Multiple Feature Selection Strategies Identified Novel Cardiac Gene Expression Signature for Heart Failure

Dan Li[1]*, Hong Lin[2] and Luyifei Li[1]

[1]Department of Cardiovascular Medicine, First Hospital Affiliated to Harbin Medical University, Harbin, China, [2]Internal Medicine-Cardiovascular Department, Harbin Chest Hospital, Harbin, China

Heart failure (HF) is a serious condition in which the support of blood pumped by the heart is insufficient to meet the demands of body at a normal cardiac filling pressure. Approximately 26 million patients worldwide are suffering from heart failure and about 17–45% of patients with heart failure die within 1-year, and the majority die within 5-years admitted to a hospital. The molecular mechanisms underlying the progression of heart failure have been poorly studied. We compared the gene expression profiles between patients with heart failure ($n = 177$) and without heart failure ($n = 136$) using multiple feature selection strategies and identified 38 HF signature genes. The support vector machine (SVM) classifier based on these 38 genes evaluated with leave-one-out cross validation (LOOCV) achieved great performance with sensitivity of 0.983 and specificity of 0.963. The network analysis suggested that the hub gene *SMOC2* may play important roles in HF. Other genes, such as *FCN3*, *HMGN2*, and *SERPINA3*, also showed great promises. Our results can facilitate the early detection of heart failure and can reveal its molecular mechanisms.

Keywords: heart failure, microarray, biomarker, network, molecular mechanism

INTRODUCTION

Heart failure (HF) is a serious condition in which the support of blood pumped by the heart is insufficient to meet the demands of body at a normal cardiac filling pressure (Ramachandra et al., 2020). Defined as a syndrome with high morbidity and mortality, HF is the major cause of death and a serious threat to human health for a long period (Jarcho, 2020). Approximately 26 million patients worldwide are suffering from heart failure, and the society faces the long-term great stresses on patients, medical stuff, and medical systems (Bowen et al., 2020). About 17–45% of patients with heart failure die within 1 year, and the majority die within 5 years admitted to a hospital in worldwide (Davison and Cotter, 2015; Zhou et al., 2020). However, the survival rates for patients with HF have improved in many parts of the world in recent years along with the advanced therapies and patient management systems. Heart failure is a complex disease, and so many factors are responsible that it is hard to blame it on one specific issue (McMurray and Pfeffer, 2005).

Over the past decades, the genetic causes and molecular mechanism underlying the progression of heart failure have been partially illustrated. Most previous studies in heart failure are limited by inadequate biological samples from patients with heart failure (Prohászka et al., 2013). Since then, studies have focused on the molecular mechanism of heart failure by virtue of animal models in combination with molecular biological techniques. Previous studies suggested that classification of disease status for HF is much important for the decision of treatment and improvement of prognosis (van Oort et al., 2011). They have discovered that novel gene biomarkers play a vital role in various diseases depending on the leapfrog development of RNA-Seq technology (Asakura and Kitakaze, 2009). According to previous reports, the specific gene expression is related to the pathological conditions of HF.

Liu et al. (2015) collected six samples from three controls, one ischemic heart disease (ISCH), and two dilated cardiomyopathies (DCMs) and used RNA-Seq to filter novel gene signatures for HF, and precisely categorize HF status in larger samples of 313 patients. Vigil-Garcia et al. (2020) selected novel genes induced during pathological cardiac hypertrophy that are relevant for human HF through cardiomyocyte-specific gene expression analysis. These results recognized PFKP as a novel potential therapeutic target to prohibit the succession of HF. Tan et al. (2002) used microarrays to describe gene expression fingerprints of HF etiologies based on seven non-failing human hearts and eight failing human hearts with a diagnosis of end-stage dilated cardiomyopathy. Zhou et al. (2020) proposed that valosin-containing protein could protect the heart against pressure overload-induced heart failure using RNA-Seq and a comprehensive bioinformatics analysis. Kittleson et al. (2004) used microarrays of 48 myocardial samples and gene expression profiling to predict biomarkers in determining prognosis and response to therapy in HF precisely. All these studies were based on microarrays, which have been the remarkable method for gene expression studies because of their ability to filter thousands of transcripts.

In our study, we tried to detect the novel HF signature genes and their networks from previous transcriptomic data which included the gene expression profiles in patients with heart failure (n = 177) and without heart failure (n = 136) using advanced bioinformatics methods. Compared with previous studies, which are intended to find the biomarker for HF put the focus on separated gene, our study focused on the linkage among them. We built the support vector machine (SVM) model with the application of multiple feature selection methods: Monte Carlo Feature Selection (MCFS; Draminski et al., 2008; Chen et al., 2018a, 2020; Pan et al., 2019b; Li et al., 2020a) and incremental feature selection (IFS; Zhang et al., 2016; Chen et al., 2018b, 2020; Wang et al., 2018; Pan et al., 2019a). What is more, we used the Search Tool for the Retrieval of Interacting Genes (STRING) database (Szklarczyk et al., 2018) to explore the protein interaction networks. A remarkable result of our study is that 38 selected genes can serve as novel

biomarkers for HF and can conduce to revealing the pathological mechanism of HF.

MATERIALS AND METHODS

The Microarray Data of Heart Failure Patients

We downloaded the microarray gene expression data of 177 patients with heart failure and 136 patients without heart failure from Gene Expression Omnibus (GEO) at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE57338 (Liu et al., 2015). The expression levels of 33,297 probes corresponding to 20,254 genes in the cardiac tissue were measured with Affymetrix Human Gene 1.1 ST Array. The probes corresponding to the same gene were averaged to obtain the gene expression levels, and the gene expression levels were quantile normalized using function normalize.quantiles from R/Bioconductor package preprocessCore[1] to minimize the systematic variance. The normalized data were used for further feature selections.

Select the Genes Based on Their Importance to Classify the Heart Failure Patients

There have been many methods for identifying differentially expressed genes (DEGs), such as t-test. But such methods only consider the distribution of one gene each time, and do not consider the relationship among genes (Tao et al., 2020). That leads to two limitations: (1) The distribution difference of a gene is not equivalent to its classification ability; and (2) The combinations of the most significant DEGs may not have good performance since they may be redundant and do not help each other to achieve a better performance. Therefore, we adopted machine learning based multiple feature selection strategies to objectively select the optimal heart failure signature. The machine learning-based methods have been widely used and achieved great success in biomarker discovery (Wang and Huang, 2019; Li et al., 2020a,b; Yuan et al., 2020; Zhang et al., 2020a,b; Zhu et al., 2020).

The proposed multiple feature selection strategies can be summarized as **Figure 1**. First, the expression profiles of 20,254 genes in 177 patients with heart failure and 136 patients without heart failure were normalized. Second, we randomly selected many subset data to construct the classification trees using Monte Carlo strategy (Draminski et al., 2008; Chen et al., 2018a, 2020; Pan et al., 2019b; Li et al., 2020a). To perform MCFS, we used the dmLab software version 2.3.0 from https://home.ipipan.waw.pl/m.draminski/mcfs.html. Third, all these trees were ensembled to calculate the classification importance of the genes. The important genes should appear in a large number of trees and be able to correctly classify the samples into right groups.

---

[1]https://bioconductor.org/packages/preprocessCore/

Fourth, the top ranked genes (1,000 in this study) were further analyzed using IFS strategy (Zhang et al., 2016; Chen et al., 2018b, 2019; Wang et al., 2018; Pan et al., 2019a). Each time, a gene set including the top K most important genes (K = 1, 2, 3, …, 1,000) was used to train a SVM model, and its performance was evaluated with leave-one-out cross validation (LOOCV; Li and Huang, 2018). To build the SVM, we used the function svm from R package e1071.[2] Fifth, the optimal heart failure signature was the gene set with the best performance. If the IFS curve did not reach its peak or the plateau area and kept increasing as the number of genes increased, more top genes should be analyzed. Sixth, to better understand the underlying regulatory mechanisms of the signature and increase the interpretability of the signature, we constructed the signature network based on STRING database version 11.0 (http://string-db.org; Szklarczyk et al., 2018; Shi et al., 2020).

RESULTS

The Optimal Heart Failure Signature Identification

We adopted multiple feature selection strategies (**Figure 1**) to identify the optimal heart failure signature. It integrated the strategies of MCFS and IFS. Step A was data preprocessing. MCFS included Steps B and C. IFS included Steps D and E. Step F was to interpret the biological mechanisms of the signature. As demonstrated in **Figure 1D**, the actual IFS curve was shown in **Figure 2**. The highest LOOCV accuracy was 0.974 when the top 38 MCFS genes were used to train the SVM model. Therefore, these 38 genes

were chosen as the optimal heart failure signature, which was shown in **Table 1**. The confusion matrix of the 38 optimal heart failure signature genes which compared the actual class labels and precited class labels of all samples were given in **Table 2**. Their LOOCV sensitivity, specificity, and accuracy were 0.983, 0.963, and 0.974, respectively. The performance was great.
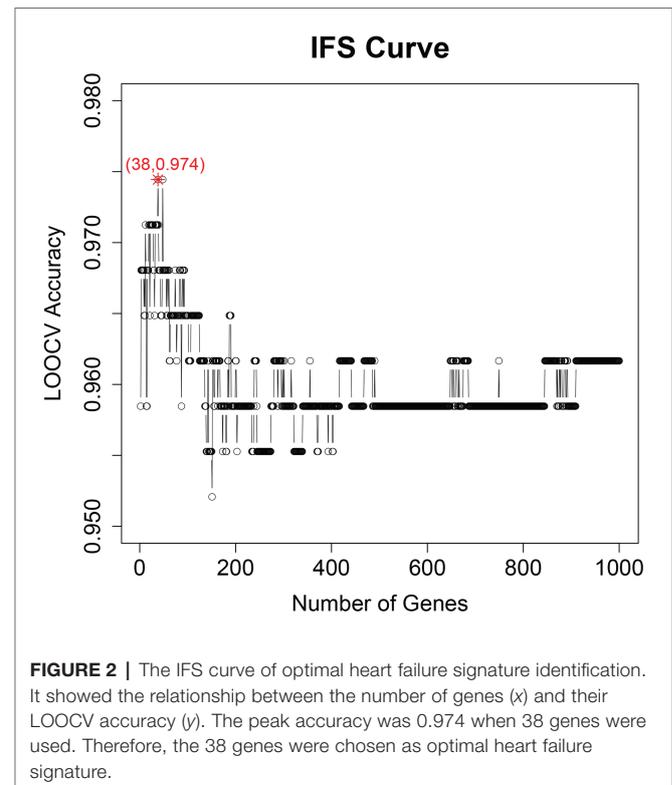
FIGURE 2 | The IFS curve of optimal heart failure signature identification. It showed the relationship between the number of genes (*x*) and their LOOCV accuracy (*y*). The peak accuracy was 0.974 when 38 genes were used. Therefore, the 38 genes were chosen as optimal heart failure signature.

---

[2]https://CRAN.R-project.org/package=e1071

FIGURE 1 | The workflow for optimal heart failure signature identification. The workflow integrated the strategies of Monte Carlo Feature Selection (MCFS) and incremental feature selection (IFS). Step **(A)**: data preprocessing. Steps **(B,C)**: MCFS. Steps **(D,E)**: IFS. Step **(F)**: signature network.
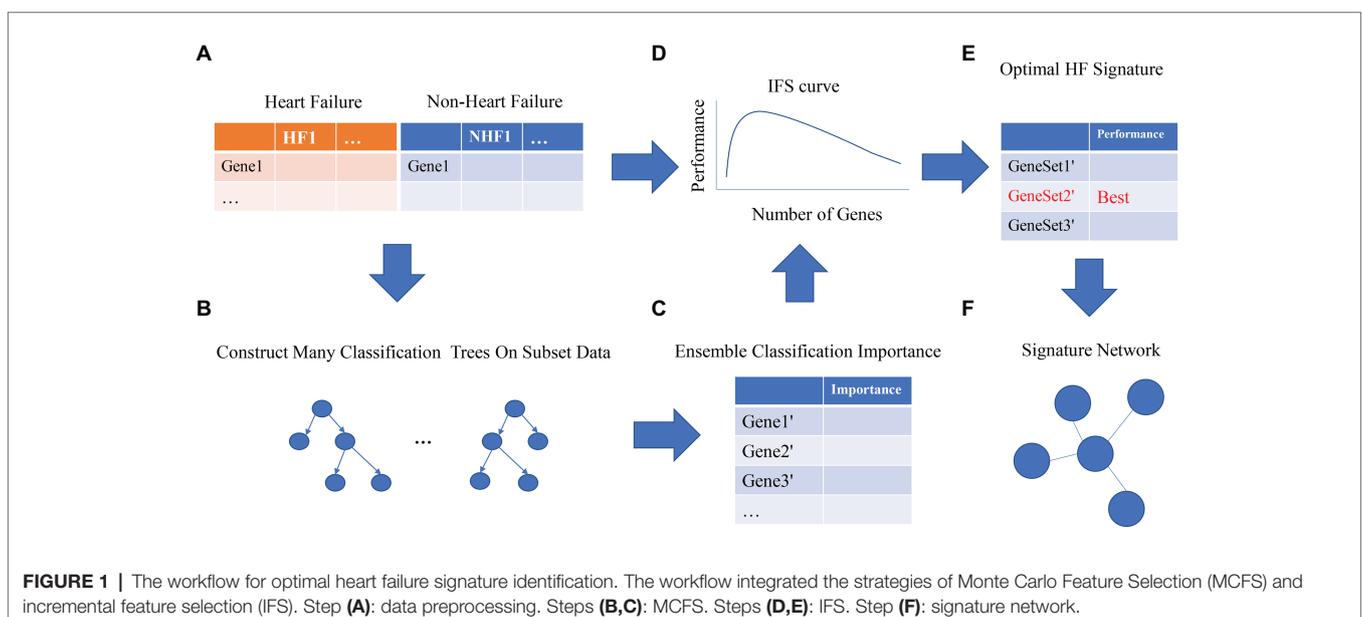
TABLE 1 | The 38 optimal heart failure (HF) signature genes.

Rank	Gene symbol	Full name	Importance
1	HMGN2	High mobility group nucleosomal binding domain 2	0.571
2	HMOX2	Heme oxygenase 2	0.527
3	SERPINA3	Serpin family A member 3	0.499
4	TUBA3D	Tubulin alpha 3d	0.489
5	ECM2	Extracellular matrix protein 2	0.481
6	FREM1	FRAS1 related extracellular matrix 1	0.461
7	FCN3	Ficolin 3	0.458
8	ZMAT1	Zinc finger matrin-type 1	0.405
9	SMOC2	SPARC related modular calcium binding 2	0.386
10	CSDC2	Cold shock domain containing C2	0.383
11	LCN6	Lipocalin 6	0.359
12	LUM	Lumican	0.356
13	FURIN	Furin, paired basic amino acid cleaving enzyme	0.349
14	LAD1	Ladinin 1	0.338
15	MNS1	Meiosis specific nuclear structural 1	0.338
16	ASPN	Asporin	0.317
17	FRZB	Frizzled related protein	0.310
18	GGT5	Gamma-glutamyltransferase 5	0.296
19	TUBA3E	Tubulin alpha 3e	0.293
20	PDE5A	Phosphodiesterase 5A	0.292
21	ISLR	Immunoglobulin superfamily containing leucine rich repeat	0.289
22	S1PR3	Sphingosine-1-phosphate receptor 3	0.279
23	SFRP4	Secreted frizzled related protein 4	0.271
24	APBB3	Amyloid beta precursor protein binding family B member 3	0.270
25	USP31	Ubiquitin specific peptidase 31	0.268
26	SLCO4A1	Solute carrier organic anion transporter family member 4A1	0.251
27	VSIG4	V-set and immunoglobulin domain containing 4	0.251
28	KCNN3	Potassium calcium-activated channel Subfamily N member 3	0.250
29	FAM58A	CCNQ cyclin Q cyclin Q	0.248
30	AP3M2	Adaptor related protein complex 3 subunit mu 2	0.247
31	C15orf59	INSYN1 inhibitory synaptic factor 1	0.243
32	BTN3A1	Butyrophilin subfamily 3 member A1	0.243
33	ZDHHC16	Zinc finger DHHC-type containing 16	0.241
34	CD163	CD163 molecule	0.238
35	SEMA4B	Semaphorin 4B	0.237
36	ST6GALNAC3	ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 3	0.228
37	TTC3	Tetratricopeptide repeat domain 3	0.228
38	MATN2	Matrilin 2	0.219

The Expression Pattern of the 38 Genes in Patients With HF and Without HF

We plotted the heatmap of the 38 genes in 177 patients with heart failure (HF) and 136 patients without heart failure [non-heart failure (NHF)] in **Figure 3**. It can be seen that most samples were clustered into the correct groups. Only very few samples were misclustered. Within the 38 genes, 17 genes (*ZMAT1*, *APBB3*, *MNS1*, *AP3M2*, *BTN3A1*, *KCNN3*, *TTC3*, *SMOC2*, *LUM*, *ASPN*, *FRZB*, *SFRP4*, *MATN2*, *ISLR*, *PDE5A*, *ECM2*, and *FREM1*) were highly expressed in HF and 20 genes (*FAM58A*, *CSDC2*, *C15orf59*, *S1PR3*, *VSIG4*, *CD163*, *SEMA4B*, *SLCO4A1*, *SERPINA3*, *GGT5*, *FURIN*,

TABLE 2 | The confusion matrix of the 38 optimal heart failure signature genes.

	Predicted HF	Predicted NHF
Actual HF	174	3
Actual NHF	5	131

ZDHHC16, *LAD1*, *USP31*, *TUBA3D*, *TUBA3E*, *ST6GALNAC3*, *LCN6*, *HMOX2*, and *FCN3*) were lowly expressed in HF.

The Network of the 38 Genes

Signature genes were not necessarily key regulators. They could be only markers. But if the signature genes have clear biological functions, they certainly can be better interpreted. Therefore, as we stated in **Figure 1F**, we searched the interaction among the STRING database (https://string-db.org/; Szklarczyk et al., 2018) and plotted the networks of the 38 genes in **Figure 4**. It can be seen that *SMOC2* is located in the hub position of the network.

SMOC2, a member of the SPARC family, which is highly expressed during embryogenesis and wound healing. Previous studies recognized that inflammatory pathways were generally dysregulated in right ventricular failure (RVF) tissue. Williams et al. (2018) analyzed mRNA datasets of human non-failing and failing heart samples from patients, and concluded that *SMOC2* was differentially expressed. *SMOC2* could be a potential significance factor that altered remodeling and inflammation for further study in the mechanism of HF. Laugier et al. (2017) found that *SMOC2*, involved in matrix remodeling, is potentially associated with the increased T-helper 1 cytokine-mediated inflammatory damage in heart, using genome-wide cardiac DNA methylation on global gene expression in myocardial samples in chronic Chagas disease cardiomyopathy, which is an inflammatory cardiomyopathy presenting with heart failure and arrhythmia.

DISCUSSION

In the present study, 38 genes were selected from our prediction model of SVM, implying strong relevance with the pathological mechanisms of HF. After literature retrieval and utilization, several evidences and analysis results have been retrieved to validate the dependability and reality of our analysis.

FCN3, a member of ficolin/opsonin p35 lectin family which consists of a collagen-like domain and a fibrinogen-like domain, which were found in all human serum. Prohászka et al. (2013) reported that the main initiator molecules of the lectin complement pathway *MBL*, *FCN2*, and *FCN3* were related to chronic heart failure (CHF). Low *FCN3* levels were related to decreased concentrations of complement factor C3 and increased complement activation product C3a (Prohászka et al., 2013). They also provided evidence for a significant association of low *FCN3* levels with advanced HF and outcome (Prohászka et al., 2013). *FCN3* is reported to be increased in microvesicles obtained

FIGURE 3 | The heatmap of the 38 genes in 177 HF and 136 non-heart failure (NHF) patients. Most samples were clustered into the correct groups. Only very few samples were misclustered. Within the 38 genes, 17 genes were highly expressed in HF, and 20 genes were lowly expressed in HF.
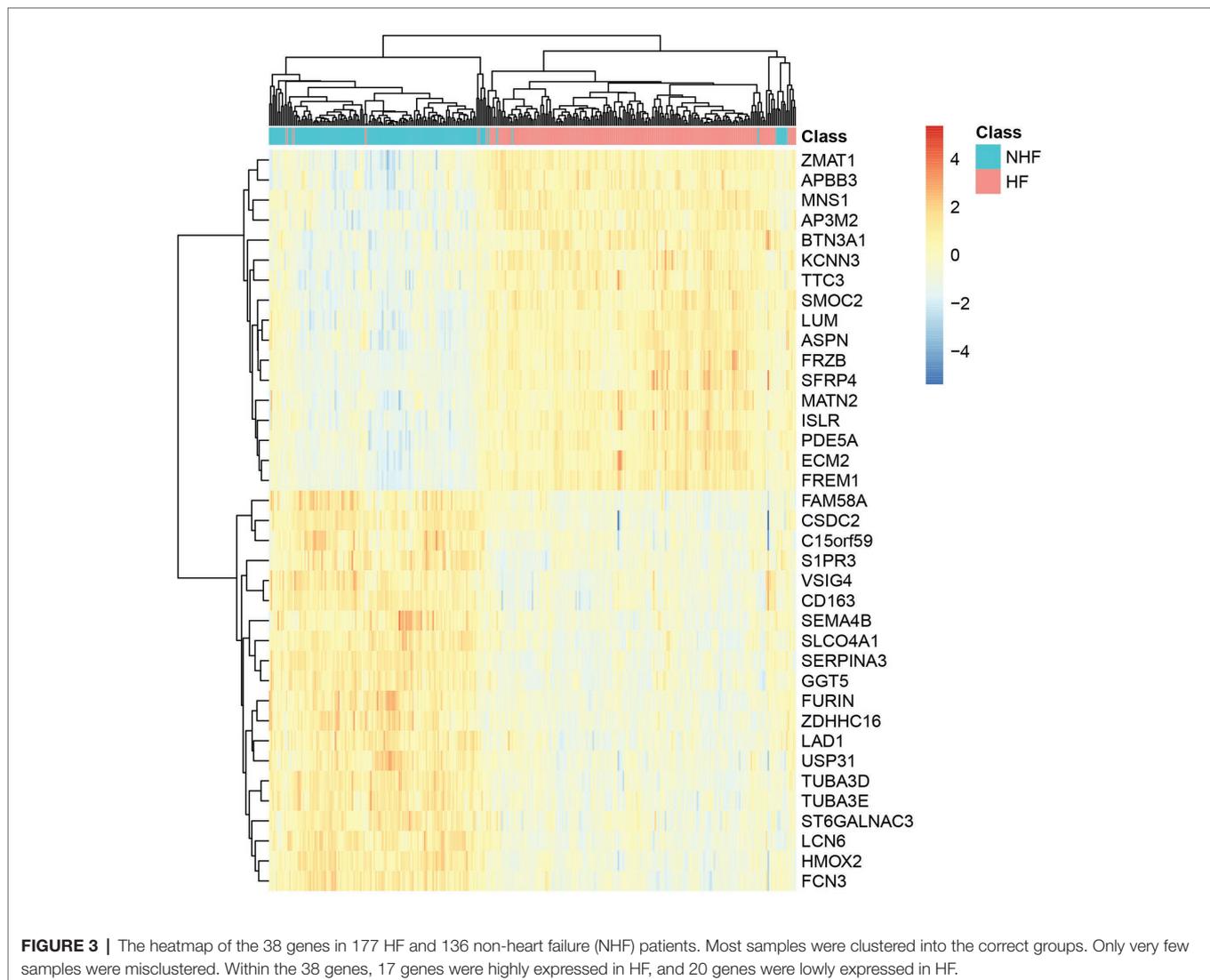
from activated platelets and abdominal aortic aneurysm (AAA) tissue (Fernandez-García et al., 2017). There is an obvious relationship between increased *FCN3* plasma levels and AAA presence and progression.

HMGN2 binds nucleosomal DNA and is associated with transcriptionally active chromatin, which is the top-ranked feature recognized by our bioinformatics analysis. HMGN protein family could regulate chromatin structure and could influence epigenetic modifications. *HMGN2* regulates active and bivalent genes by promoting an epigenetic landscape of active histone modifications at promoters and enhancers (Garza-Manero et al., 2019). *HMGN2* protected corticogenesis *via* maintaining global chromatin accessibility at promoter regions, thus ensuring proper transcriptome regulation (Apelt et al., 2020; Gao et al., 2020). There are few studies to certificate the role of HMGN2 in the progress of HF.

SERPINA3 also called Alpha-1-Antichymotrypsin or ACT, is first discovered as a plasma protease inhibitor and a member of the serine protease inhibitor (Jiang et al., 2020).

Previous study showed that *SERPINA3* emerged as a responsible cardiac secreted factor that is increased in HF patients could be the most robust and promising culprit and were related to long-term mortality. Additionally, several researches thought that mineralocorticoid receptor antagonists (MRAs) were associated to *SERPINA3* (Meijers et al., 2018). Gene expression of *SERPINA3* was significantly increased in the HF group. In circulating plasma, the level of *SERPINA3* in the HF group was confirmed significant increase by ELISA analysis. These results suggested that *SERPINA3* might play an important role in the progression of HF (Zhao et al., 2020). Asakura and Kitakaze (2009) proved that *SERPINA3* might become novel diagnostic and therapeutic targets linked to the pathophysiology of HF using seven microarray datasets previously reported.

Due to the length limitation of the article, we cannot describe all 38 selected genes in detail. After detailed literature review, we found that all the above-mentioned genes play a vital role in the progression of HF, which also verifies the reliability of

FIGURE 4 | The network of the 38 genes searched on the Search Tool for the Retrieval of Interacting Genes (STRING) database. We searched the interactions of the 38 genes on the STRING database and plotted the network. It can be seen that *SMOC2* located in the hub position of the network and its functions in HF was intensively studied.

our prediction model. We believe that these 38 selected genes are meaningful in the development of HF. They will contribute to the study of molecular mechanism, diagnosis, and treatment of HF, and will play an enlightening role in the future molecular biology research.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

DL: conception and design, administrative support, and provision of study materials or patients. HL: collection and assembly of data. HL and LL: data analysis and interpretation. All authors contributed to the article and approved the submitted version.

FUNDING

REFERENCES

Apelt, K., Zoutendijk, I., Gout, D. Y., Wondergem, A. P., van den Heuvel, D., and Luijsterburg, M. S. (2020). Human HMGN1 and HMGN2 are not required for transcription-coupled DNA repair. *Sci. Rep.* 10:4332. doi: 10.1038/s41598-020-61243-4

Asakura, M., and Kitakaze, M. (2009). Global gene expression profiling in the failing myocardium. *Circ. J.* 73, 1568–1576. doi: 10.1253/circj.cj-09-0465

Bowen, R. E. S., Graetz, T. J., Emmert, D. A., and Avidan, M. S. (2020). Statistics of heart failure and mechanical circulatory support in 2020. *Ann. Transl. Med.* 8:827. doi: 10.21037/atm-20-1127

Chen, L., Li, J., Zhang, Y. H., Feng, K., Wang, S., Zhang, Y., et al. (2018a). Identification of gene expression signatures across different types of neural stem cells with the Monte-Carlo Feature Selection method. *J. Cell. Biochem.* 119, 3394–3403. doi: 10.1002/jcb.26507

Chen, L., Pan, X., Guo, W., Gan, Z., Zhang, Y. -H., Niu, Z., et al. (2020). Investigating the gene expression profiles of cells in seven embryonic stages with machine learning algorithms. *Genomics* 112, 2524–2534. doi: 10.1016/j.ygeno.2020.02.004

Chen, L., Pan, X., Hu, X., Zhang, Y. -H., Wang, S., Huang, T., et al. (2018b). Gene expression differences among different MSI statuses in colorectal cancer. *Int. J. Cancer* 143, 1731–1740. doi: 10.1002/ijc.31554

Chen, L., Pan, X., Zeng, T., Zhang, Y., Huang, T., and Cai, Y. (2019). Identifying essential signature genes and expression rules associated with distinctive development stages of early embryonic cells. *IEEE Access* 7, 128570–128578. doi: 10.1109/ACCESS.2019.2939556

Davison, B., and Cotter, G. (2015). Why is heart failure so important in the 21st century? *Eur. J. Heart Fail.* 17, 122–124. doi: 10.1002/ejhf.219

Draminski, M., Rada-Iglesias, A., Enroth, S., Wadelius, C., Koronacki, J., and Komorowski, J. (2008). Monte Carlo Feature Selection for supervised classification. *Bioinformatics* 24, 110–117. doi: 10.1093/bioinformatics/btm486

Fernandez-García, C. E., Burillo, E., Lindholt, J. S., Martinez-Lopez, D., Pilely, K., Mazzeo, C., et al. (2017). Association of ficolin-3 with abdominal aortic

aneurysm presence and progression. *J. Thromb. Haemost.* 15, 575–585. doi: 10.1111/jth.13608

Gao, X. L., Tian, W. J., Liu, B., Wu, J., Xie, W., and Shen, Q. (2020). High-mobility group nucleosomal binding domain 2 protects against microcephaly by maintaining global chromatin accessibility during corticogenesis. *J. Biol. Chem.* 295, 468–480. doi: 10.1074/jbc.RA119.010616

Garza-Manero, S., Sindi, A. A. A., Mohan, G., Rehbini, O., Jeantet, V. H. M., Bailo, M., et al. (2019). Maintenance of active chromatin states by HMGN2 is required for stem cell identity in a pluripotent stem cell model. *Epigenetics Chromatin* 12:73. doi: 10.1186/s13072-019-0320-7

Jarcho, J. A. (2020). More evidence for SGLT2 inhibitors in heart failure. *N. Engl. J. Med.* 383, 1481–1482. doi: 10.1056/NEJMe2027915

Jiang, Z., Guo, N., and Hong, K. (2020). A three-tiered integrative analysis of transcriptional data reveals the shared pathways related to heart failure from different aetiologies. *J. Cell. Mol. Med.* 24, 9085–9096. doi: 10.1111/jcmm.15544

Kittleson, M. M., Ye, S. Q., Irizarry, R. A., Minhas, K. M., Edness, G., Conte, J. V., et al. (2004). Identification of a gene expression profile that differentiates between ischemic and nonischemic cardiomyopathy. *Circulation* 110, 3444–3451. doi: 10.1161/01.Cir.0000148178.19465.11

Laugier, L., Frade, A. F., Ferreira, F. M., Baron, M. A., Teixeira, P. C., Cabantous, S., et al. (2017). Whole-genome cardiac DNA methylation fingerprint and gene expression analysis provide new insights in the pathogenesis of chronic Chagas disease cardiomyopathy. *Clin. Infect. Dis.* 65, 1103–1111. doi: 10.1093/cid/cix506

Li, J., and Huang, T. (2018). Predicting and analyzing early wake-up associated gene expressions by integrating GWAS and eQTL studies. *Biochim. Biophys. Acta Mol. basis Dis.* 1864, 2241–2246. doi: 10.1016/j.bbadis.2017.10.036

Li, J., Lu, L., Zhang, Y. -H., Xu, Y., Liu, M., Feng, K., et al. (2020a). Identification of leukemia stem cell expression signatures through Monte Carlo Feature Selection strategy and support vector machine. *Cancer Gene Ther.* 27, 56–69. doi: 10.1038/s41417-019-0105-y

Li, J., Xu, Q., Wu, M., Huang, T., and Wang, Y. (2020b). Pan-cancer classification based on self-normalizing neural networks and feature selection. *Front. Bioeng. Biotechnol.* 8:766. doi: 10.3389/fbioe.2020.00766

Liu, Y., Morley, M., Brandimarto, J., Hannenhalli, S., Hu, Y., Ashley, E. A., et al. (2015). RNA-Seq identifies novel myocardial gene expression signatures of heart failure. *Genomics* 105, 83–89. doi: 10.1016/j.ygeno.2014.12.002

McMurray, J. J., and Pfeffer, M. A. (2005). Heart failure. *Lancet* 365, 1877–1889. doi: 10.1016/S0140-6736(05)66621-4

Meijers, W. C., Maglione, M., Bakker, S. J. L., Oberhuber, R., Kieneker, L. M., de Jong, S., et al. (2018). Heart failure stimulates tumor growth by circulating factors. *Circulation* 138, 678–691. doi: 10.1161/circulationaha.117.030816

Pan, X., Chen, L., Feng, K. Y., Hu, X. H., Zhang, Y. H., Kong, X. Y., et al. (2019a). Analysis of expression pattern of snoRNAs in different cancer types with machine learning algorithms. *Int. J. Mol. Sci.* 20:2185. doi: 10.3390/ijms20092185

Pan, X., Hu, X., Zhang, Y. -H., Chen, L., Zhu, L., Wan, S., et al. (2019b). Identification of the copy number variant biomarkers for breast cancer subtypes. *Mol. Genet. Genomics* 294, 95–110. doi: 10.1007/s00438-018-1488-4

Proházska, Z., Munthe-Fog, L., Ueland, T., Gombos, T., Yndestad, A., Förhécz, Z., et al. (2013). Association of ficolin-3 with severity and outcome of chronic heart failure. *PLoS One* 8:e60976. doi: 10.1371/journal.pone.0060976

Ramachandra, C. J. A., Hernandez-Resendiz, S., Crespo-Avilan, G. E., Lin, Y. H., and Hausenloy, D. J. (2020). Mitochondria in acute myocardial infarction and cardioprotection. *EBioMedicine* 57:102884. doi: 10.1016/j.ebiom.2020.102884

Shi, X., Shao, X., Liu, B., Lv, M., Pandey, P., Guo, C., et al. (2020). Genome-wide screening of functional long noncoding RNAs in the epicardial adipose tissues of atrial fibrillation. *Biochim. Biophys. Acta Mol. Basis Dis.* 1866:165757. doi: 10.1016/j.bbadis.2020.165757

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2018). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. doi: 10.1093/nar/gky1131

Tan, F. L., Moravec, C. S., Li, J., Apperson-Hansen, C., McCarthy, P. M., Young, J. B., et al. (2002). The gene expression fingerprint of human heart failure. *Proc. Natl. Acad. Sci. U. S. A.* 99, 11387–11392. doi: 10.1073/pnas.162370099

Tao, X., Wu, X., Huang, T., and Mu, D. (2020). Identification and analysis of dysfunctional genes and pathways in CD8+ T cells of non-small cell lung cancer based on RNA sequencing. *Front. Genet.* 11:352. doi: 10.3389/fgene.2020.00352

van Oort, R. J., Garbino, A., Wang, W., Dixit, S. S., Landstrom, A. P., Gaur, N., et al. (2011). Disrupted junctional membrane complexes and hyperactive ryanodine receptors after acute junctophilin knockdown in mice. *Circulation* 123, 979–988. doi: 10.1161/Circulationaha.110.006437

Vigil-Garcia, M., Demkes, C. J., Eding, J. E. C., Versteeg, D., de Ruiter, H., Perini, I., et al. (2020). Gene expression profiling of hypertrophic cardiomyocytes identifies new players in pathological remodeling. *Cardiovasc. Res.* cvaa233. doi: 10.1093/cvr/cvaa233 [Epub ahead of print]

Wang, S. B., and Huang, T. (2019). The early detection of asthma based on blood gene expression. *Mol. Biol. Rep.* 46, 217–223. doi: 10.1007/s11033-018-4463-6

Wang, D., Li, J. R., Zhang, Y. H., Chen, L., Huang, T., and Cai, Y. D. (2018). Identification of differentially expressed genes between original breast cancer and xenograft using machine learning algorithms. *Genes* 9:155. doi: 10.3390/genes9030155

Williams, J. L., Cavus, O., Loccoh, E. C., Adelman, S., Daugherty, J. C., Smith, S. A., et al. (2018). Defining the molecular signatures of human right heart failure. *Life Sci.* 196, 118–126. doi: 10.1016/j.lfs.2018.01.021

Yuan, F., Pan, X., Zeng, T., Zhang, Y. -H., Chen, L., Gan, Z., et al. (2020). Identifying cell-type specific genes and expression rules based on single-cell transcriptomic atlas data. *Front. Bioeng. Biotechnol.* 8:350. doi: 10.3389/fbioe.2020.00350

Zhang, Y. -H., Pan, X., Zeng, T., Chen, L., Huang, T., and Cai, Y. -D. (2020b). Identifying the RNA signatures of coronary artery disease from combined lncRNA and mRNA expression profiles. *Genomics* 112, 4945–4958. doi: 10.1016/j.ygeno.2020.09.016

Zhang, N., Wang, M., Zhang, P., and Huang, T. (2016). Classification of cancers based on copy number variation landscapes. *Biochim. Biophys. Acta* 1860, 2750–2755. doi: 10.1016/j.bbagen.2016.06.003

Zhang, S., Zeng, T., Hu, B., Zhang, Y. -H., Feng, K., Chen, L., et al. (2020a). Discriminating origin tissues of tumor cell lines by methylation signatures and dys-methylated rules. *Front. Bioeng. Biotechnol.* 8:507. doi: 10.3389/fbioe.2020.00507

Zhao, L., Guo, Z., Wang, P., Zheng, M., Yang, X., Liu, Y., et al. (2020). Proteomics of epicardial adipose tissue in patients with heart failure. *J. Cell. Mol. Med.* 24, 511–520. doi: 10.1111/jcmm.14758

Zhou, N., Chen, X., Xi, J., Ma, B., Leimena, C., Stoll, S., et al. (2020). Genomic characterization reveals novel mechanisms underlying the valosin-containing protein-mediated cardiac protection against heart failure. *Redox Biol.* 36:101662. doi: 10.1016/j.redox.2020.101662

Zhu, J., Yan, Q., Wang, J., Chen, Y., Ye, Q., Wang, Z., et al. (2020). The key genes for perineural invasion in pancreatic ductal adenocarcinoma identified with Monte-Carlo Feature Selection method. *Front. Genet.* 11:554502. doi: 10.3389/fgene.2020.554502