely related to AF and ICM were screened by bioinformatic analysis, and used in clinic to diagnose AF combined with ICM, and confirm the cause of AF is ICM.

Declarations

Ethics Statement

The present work gained approval from Ethics Committee of the First Affiliated Hospital of Xinjiang Medical University (Approval No. 2021D01D17). All data utilized in this study is obtained from GEO database thus there is no informed consents.

Conflict of Interest

All authors claimed that there existed no competing interest.

Author Contributions

All authors contributed to this work equally Tuersunjiang Naman and Ailiman Mahemuti should be regarded as co-first author.

Found

This study was supported by the grant from the “Key Project of Natural Science Foundation of Xinjiang Uygur Autonomous Region” (No.2021D01D17).

Data availability statement

All datasets utilized during the presentstudy can be obtained in GEO (http://www.ncbi.nlm.nih.gov/geo/).

References

1. Ball J, Carrington M J, Mcmurray J J, et al. Atrial fibrillation: profile and burden of an evolving epidemic in the 21st century[J]. Int J Cardiol, 2013, 167(5): 1807-24.
2. Chugh S S, Havmoeller R, Narayanan K, et al. Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study[J]. Circulation, 2014, 129(8): 837-47.
3. Emdin C A, Wong C X, Hsiao A J, et al. Atrial fibrillation as risk factor for cardiovascular disease and death in women compared with men: systematic review and meta-analysis of cohort studies[J]. BMJ, 2016, 532: h7013.
4. Rienstra M, Lubitz S A, Mahida S, et al. Symptoms and functional status of patients with atrial fibrillation: state of the art and future research opportunities[J]. Circulation, 2012, 125(23): 2933-43.
5. Gallagher C, Hendriks J M, Giles L, et al. Increasing trends in hospitalisations due to atrial fibrillation in Australia from 1993 to 2013[J]. Heart, 2019, 105(17): 1358-1363.
6. Wong C X, Brooks A G, Leong D P, et al. The increasing burden of atrial fibrillation compared with heart failure and myocardial infarction: a 15-year study of all hospitalizations in Australia[J]. Arch

Intern Med, 2012, 172(9): 739-41.
7. Wb K, Rd A, Dd S, et al. Coronary heart disease and atrial fibrillation: the Framingham Study[J]. American heart journal, 1983, 106(2): 389-96.
8. Aronow W S, Ahn C, Mercando A D, et al. Correlation of atrial fibrillation, paroxysmal supraventricular tachycardia, and sinus rhythm with incidences of new coronary events in 1,359 patients, mean age 81 years, with heart disease[J]. Am J Cardiol, 1995, 75(2): 182-4.
9. Soliman E Z, Safford M M, Muntner P, et al. Atrial fibrillation and the risk of myocardial infarction[J]. JAMA Intern Med, 2014, 174(1): 107-14.
10. Conen D, Chae C U, Glynn R J, et al. Risk of death and cardiovascular events in initially healthy women with new-onset atrial fibrillation[J]. JAMA, 2011, 305(20): 2080-7.
11. Kralev S, Schneider K, Lang S, et al. Incidence and severity of coronary artery disease in patients with atrial fibrillation undergoing first-time coronary angiography[J]. PLoS One, 2011, 6(9): e24964.
12. Kannel W B, Abbott R D, Savage D D, et al. Coronary heart disease and atrial fibrillation: the Framingham Study[J]. Am Heart J, 1983, 106(2): 389-96.
13. Shiba T, Kondo Y, Senoo K, et al. Proximal Occlusion in the Right Coronary Artery Involving the Atrial Branch as a Strong Predictor of New-Onset Atrial Fibrillation in Acute Myocardial Infarction[J]. Int Heart J, 2019, 60(6): 1308-1314.
14. Lau D H, Huynh L T, Chew D P, et al. Prognostic impact of types of atrial fibrillation in acute coronary syndromes[J]. Am J Cardiol, 2009, 104(10): 1317-23.
15. Shiyovich A, Axelrod M, Gilutz H, et al. Early Versus Late New-Onset Atrial Fibrillation in Acute Myocardial Infarction: Differences in Clinical Characteristics and Predictors[J]. Angiology, 2019, 70(10): 921-928.
16. Yao M, Zhang C, Gao C, et al. Exploration of the Shared Gene Signatures and Molecular Mechanisms Between Systemic Lupus Erythematosus and Pulmonary Arterial Hypertension: Evidence From Transcriptome Data[J]. Front Immunol, 2021, 12: 658341.
17. Bardou P, Mariette J, Escudie F, et al. jvenn: an interactive Venn diagram viewer[J]. BMC Bioinformatics, 2014, 15(1): 293.
18. Liang F, Wang Y. Coronary heart disease and atrial fibrillation: a vicious cycle[J]. Am J Physiol Heart Circ Physiol, 2021, 320(1): H1-H12.
19. Sagris M, Antonopoulos A S, Theofilis P, et al. Risk factors profile of young and older patients with myocardial infarction[J]. Cardiovasc Res, 2022, 118(10): 2281-2292.
20. Diavati S, Sagris M, Terentes-Printzios D, et al. Anticoagulation Treatment in Venous Thromboembolism: Options and Optimal Duration[J]. Curr Pharm Des, 2022, 28(4): 296-305.
21. Flaus A, Martin D M, Barton G J, et al. Identification of multiple distinct Snf2 subfamilies with conserved structural motifs[J]. Nucleic Acids Res, 2006, 34(10): 2887-905.
22. Smolle M, Venkatesh S, Gogol M M, et al. Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange[J]. Nat Struct Mol Biol, 2012, 19(9):

884-92.

23. Stockdale C, Flaus A, Ferreira H, et al. Analysis of nucleosome repositioning by yeast ISWI and Chd1 chromatin remodeling complexes[J]. J Biol Chem, 2006, 281(24): 16279-88.
24. Zhao D, Lu X, Wang G, et al. Synthetic essentiality of chromatin remodelling factor CHD1 in PTEN-deficient cancer[J]. Nature, 2017, 542(7642): 484-488.
25. Li H, Wang Y, Lin K, et al. CHD1 Promotes Sensitivity to Aurora Kinase Inhibitors by Suppressing Interaction of AURKA with Its Coactivator TPX2[J]. Cancer Res, 2022, 82(17): 3088-3101.
26. Zhao D, Cai L, Lu X, et al. Chromatin Regulator CHD1 Remodels the Immunosuppressive Tumor Microenvironment in PTEN-Deficient Prostate Cancer[J]. Cancer Discov, 2020, 10(9): 1374-1387.
27. Vasen H F, Blanco I, Aktan-Collan K, et al. Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts[J]. Gut, 2013, 62(6): 812-23.
28. Gao D, Zhu B, Cao X, et al. Roles of NIPBL in maintenance of genome stability[J]. Semin Cell Dev Biol, 2019, 90: 181-186.

Figures

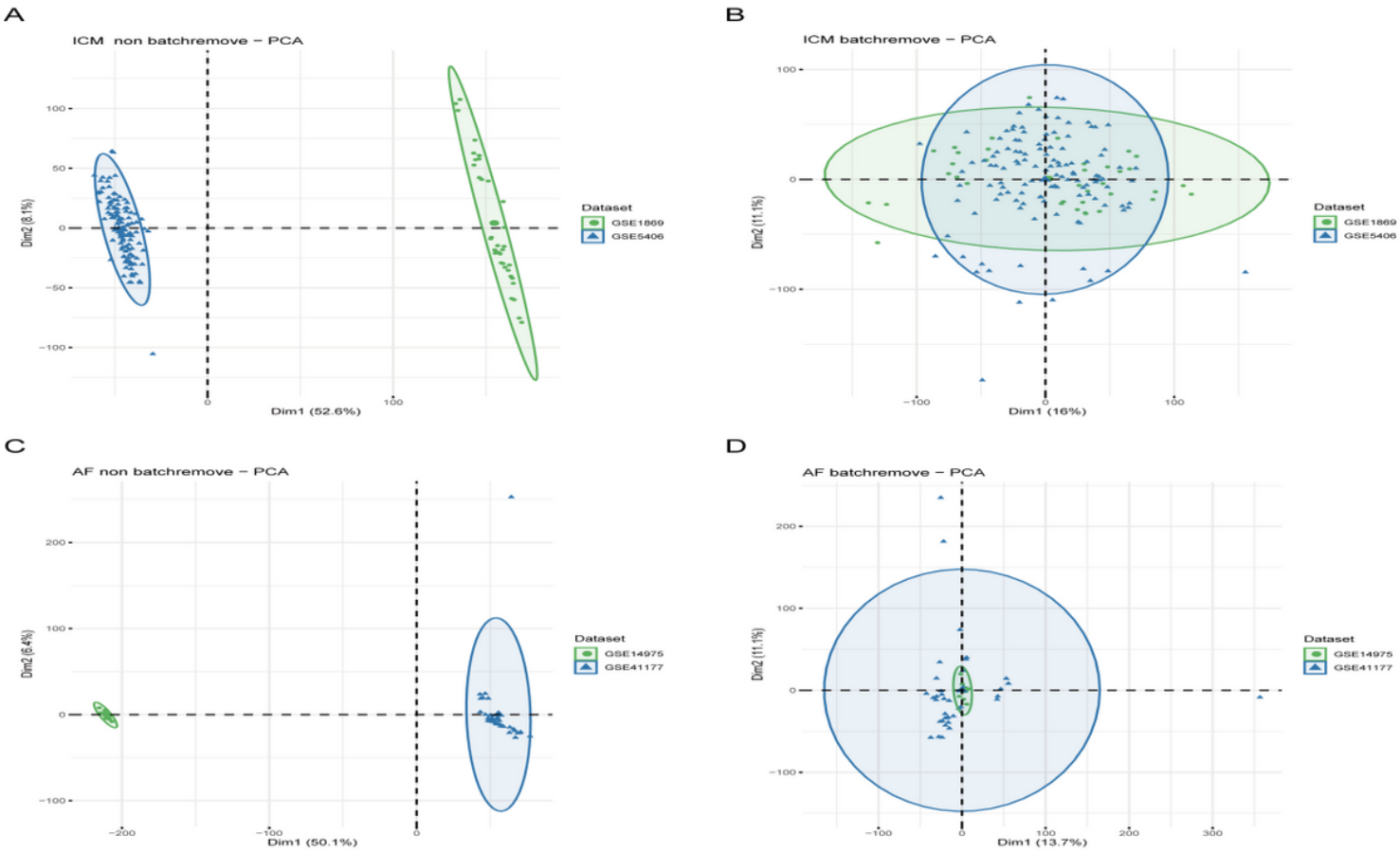

Figure 1

A. PCA map of ICM dataset before performing batch effects. **B.** PCA map of ICM dataset after performing batch effects. **C.**PCA map of AF dataset before performing batch effects. **D.** PCA map of AF dataset after performing batch effects

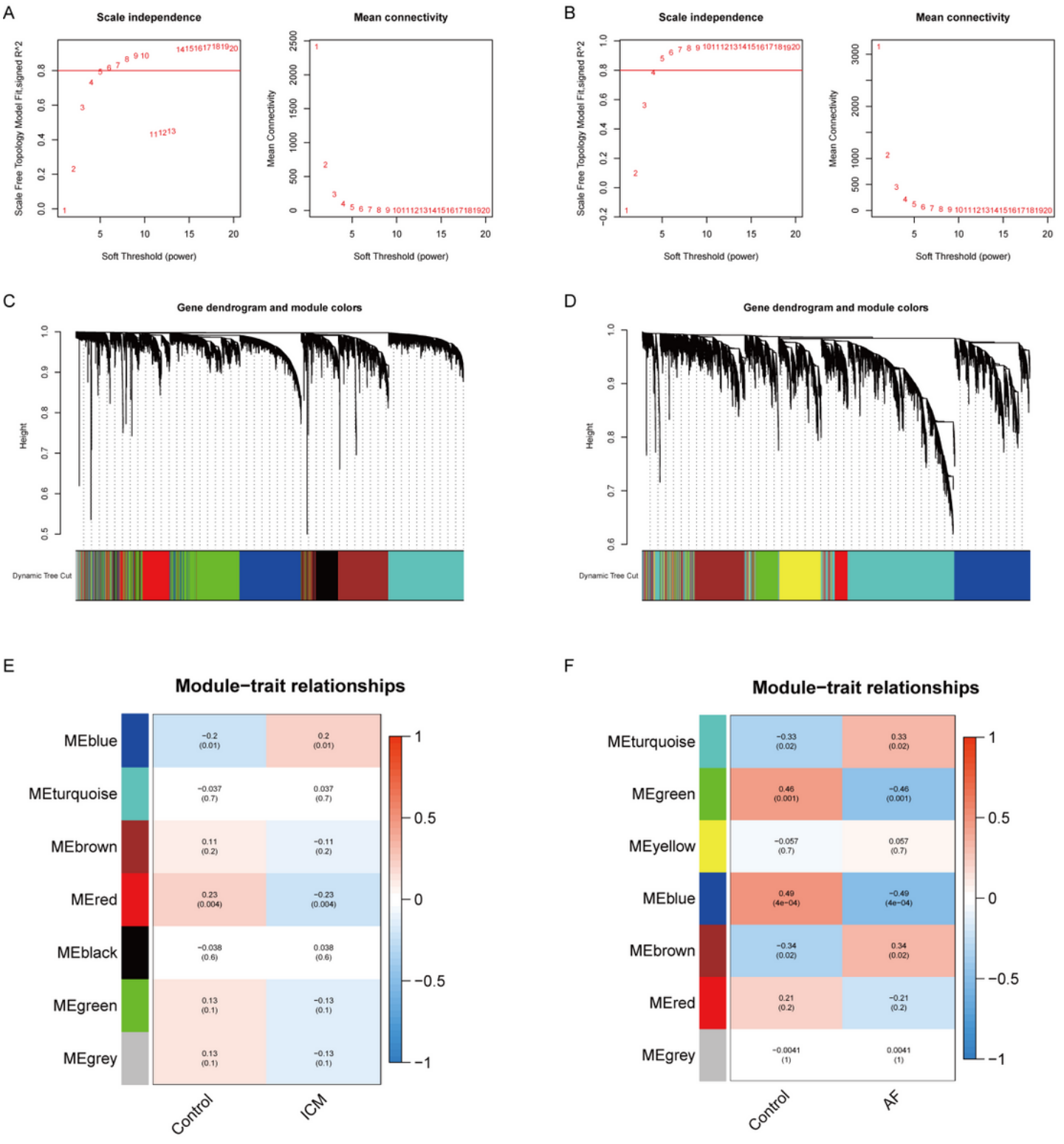

Figure 2

(A) Analysis of the scale-free fit index for various soft-thresholding powers in ICM(Left) and analysis of the mean connectivity for various soft-thresholding powers (Right); **(B)** Analysis of the scale-free fit index for various soft-thresholding powers in AF(Left) and analysis of the mean connectivity for various soft-thresholding powers in AF (Right);**(C)** cluster dendrogram of co-expression genes in ICM. **(D)**cluster dendrogram of co-expression genes in AF.**(E)** Module-trait relationship in ICM,Each cell contains the corresponding correlation and *P*values. **(E)** Module-trait relationship in Af, Each cell contains the corresponding correlation and *P* values.

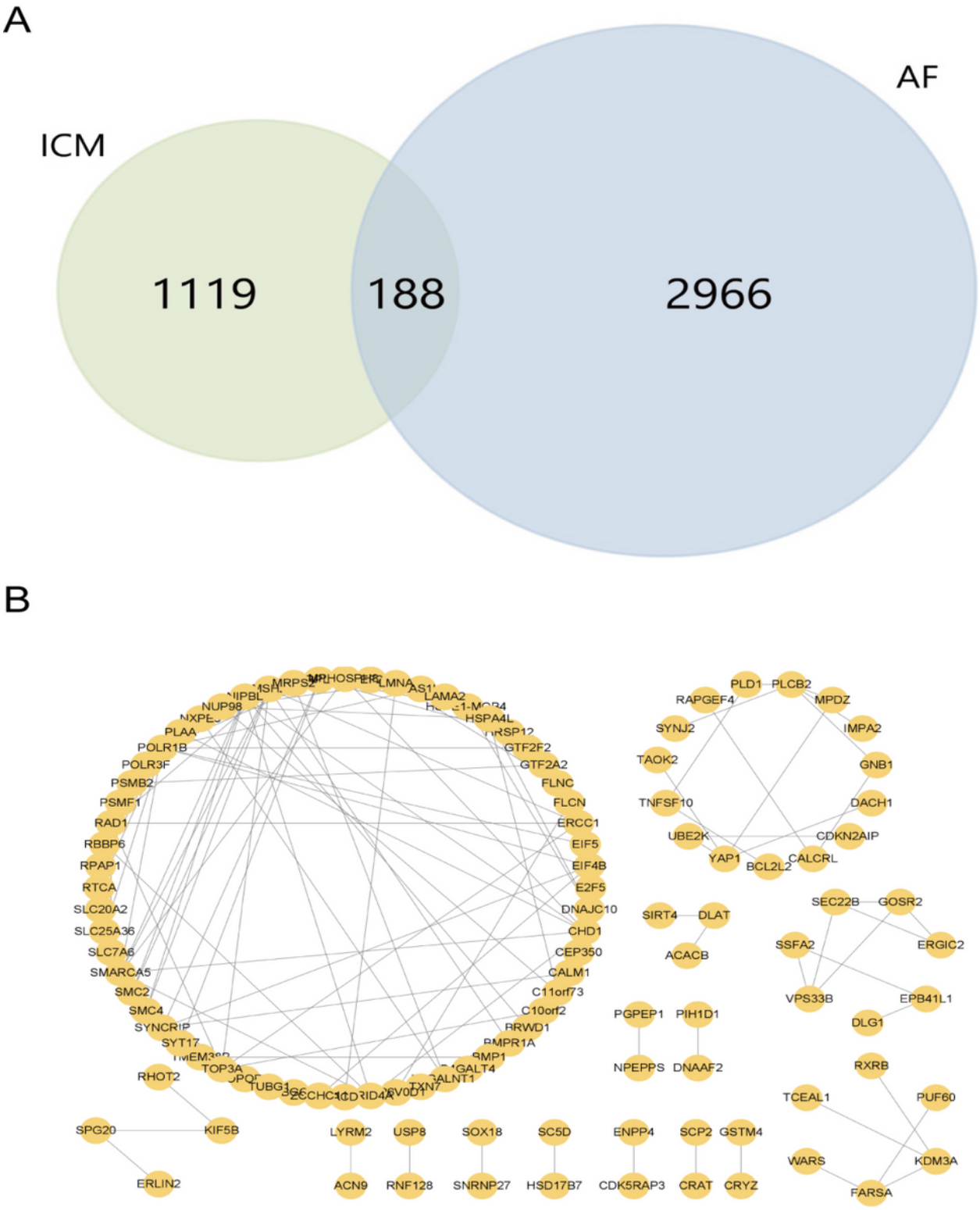

Figure 3

(A) venn map of correlated genes of ICM filtered by the highest correlated module(red) of WGCNA and correlated genes of AF filtered by the highest correlated module(red) of WGCNA. (B)PPI map of these intersected genes.

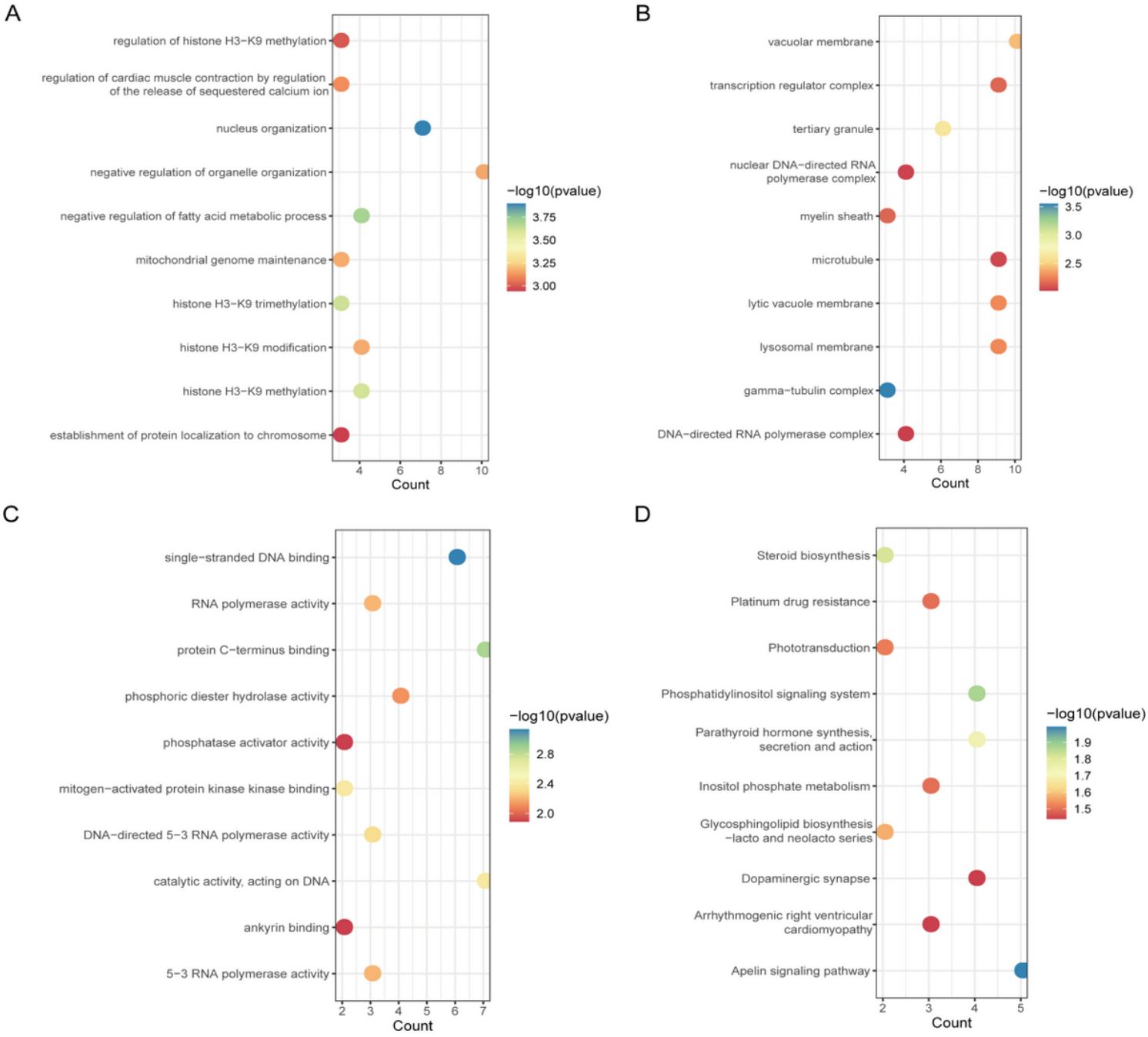

Figure 4

GO enrichment analysis(A:biological process, B:cellular component, C:molecular function) of 188 shared genes of ICM and AF. D:KEGG enrichment analysis of shared genes.

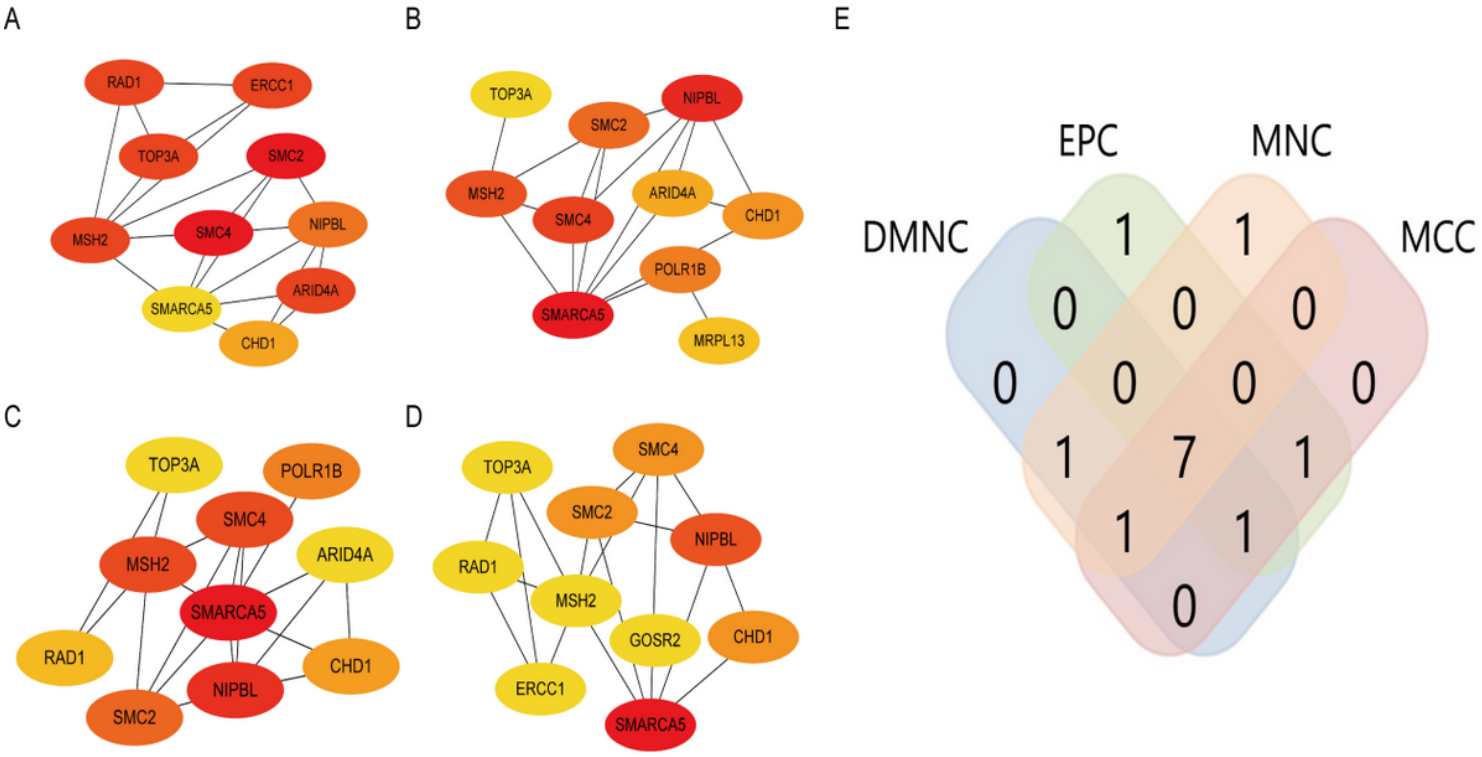

Figure 5

Topological analysis methods performed by the “cytohubba” algorithm in cytoscape, **(A)** Edge Percolated component, **(B)** Density of Maximum Neighborhood Component, **(C)**Maximal Clique Centrality, **(D)**Maximum neighborhood component, **(E)** venn map of screened genes by four topological analysis methods.

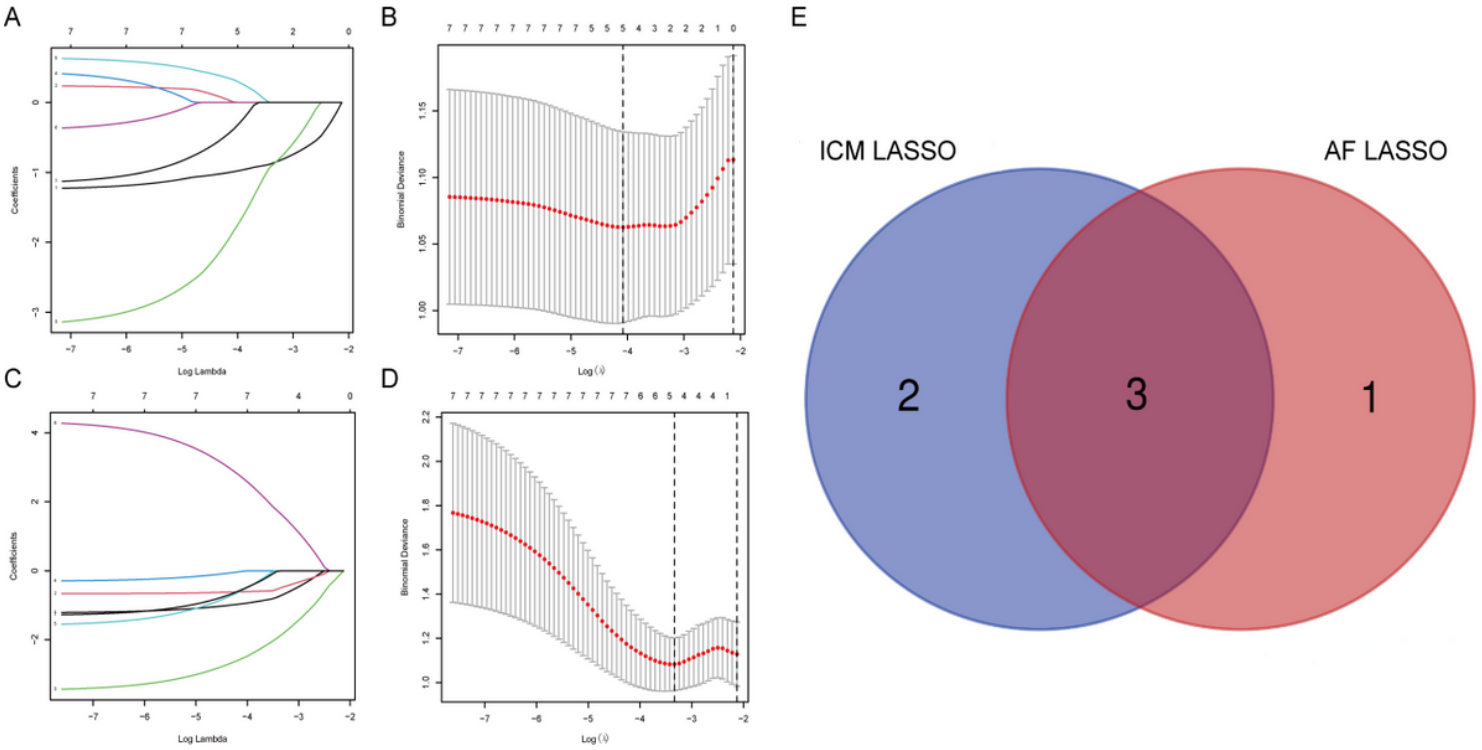

Figure 6

A Lasso regression analysis of ICM combined dataset B , LASSO regression tenfold cross validation of ICM, lambda value is 0.01692637 C Lasso regression analysis of AF combined dataset.(D) LASSO regression tenfold cross validation of ICM, lambda value is 0.03566359.

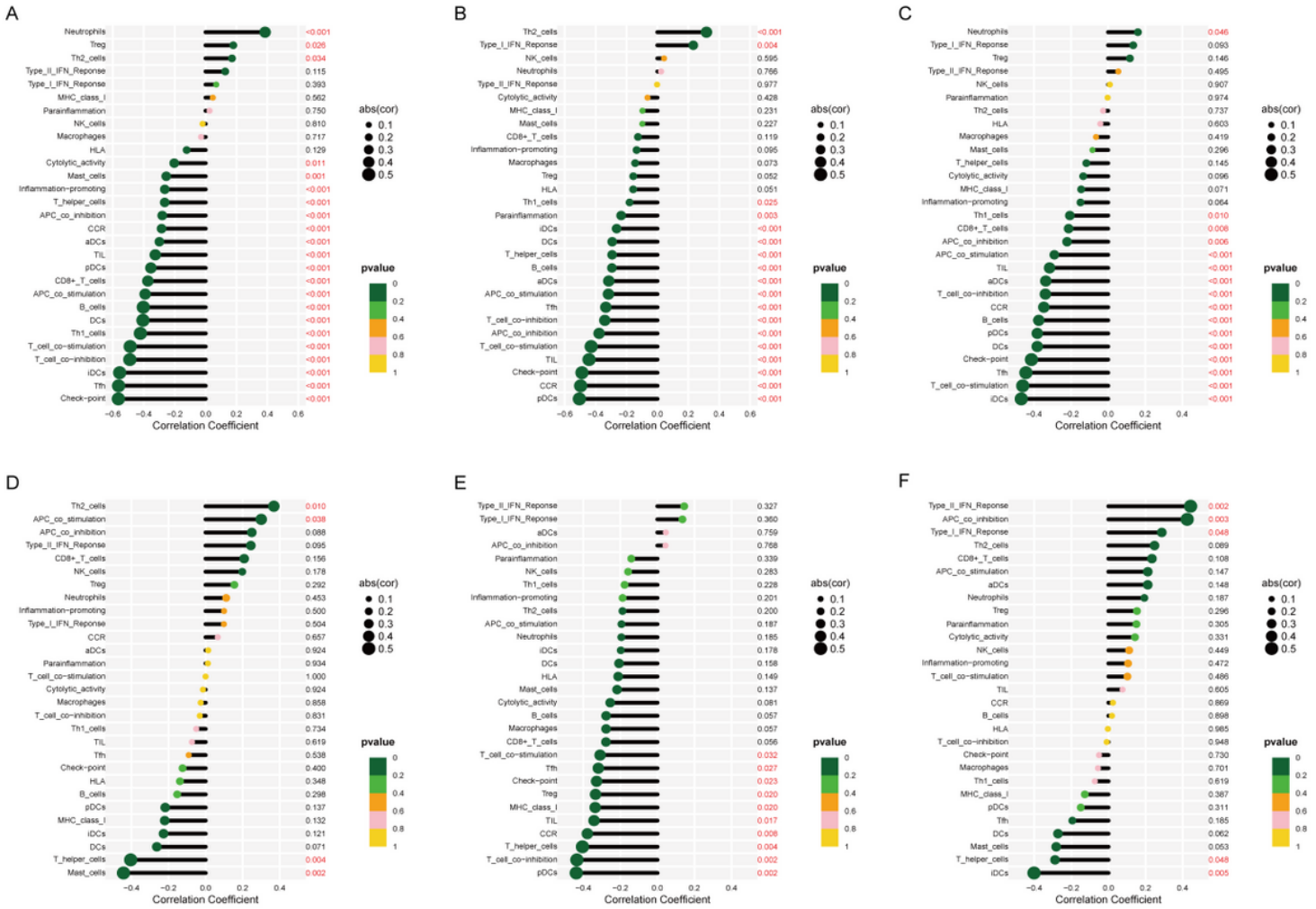

Figure 7

Lollipop chart of correlation between the enrichment fraction of 29 immune cells and immune processes in the ICM combined gene set and expression level of hub genes (A) CHD1, B MSH2,(C) NIPGL. In the AF combined gene set (D) CHD1, E MSH2, (F) NIPGL.

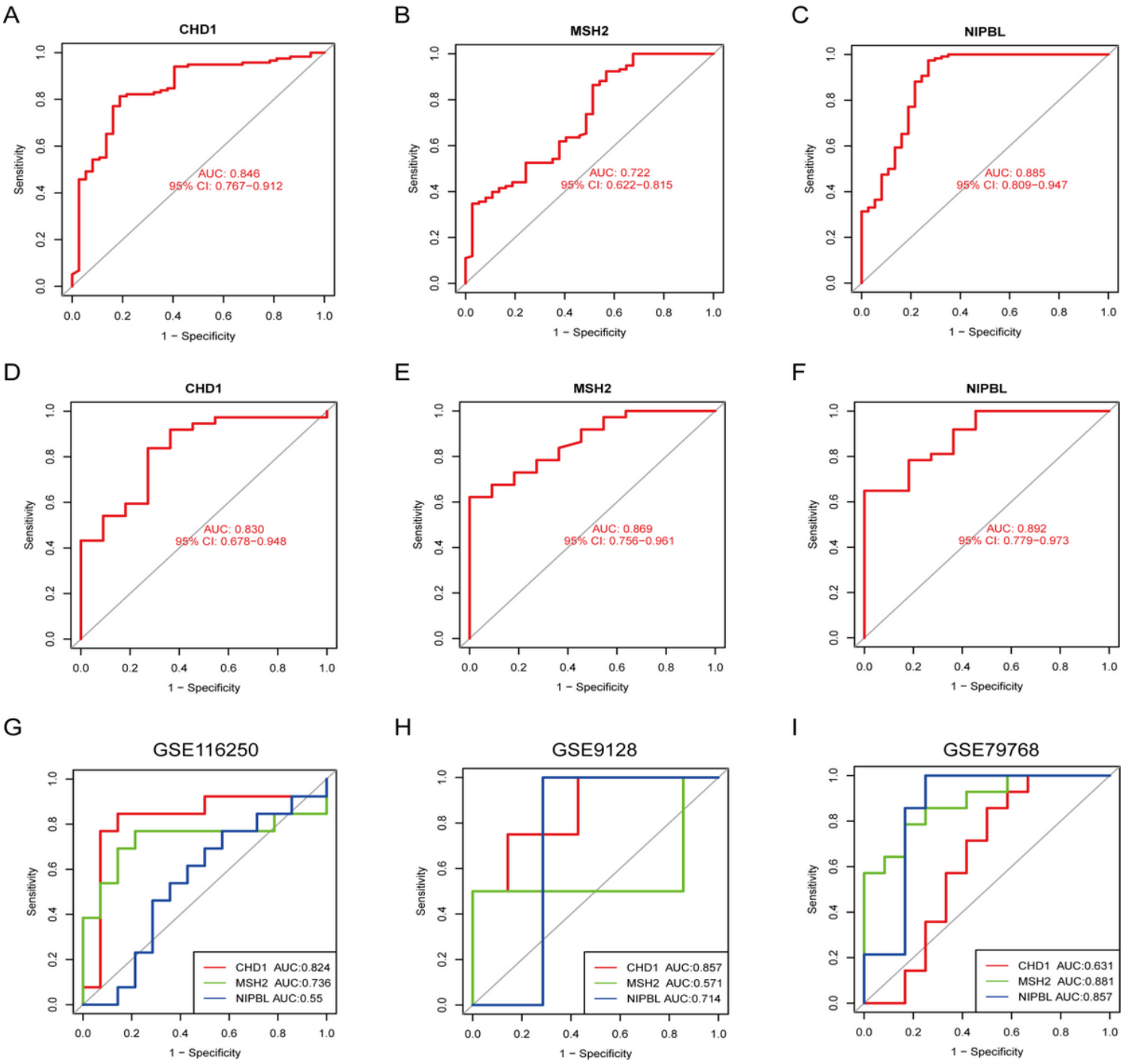

Figure 8

ROC curve and AUC value of three hub genes in ICM combined gene set, (A)CHD1 (B)MSH2 (C) NIPGL. ROC curve and AUC value of three hub genes in AF combined gene set, (D)CHD1, (E)MSH2, (F) NIPGL.ROC curve and AUC value of three hub genes in ICM validation dataset(G and H), ROC curve and AUC value of three hub genes in AF validation dataset(I)

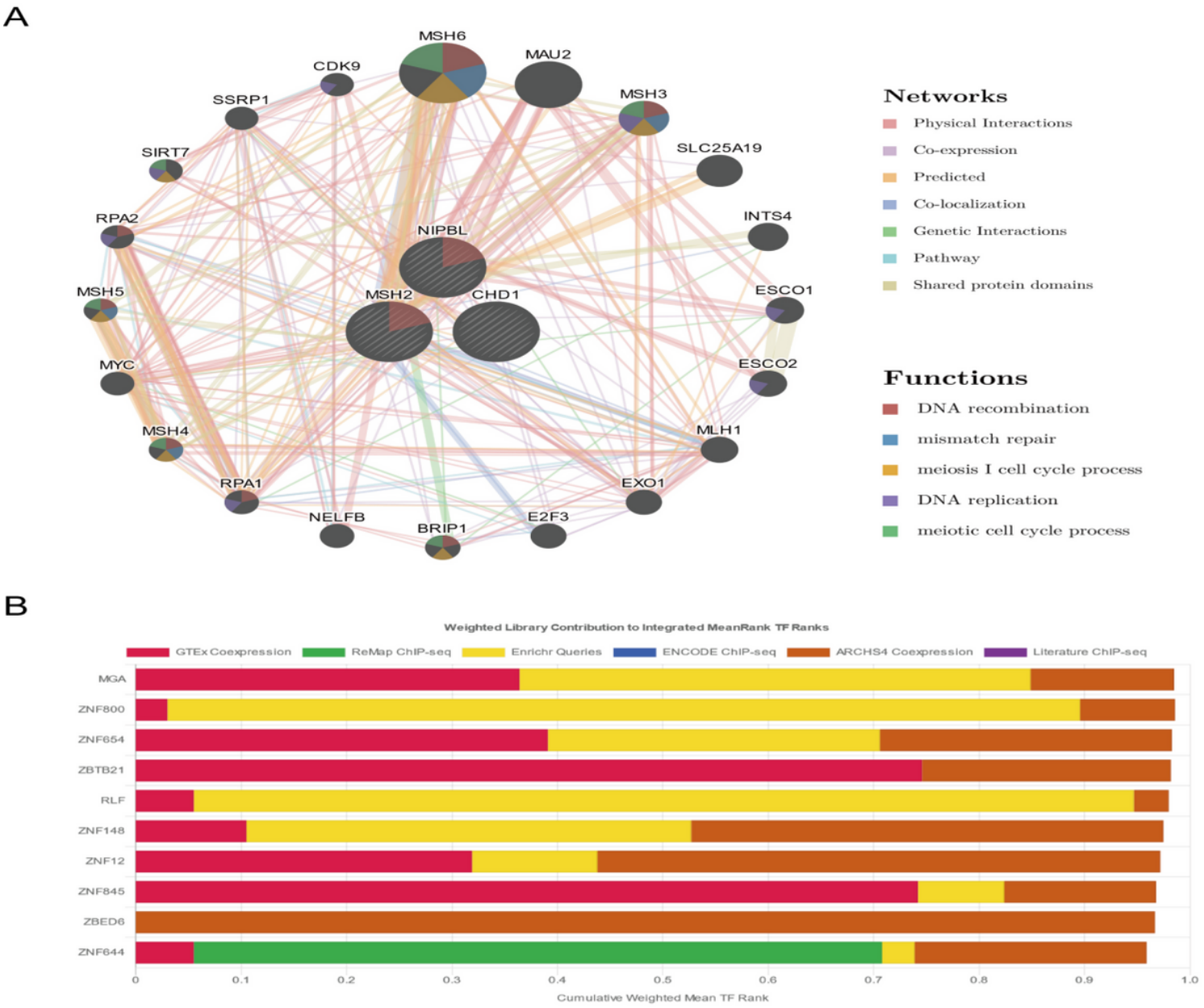

Figure 9

A: PPI map of the genes which have the protein-protein interaction with three hub genes B: Histogram of potential transcription factors of three hub genes.