# Connecting intermediate phenotypes to disease using multi-omics in heart failure

Anni Moore [*1], Rasika Venkatesh[*1], Michael G. Levin[2], Scott M. Damrauer[3,4,], Nosheen Reza[2], Thomas P. Cappola[1,2], Marylyn D. Ritchie[1,4]**

*¹Genomics and Computational Biology, University of Pennsylvania Perelman School of Medicine, 3700 Hamilton Walk*
*Philadelphia, PA, 19104, USA*

*²Division of Cardiovascular Medicine, University of Pennsylvania Perelman School of Medicine, 3400 Civic Center Blvd*
*Philadelphia, PA, 19104, USA*

*³Department of Surgery, University of Pennsylvania Perelman School of Medicine, 3400 Spruce St Philadelphia, PA 19104*

*⁴Department of Genetics, University of Pennsylvania Perelman School of Medicine, 3700 Hamilton Walk Philadelphia, PA, 19104, USA*
**Email: marylyn@pennmedicine.upenn.edu**

*Authors contributed equally to the manuscript*
**Corresponding author*

Heart failure (HF) is one of the most common, complex, heterogeneous diseases in the world, with over 1-3% of the global population living with the condition. Progression of HF can be tracked via MRI measures of structural and functional changes to the heart, namely left ventricle (LV), including ejection fraction, mass, end-diastolic volume, and LV end-systolic volume. Moreover, while genome-wide association studies (GWAS) have been a useful tool to identify candidate variants involved in HF risk, they lack crucial tissue-specific and mechanistic information which can be gained from incorporating additional data modalities. This study addresses this gap by incorporating transcriptome-wide and proteome-wide association studies (TWAS and PWAS) to gain insights into genetically-regulated changes in gene expression and protein abundance in precursors to HF measured using MRI-derived cardiac measures as well as full-stage all-cause HF. We identified several gene and protein overlaps between LV ejection fraction and end-systolic volume measures. Many of the overlaps identified in MRI-derived measurements through TWAS and PWAS appear to be shared with all-cause HF. We implicate many putative pathways relevant in HF associated with these genes and proteins via gene-set enrichment and protein-protein interaction network approaches. The results of this study (1) highlight the benefit of using multi-omics to better understand genetics and (2) provide novel insights as to how changes in heart structure and function may relate to HF.

*Keywords:* Heart Failure; Cardiovascular; Multi-omics; Gene Transcription; Proteome; Network

## 1. Introduction

### 1.1. *Heart failure has a high disease burden.*

Heart failure (HF) carries one of the highest disease burdens in the world, with 1-3% of the global population estimated to be living with HF. This includes 6.7 million people in the United States (US) alone, and does not include the 33% of the US population at-risk of developing HF[1]. The etiology of HF is heterogeneous and complex, but has ultimately been defined as a clinical syndrome with symptoms and signs caused by structural and functional cardiac abnormalities[2]. Its

risk is promoted by increasing age and by the presence of comorbidities such as myocardial infarction, diabetes, hypertension, obesity, arrhythmias, infiltrative and inflammatory disorders, and exposure to drugs or environmental toxins[3–5]. Despite the complexity of HF, it has been demonstrated that risk is impacted by genetic predisposition to some degree[6]. While an exact consensus of heritability for HF has not been reached, some have estimated that the heritability of HF is around 26%[7].

The overall progression of HF can be broken up into four stages:

Stage A: risk of HF but patients have no symptoms or structural heart changes

Stage B: no symptoms of HF or asymptomatic but patients do have structural heart changes

Stage C: patients experiencing symptoms of HF

Stage D: advanced heart failure requiring specialized interventions[8,9].

As noted in Stage A and B, HF is often preceded by a phase of undetected progression, highlighting the need for better insight into the changes, such as structural heart changes[9,10]. These changes often appear specifically at the left ventricle (LV), and include decreased left ventricular ejection fraction (LVEF), LV dilation and/or hypertrophy, and valvular disease in which the heart cannot pump as effectively, losing function. LV mass (LVM) has been shown to be an independent predictor of HF, with risk for HF increasing by 1% for every 1% increase in excess LV mass[11]. Likewise, LVEF, which measures LV contractile function (the percentage of blood leaving the heart with each contraction) has been shown to be associated with HF prognosis[12,13]. Abnormal measurements of several of these parameters measuring both structural and functional changes together are reliable markers of cardiovascular risk and eventual HF diagnosis.

To quantify changes in the volume of blood in the heart before and after contraction, we can use LV end-diastolic volume (LVEDV) and LV end-systolic volume (LVESV) respectively. Together these four measures of heart structure and function (LVM, LVEF, LVEDV, and LVESV) can provide an overall characterization of progression towards potential HF and represent an intermediate phenotype or endophenotype. Identifying overlaps in changes seen in both intermediate MRI trait measures and HF could give us a better idea of vital aspects that lead towards full, advanced HF.

## 1.2. *Using multi-omics to increase knowledge gained from GWAS.*

Given the known genetic contribution to HF and the prevalence of patients with the disease, many groups have performed genome-wide association studies (GWAS) to identify genetic variants associated with HF[14–21]. While this approach allows us to gain valuable insights into potential genetic variation that contributes to the disease, it still leaves a crucial gap in connecting how these variants are actually resulting in mechanistic change, and in which specific tissues. This is especially relevant in quantitative phenotypes, where GWAS is insufficient to capture the full heterogeneity measured by the trait. Transcriptome-wide association studies (TWAS) use GWAS summary statistics along with reference gene expression from specific tissues to predict how genetic variants affect gene expression within those tissues. TWAS and PWAS also provide a boost in overall statistical power, as they are less affected by multiple test corrections due to being

a gene-based test of association[22,23]. These methods are also more portable than GWAS; they are less impacted by population structure in datasets as they operate on a gene and protein level[24].

In this study, we make use of TWAS and PWAS methods to investigate genetic-derived gene and protein changes among cardiovascular related-tissues using the largest published GWAS summary statistics of HF and MRI measures of LV structure and function to date[25–27]. Our goals are: 1) to integrate multi-omics data in the form of reference gene expression and protein expression datasets to identify novel HF and related trait associated genes, 2) to evaluate whether TWAS and PWAS approaches uncover the same association signals or provide novel gene-based associations, and 3) determine whether these genes associated with HF and related traits are part of shared pathways and/or networks between traits. This study is also, to our knowledge, one of the first times that both TWAS *and* PWAS have been performed simultaneously on quantitative traits.

## 2. Methods
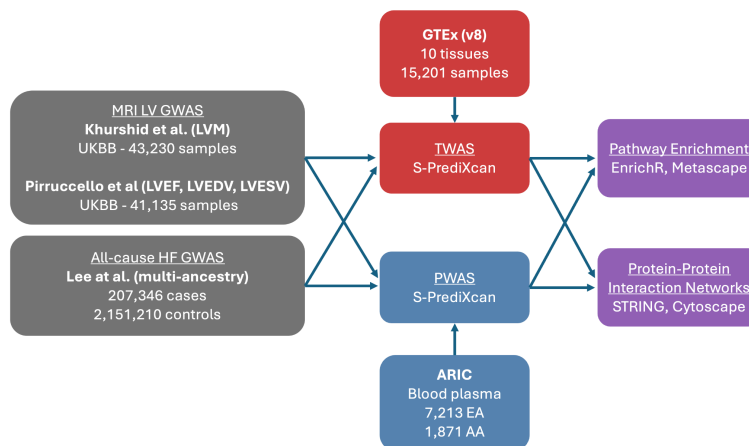


**Figure 1. Overview of study analysis to identify genes, proteins, and related interactions between MRI-derived cardiac intermediate traits and heart failure GWAS. (UKBB: UK Biobank, LVM: Left ventricular mass, LVEF: Left ventricular ejection fraction, LVEDV: Left ventricular end-diastolic volume, LVESV: Left ventricular end-systolic volume, EA: European American, AA: African American)**

### 2.1. *Cardiovascular data*

#### 2.1.1. *MRI traits*

Of all cardiac chambers, dysfunction of the left ventricle is the most common structural abnormality in HF cases. We chose four measurements taken from the left ventricle derived from MRI imaging with previously published GWAS data to characterize potential associations with HF: LVM indexed to body surface area, LVEDV, LVESV, and LVEF[26,27]. LVM measurements were taken from Khurshid et al, and includes 43,230 samples (91% European ancestry) with MRI imaging and genotype data from the UK Biobank[26]. LVEDV, LVESV, and LVEF association studies were performed on 41,135 samples also from the UK Biobank with MRI imaging and genotype data by Pirruccello et al[27].

### 2.1.2. *Heart failure (HF)*

We identified the largest all-cause HF GWAS study to date including 207,346 non-overlapping samples of cases and 2,151,210 controls meta-analyzed from HERMES, the Million Veterans Project (MVP), FinnGen, Mount Sinai BioMe (BIOME), Global Biobank Meta-analysis Initiative (GBMI), eMERGE, Geisinger DiscovEHR, and Penn Medicine BioBank (PMBB)[25]. This included an overall sample of 81.1% European ancestry, 9.7% African American, 6.5% East Asian, and 2.6% Admixed American.

## 2.2. *Transcriptome-wide association study (TWAS)*

To provide tissue-specific context to GWAS results from the selected MRI traits and HF studies we conducted transcriptome-wide association studies (TWASs) using S-PrediXcan[28] and multivariate adaptive shrinkage (MASHR) eQTL models from the Genotype-Tissue Expression (GTEx) Project v8, available in PredictDB[29,30] GTEx eQTLs were derived from a sample group of mostly European ancestry (84.6% European ancestry, 12.9% African American, 1.3% Asian and 1.1% unknown) that closely parallels the composition of the HF multi-ancestry cohort. Using this reference, we imputed genetically regulated gene expression (GReX) for all genes in ten tissues known to be relevant in the cardiovascular system and heart failure (aorta, coronary artery, tibial artery, atrial appendage, left ventricle, whole blood, visceral adipose, subcutaneous adipose, liver, and kidney)[31–37]. Associations for each of the intermediate MRI traits from the UK Biobank, as well as multi-ancestry and EUR populations from all-cause HF were calculated independently for each of these ten tissues. Significant genes were determined using a Bonferroni threshold of (p<0.05/(# genes x 10 tissues tested) per trait. All code is available at https://github.com/RitchieLab/HFmultiomics_PSB2025.

## 2.3. *Proteome-wide association study (PWAS)*

We performed a proteome-wide association study (PWAS) using S-PrediXcan[28] with the GWAS summary statistics for MRI traits from UK Biobank and for the multi-ancestry and European (EUR) population all-cause HF studies. PWAS identifies genetic associations that may influence complex traits, such as all-cause HF and MRI traits, by regulating protein abundance in tissue[38]. Blood plasma-derived protein quantitative trait loci (cis-pQTLs) from the Atherosclerosis Risk in Communities (ARIC)[39] study were used to construct the models. This large bi-ethnic study was made up of 9,084 participants, consisting of 7,213 European Americans (EA) and 1,871 African Americans (AA). S-PrediXCan PWAS EA and AA models were identified in PredictDB and were constructed using ARIC consortium data by utilizing PEERS covariates, expression information from eQTL associations, gene and SNP annotations[40,41]. PWAS was conducted on multi-ancestry and EUR studies of all-cause HF, as well as on traits from UK Biobank with the intermediate MRI traits using the EA cohort information, and additionally the AA cohort for the multi-ancestry HF study. The resulting PWAS associations were assessed for statistical significance using a Bonferroni significance threshold (p<0.05/# proteins tested) for each trait.

## 2.4. *Network and pathway analyses*

### 2.4.1. *Pathway enrichment analysis*

Gene set enrichment was performed using EnrichR[42,43] for the significant results from TWAS and PWAS for each MRI trait and HF phenotype, respectively. Enrichment analysis explored the specific pathways and processes associated with the statistically significant genes and proteins from the TWAS and/or PWAS. Pathway results were annotated with KEGG 2021, Reactome 2022, and Gene Ontology (GO) Biological Process 2023 pathways. The significant pathways were identified as having Fisher's exact test p-value < $0.05^{38,44}$.

### 2.4.2. *Network analysis and identification of hub genes and proteins*

The statistically significant genes and proteins identified via TWAS and PWAS were used to construct a protein-protein interaction (PPI) network using the online Search Tool for the Retrieval of Interacting Genes (STRING v11)[45], where the number of interactions present was assessed for significance. Network interactions were thresholded by a minimum confidence score of > 0.4, as calculated by STRING[45]. The networks were then visualized using Cytoscape 3.10.2[46], and degree centrality analysis was performed using the cytoHubba module to identify and visualize the hub genes and proteins[47,48].

### 2.4.3. *Classification of of sub-clusters*

Additionally, the Molecular Complex Detection (MCODE)[49] module in Cytoscape was used to screen modules of the larger PPI networks and construct clusters by identifying densely-connected regions of the network[50]. The networks were thresholded to have an MCODE degree cutoff of 3, node density cutoff of 0.1, node score cutoff of 0.2, number of nodes > $3^{47,51}$. Gene set enrichment analysis using KEGG 2021[52–54], Reactome 2022[55,56], and Gene Ontology (GO) Biological Process 2023[57,58] of each cluster was then conducted using Metascape[59], using the default parameters of minimum overlap of 3, p-value cutoff of 0.01, and minimum enrichment score of 1.5.

## 3. Results

## 3.1. *TWAS and PWAS Association Analyses*

### 3.1.1 *MRI trait gene and protein associations*

After fine-mapping, 35 unique genes within ten tissues and three proteins from blood plasma (SPON1, C2, PACAP) were significant for LVEF based on a Bonferroni threshold (TWAS:p<3.814E-07, PWAS:3.75E-05) (**Figure 2A**). 16 of these genes were significant in three or more tissues, and one gene, *SPON1*, replicated in both TWAS and PWAS for LVEF. 35 genes and one protein (THBS4) appeared significantly associated with LVM measures (TWAS:p<3.842E-07, PWAS:3.79E-05). Five of these significant genes (*FKBP7, WNT3, HSPQ4, PSMC3,* and *PRKRA*) appeared in three or more tissues tested. Finally, amongst the ten tissues tested, 33 genes and three proteins (ENG, QPCTL, SPON1) were significant for LVEDV (TWAS:p<3.815E-07, PWAS:3.79E-05) and 48 genes along with four proteins (RAB5A, SRL, PACAP, SPON1) for LVESV (TWAS:p<3.815E-07), protein:3.79E-05). *SPON1* was also

significantly associated with LVESV for both TWAS and PWAS. Figures for LVM, LVEDV, LVESV are available in **Supplemental Figure 1A-C**. The full significant results of the TWAS and PWAS for MRI traits are available in **Supplemental Table 1 and 2**.

Between MRI traits, several genes appeared significant between measures. Genes *FKBP7*, *PRKRA*, and *RP11-171I2.3* were associated with all four MRI-based traits in at least one tissue. 15 genes overlapped between LVEDV and LVESV, four genes were shared between LVEDV and LVM, four between LVEDV and LVM, five genes between LVESV and LMV, and finally 30 genes between LVESV and LVEF. Amongst protein results, SPON1 was the only protein shared between traits and was significantly associated with LVEF, LVEDV, and LVESV.

## MRI – LV Ejection Fraction

**A**
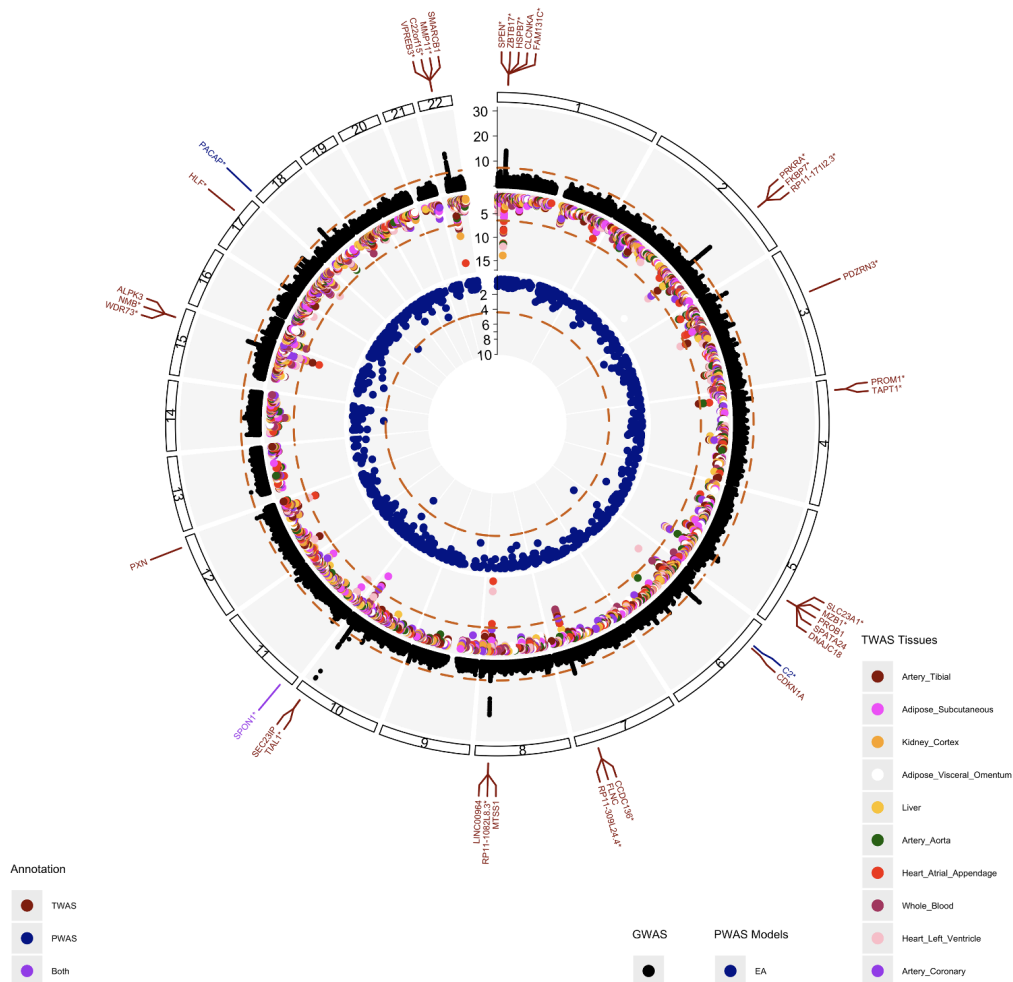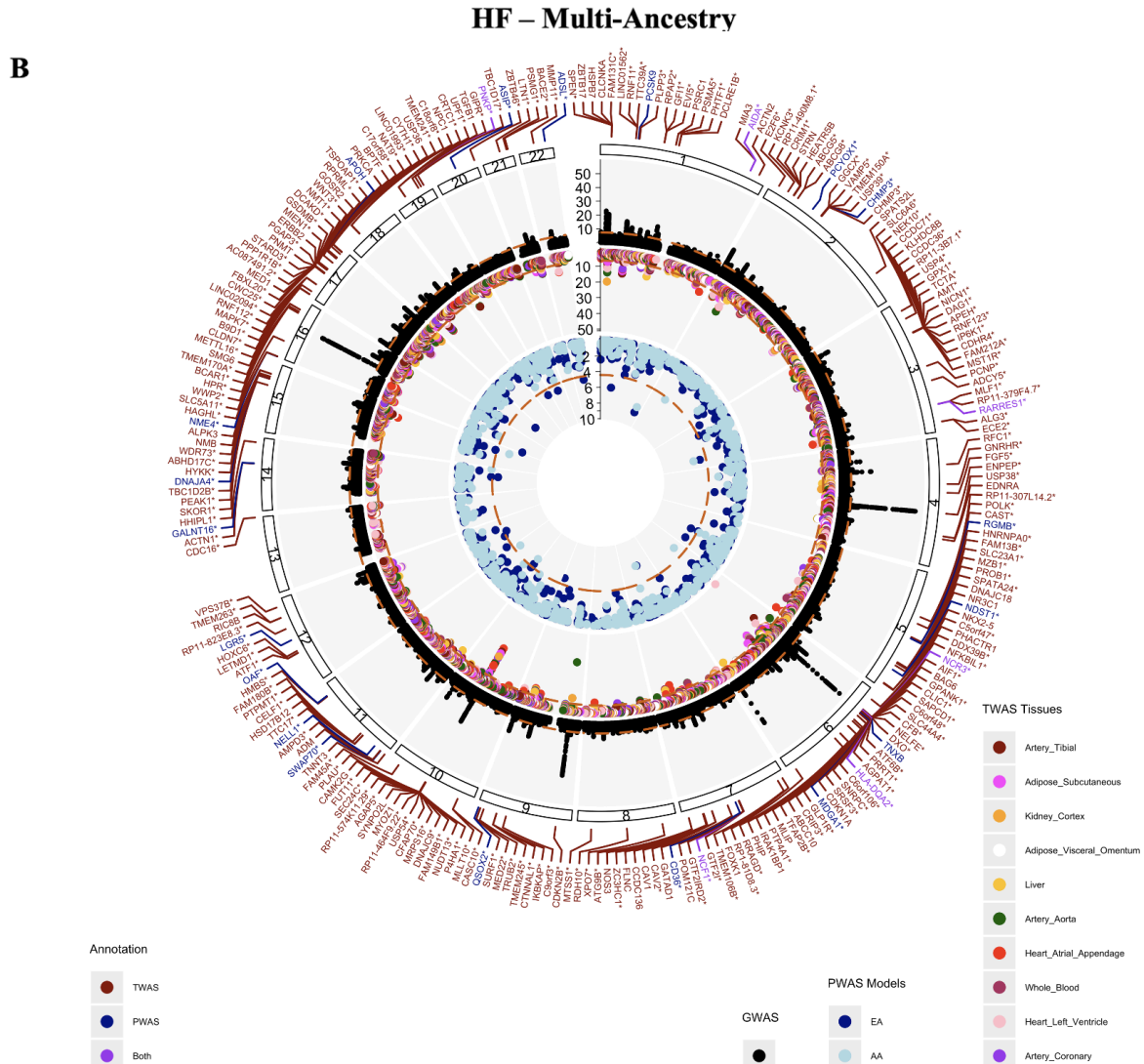
**Figure 2. Circos plots for A) MRI LV Ejection Fraction and B) HF - Multi-ancestry representing identified associations through GWAS (black), TWAS (red), and PWAS (blue). The outermost track of annotations depicts genes and proteins identified through TWAS (red) and PWAS (blue), or both analyses (purple), with the asterisk denoting novel results not previously reported from the source GWAS, GWAS Catalog, or NCBI.**

### 3.1.2  *HF gene and protein associations*

231 unique genes and 29 proteins significantly associated with HF in a multi-ancestry population (TWAS:p<3.806E-07, PWAS:p<3.79E-05) (**Figure 2B**). Six genes replicated across genes and proteins, including *RARRES1*, *NCF1*, *AIDA*, *HLA-DQA2*, *PNKP*, and *NCR3*. 185 of the total 231 associating genes were significant in at least one vascular tissue (heart atrial appendage, heart left ventricle, tibial artery, coronary artery, whole blood). 90 genes were significant in at least one vascular tissue and one peripheral tissue (liver, kidney, subcutaneous adipose, visceral adipose). Of the ten tissues tested, heart atrial appendage and heart left ventricle tissues had the largest number of genes significantly associating with HF. Genes *CRIP3* and *USP54* were significant in

all ten tissues tested. Similar associations with HF were noted in the European population (**Supplemental Figure 1D).**

## 3.2. *Network and Pathway analyses*

### 3.2.1. *MRI trait gene-set enrichment*

In order to identify the known biologically relevant pathways associated with the statistically significant TWAS and PWAS genes for each phenotype, gene-set enrichment analysis was performed using EnrichR for Reactome 2022, KEGG 2021, and Gene ontology (GO) 2023 pathways. Significant pathways were identified at a p-value < 0.05; the full set of significant pathways for each phenotype are available in **Supplemental Table 3**. For LVEF, the most significant pathway by p-value was positive regulation of actin filament bundle assembly (p = 4.23E-03), made up of genes *PXN* and *MTSS1* (**Figure 3A**). Several pathways involved in kidney development function were also identified to be significant, such as renal cell filtration differentiation (p-value = 9.47E-03), and nephron tubule development (p = 9.47E-03).

LVM (**Supplemental Figure 2A**) was enriched for a variety of cell proliferation and differentiation pathways such as positive regulation of endothelial cell proliferation (p = 1.11E-03) and negative regulation of muscle cell differentiation (p = 4.54E-03), identifying the gene *IGF2* and protein THBS4 as important. Similarly developmentally important cardiovascular pathways were identified for the LVEDV **(Supplemental Figure 2B)**, including embryonic hemopoiesis (p = 2.43E-04), megakaryocyte differentiation (p = 4.211E-04), and cardiac atrium morphogenesis (p = 4.73E-04). The most significant pathway for LVESV (**Supplemental Figure 2C**) was modulation by host of symbiont process (p = 5.34E-03), in addition to developmental pathways - glomerular epithelial cell differentiation (p = 1.29E-02), and renal filtration cell differentiation (p = 1.29E-02).
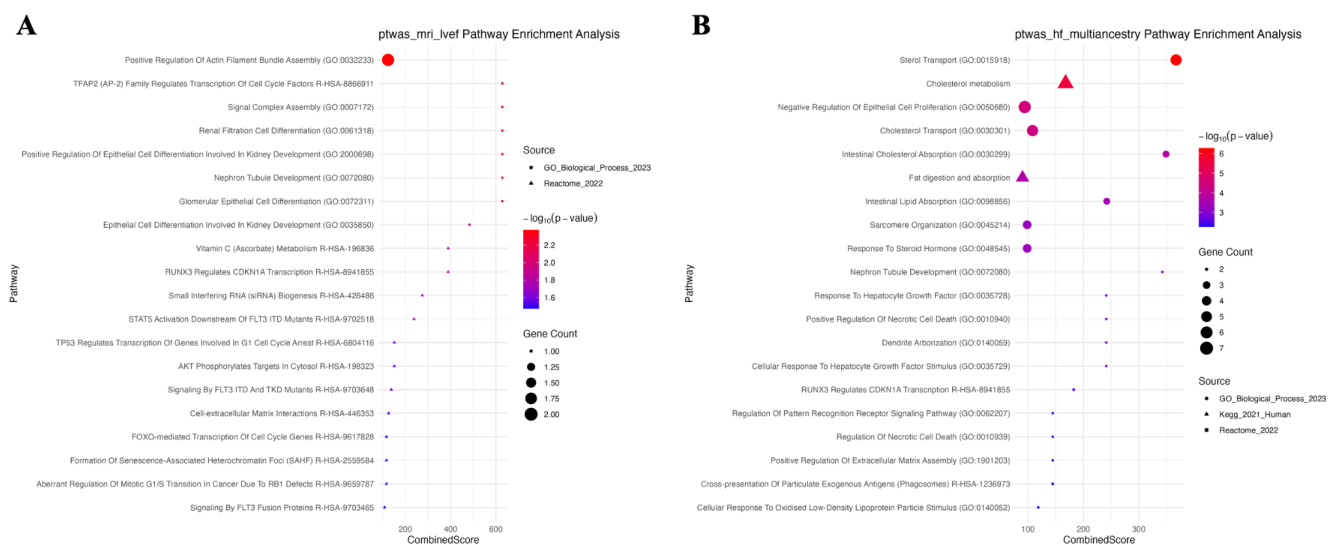


**Figure 3. Gene-set enrichment results of TWAS and PWAS significant hits for A) MRI LVEF and B) multi-ancestry HF phenotypes.**

### 3.2.2.  *HF gene-set enrichment*

A variety of relevant gene-sets were found to have overrepresented pathways previously identified as important in all-cause HF[60]. For the HF multi-ancestry cohort (**Figure 3B**), the most significant pathways include sterol transport (p = 5.34E-07) and cholesterol metabolism (p = 2.84E-06), which are known to be impacted in a variety of cardiovascular disease states, including heart failure[61,62].  Similar pathways were enriched in the EUR population for HF (**Supplemental Figure 2D**). Genes implicated in these pathways include *ABCG8, STARD3, ABCG5, NPC1, CAV1, APOH, PCSK9,* and *CD36*.

### 3.2.3.  *PPI network analysis of MRI trait genes and proteins*

To evaluate the association of candidate genes and proteins identified by TWAS and PWAS, PPI networks were constructed for each MRI trait phenotype using the STRING database. The PPI network for LVEF contained 32 nodes, 16 of which were connected, and 14 edges at a confidence threshold of > 0.4, with a PPI enrichment p-value = 5.24E-07, indicating that there were significantly more interactions observed than expected by random chance. The hub nodes identified via the cytohubba plugin by degree centrality were *FLNC, ALPK3, SPATA24*, and *HSPB7* (**Figure 4A**). MCODE identified 1 cluster in the network, with nodes *FLNC, HSPB7*, and *ALPK3*, at a score of 1.5, as computed by multiplying node density by the number of members.

The network for LVM (**Supplemental Figure 3A**) contained 10 nodes (3 connected) and 2 edges, all corresponding to hits from TWAS. The network had a PPI enrichment score of p = 0.303. The hub node identified was FKBP7, a protein which functions as a molecular chaperone to accelerate protein folding. MCODE did not identify clusters in this network.

The network for LVESV (**Supplemental Figure 3B**) consisted of 45 nodes (19 connected) and 16 edges, with a PPI enrichment p-value = 3.14E-06. Similarly to LVEF, the genes *FLNC, ALPK3*, *HSPB7*, and *BHMG1* were identified as hub nodes. MCODE once again identified 1 cluster in the network, with nodes *FLNC, HSPB7*, and *ALPK3*, at a score of 3.

The LVEDV network (**Supplemental Figure 3C**) contained 34 nodes (13 connected) and 11 edges. Cytohubba identified BHMG1 as a hub node. MCODE did not identify clusters in this network. Metascape pathway enrichment of clusters for the MRI phenotypes did not yield additional enriched terms; the full cluster and pathway enrichment results are available in **Supplemental Table 4**.
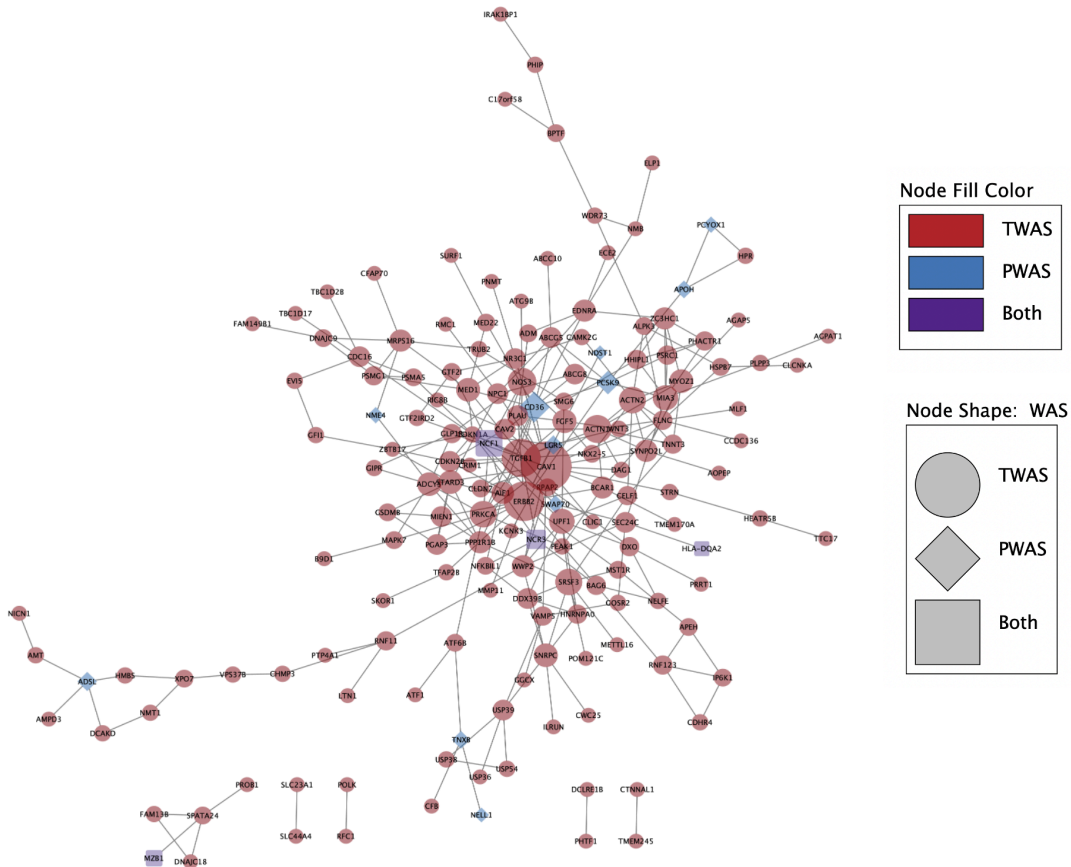
**MRI – LV Ejection Fraction PPI Network**

**A**
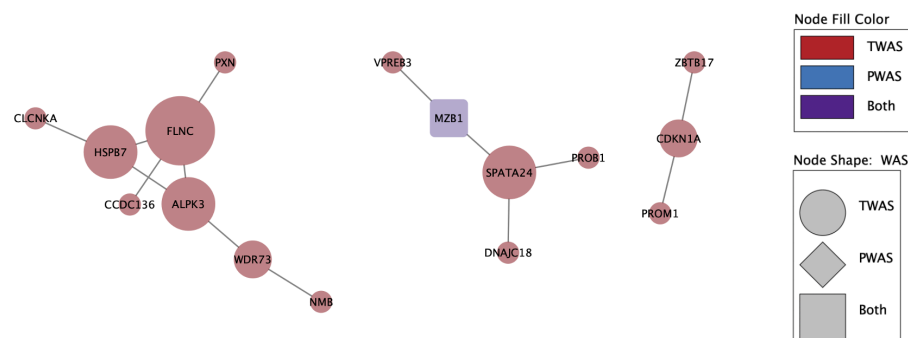
# HF – Multi-Ancestry PPI Network



**Figure 4. PPI networks constructed using TWAS and PWAS significant hits for A) MRI LVEF and B) multi-ancestry HF phenotypes. Size of nodes denotes degree centrality, with largest nodes identified as hub nodes.**

### 3.2.4. *PPI network analysis of HF genes and proteins*

The PPI network for the HF multi-ancestry cohort consisted of 234 nodes and 294 edges at the medium confidence score > 0.4 in STRING, with an average local clustering coefficient of 0.377 (**Figure 4B**). STRING found that the network had a PPI enrichment p-value = 1.0E-16, suggesting that the network had significantly more edges than expected by random chance. In Cytoscape, the hub nodes identified by degree centrality were *CAV1*, *ERBB2*, *TGFB1*, and *CD36*. Additionally, 5 non-overlapping clusters were identified in MCODE, with a max cluster score of 4.889 and minimum score of 3.0, where a higher score denotes a greater number of nodes in the cluster. Metascape was used to evaluate pathways for the genes present in each cluster, identifying pathways in sarcomere organization ($LOG_{10}P=-9.09$), Hemostasis ($LOG_{10}P=-7.84$), VEGFA-VEGFR2 pathway ($LOG_{10}P=-7.65$), and heart development ($LOG_{10}P=-3.14$), among several others. The PPI network for HF in the EUR population (**Supplemental Figure 3D**) yielded similar findings.

## 4. Discussion

We performed the first ever intermediate cardiac imaging trait TWAS at the gene level and first-ever protein imputation and largest for all-cause HF, followed by enriched gene sets and constructed interaction networks to contextualize our findings. Current established cardiomyopathy (CM) genes have been found studying familial forms of disease, however, here we focused on identifying associations based on structural and functional cardiac changes and expect to see differences. Overall we demonstrated (1) an increase in information gain using TWAS and PWAS in addition to GWAS in connecting intermediate cardiac phenotypes and full-stage HF, (2) the value in using quantitative intermediate measures for interpreting the changes that occur during disease progression, and (3) how these intermediate measures may relate to changes seen in full-stage HF.

### 4.1.    *Information gain from multi-omics approaches*

While GWAS has value in identifying singular variants associated with traits, variant results are unable to give clues toward mechanistic effects as it relates to the traits of interest. Genes deemed significant from GWAS are usually only identified based on their proximity to significant variants, but we know that proximity actually is not always a good indicator of effect on gene expression and often varies per tissue[63,64]. However, our results are derived from published variant to gene expression values (eQTLs). Our TWAS showed an overall increase in the number of genes associated for every MRI-trait tested compared to GWAS. For example, from Pirruccello et al., 19 genes were identified as proximal to significant variants from GWAS for LVEF. We replicated 8 of those in our TWAS and PWAS, and also identified 29 additional genes and proteins (4 within GWAS Catalog[65] and NCBI[66], 25 novel) that were significant in a tissue specific context (**Figure 2A**)[27]. This trend continued for LVEDV, LVESV, and LMV from Khurshid et al (**Supplemental Figures 1A-C**)[26]. For all-cause multi-ancestry HF GWAS, we replicated 50 genes from the source GWAS in our TWAS results. 199 genes and proteins were not identified in GWAS, indicating that GWAS variants may not be affecting closest proximity genes. Of the remaining 199 genes and proteins identified in our TWAS, 179 are considered novel associations and are not found in previous HF GWAS within GWAS Catalog (**Figure 2B**).

Additionally, to our knowledge this is the first study looking at the protein level based on quantitative cardiac traits, making all protein results novel. By adding proteins to genes in our gene-protein networks, these results 1) give higher confidence associations to those replicating at the gene and protein level, and 2) connect gene clusters with missing information. For example, SPON1 replicated at the gene and protein level for LVEF and LVESV and has been implicated in cardiomyopathies[67–69], blood pressure[70], cognitive decline[71–73], cancers[74,75], as well as EF in a recent study[76]. NCF1, replicated at the gene and protein level for HF and does not appear to have previously been associated with HF in literature, but has been reported for Williams Syndrome which can cause structural vascular changes[77]. NCF1 also has ties to immune response and is a key regulator of reactive oxygen species[78,79]. Other replicating proteins for HF (APOH, TNXB, PCSK9, and RGMB), have previously been associated with HF[16,80–82].

### 4.2.    *Relationships between intermediate cardiac measures and HF*

The four MRI derived cardiac measures explored in this study are often collectively used to mark structural and functional changes in the heart, and are reliable indicators of HF risk and eventual

diagnosis. Therefore the relationship between these measures as well as the overlap between each and HF at the gene, protein, and pathway level is of interest.

Between the MRI cardiac measures we see overlap at the gene and protein level (**Figure 2A** and **Supplemental Figures 1A-C**). Three genes, *FKBP7*, *RP11-171I2.3*, and *PRKRA* were significant in at least one tissue for each of the four MRI cardiac traits. Both *FKBP7* and *PRKRA* have been discussed in relation to cardiac traits before. *PRKRA* has been implicated in studies impacting heart structure, and while *FKBP7* has been less discussed, was linked to atrial fibrillation in one study[83–87]. *PRKRA* and *FKBP7* are both involved in stress response and could be indicative of problematic changes beginning to occur. It is worth noting that *FKBP7*, *PRKRA*, and third gene (*PLEKHA3,* significant with LVM, LVEDV, and LVESV), all lie within a region of chromosome 2 that also includes the gene *TTN*, which is well established heritable cause of dilated cardiomyopathy, a leading cause of heart failure[88–90].

Of the four traits LVEF and LVESV had the most overall overlaps, with 30 different genes and one protein, SPON1 appearing for both traits. When we clustered the results, one PPI cluster for LVEF is dominated by genes previously implicated with dilated cardiomyopathy and heart failure risk (*HSPB7*, *FLNC*, *ALPK3*, *CLCNKA*) **(Figure 4A)**, as well as links to the brain via *WDR73*[91–97]. We also see a cluster composed of much of the same genes for LVESV, and overlapping enriched pathways for cardiac cell development (*ALPK3*), and transepithelial chloride transport (*CLCNKA*). Renal transport also appears as a significantly enriched pathway for LVEF as a result of *CLCNKA,* as well as renal filtration cell differentiation and nephron tubule as top pathway results. Renal impairment is common among HF patients and is documented to increase mortality risk[98,99]. The significance of structural LV genes contributing to these pathways could indicate that these genes impact both disorders, or that there may be some interplay between HF and renal conditions that contribute to progression.

Within our multi-ancestry HF results there appears to be substantial overlap in the MRI derived cardiac measures. 30 genes that appeared significant in one of the MRI cardiac traits also appeared to be associated with all-cause HF. LVEF and LVESV had the largest number of overlapping genes with the disease respectively (18 genes for LVEF and 18 genes for LVESV), further supporting their use as HF indicators. AIDA was the top significant protein from blood plasma, significant with both the European American and African American pQTL reference panels (**Supplemental Table 2**), a gene previously discussed as part of inflammatory response that also promotes atherosclerosis and coronary artery disease[100]. Overall, HF associated gene and protein results tend to enrich lipid and specifically cholesterol related pathways (**Figure 3B**). Nephron tubule development also appears as a top pathway for GF, implicating the kidneys again. We did not see top CM genes such as *MYBPC3*, *MYH7*, *MYL2/3*, *DSP*, but these also were not significant in our source GWASs and not surprising given that our studies were focused on cardiac structure and functional changes rather than familial types. However, we did find some known CM genes: *FLNC* (LVEF, LVESV, HF), *PLN* (LVEDV), and *ACTIN2* (HF), as well as genes related to known CM genes like *TNNT2* (*TNNT3* - HF) and *TMEM43* (*TMEM170A*, *TMEM241*, *TMEM150A*, *TMEM245* - HF).

### 4.3. *Limitations and future directions*

Limitations exist within this study. First, this analysis only encompasses the genetic factors of HF and intermediate cardiac measures. Genetics play a significant role in the development of HF;

however, it has been shown that environment and comorbidities play a large role in increasing risk as well[101,102]. We did not consider social factors or health records of the patients used in the source GWASs, and therefore may not have fully been able to characterize all potential sources of disease progression. Future studies attempting to characterize or stratify individual-level risk of developing HF would benefit from including these data modalities, in addition to those used in this study.

Additionally, we used a GWAS study with an all-case HF multi-ancestry cohort to represent HF in our study. While this is beneficial for increasing sample size and increasing power, this might also muddle distinct signals within known phenotypic subgroups of HF or specific ancestries given the complexity of the disease as earlier stated. Future studies may find improvement by stratifying HF by subgroups, such as diastolic vs systolic dysfunction, to achieve more informed results. Beyond our cohort limitations, we also acknowledge that the imputation of gene and protein expression is influenced by the ancestry and completeness of the reference eQTL and pQTL sets we used (GTEx v8 and ARIC). For example, known cardiomyopathy genes *LMNA*, *FHL1*, and *CALR3* were not in the GTEx eQTL set used[25]. These QTL reference sets were also not disease specific. The creation of disease specific models using disease-specific RNAseq and proteomic data may be an avenue for further improvement. Lastly, here we only considered imputed gene and protein data modalities, based on multi-omics data from well characterized reference populations. The use of additional modalities, such as RNAseq, protein abundance, or methylation information would provide stronger evidence for our conclusions.

## 5. Acknowledgments

## 6. Appendix

Supplemental figures and tables are available at:
https://ritchielab.org/publications/supplementary-data/psb-2025/hfmultiomics
Code is available at: https://github.com/RitchieLab/HFmultiomics_PSB2025

## References

1. Bozkurt, B. *et al.* Heart Failure Epidemiology and Outcomes Statistics: A Report of the Heart Failure Society of America. *J. Card. Fail.* **29**, 1412–1451 (2023).
2. Bozkurt, B., Coats, A. & Tsutsui, H. Universal definition and classification of heart failure. *J. Card. Fail.* **27**, 387–413 (2021).
3. Ziaeian, B. & Fonarow, G. C. Epidemiology and aetiology of heart failure. *Nat. Rev. Cardiol.* **13**, 368–378 (2016).
4. Kenchaiah, S. *et al.* Obesity and the risk of heart failure. *N. Engl. J. Med.* **347**, 305–313 (2002).
5. Roger, V. L. Epidemiology of Heart Failure: A Contemporary Perspective. *Circ. Res.* **128**, 1421–1434 (2021).
6. Lopes, L. R. & Elliott, P. M. Genetics of heart failure. *Biochim. Biophys. Acta* **1832**, 2451–2461 (2013).

7. Lindgren, M. P. *et al.* A Swedish Nationwide Adoption Study of the Heritability of Heart Failure. *JAMA Cardiol* **3**, 703–710 (2018).

8. Ammar, K. A. *et al.* Prevalence and prognostic significance of heart failure stages: application of the American College of Cardiology/American Heart Association heart failure staging criteria in the community. *Circulation* **115**, 1563–1570 (2007).

9. Goldberg, L. R. & Jessup, M. Stage B heart failure: management of asymptomatic left ventricular systolic dysfunction. *Circulation* **113**, 2851–2860 (2006).

10. Kosmala, W. & Marwick, T. H. Asymptomatic Left Ventricular Diastolic Dysfunction: Predicting Progression to Symptomatic Heart Failure. *JACC Cardiovasc. Imaging* **13**, 215–227 (2020).

11. de Simone, G., Gottdiener, J. S., Chinali, M. & Maurer, M. S. Left ventricular mass predicts heart failure not related to previous myocardial infarction: the Cardiovascular Health Study. *Eur. Heart J.* **29**, 741–747 (2008).

12. Aurigemma, G. P., Zile, M. R. & Gaasch, W. H. Contractile behavior of the left ventricle in diastolic heart failure: with emphasis on regional systolic function. *Circulation* **113**, 296–304 (2006).

13. Solomon, S. D. *et al.* Influence of ejection fraction on cardiovascular outcomes in a broad spectrum of heart failure patients. *Circulation* **112**, 3738–3744 (2005).

14. Smith, J. G. *et al.* Discovery of Genetic Variation on Chromosome 5q22 Associated with Mortality in Heart Failure. *PLoS Genet.* **12**, e1006034 (2016).

15. Villard, E. *et al.* A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. *Eur. Heart J.* **32**, 1065–1076 (2011).

16. Rasooly, D. *et al.* Genome-wide association analysis and Mendelian randomization proteomics identify drug targets for heart failure. *Nat. Commun.* **14**, 3826 (2023).

17. Levin, M. G. *et al.* Genome-wide association and multi-trait analyses characterize the common genetic architecture of heart failure. *Nat. Commun.* **13**, 6914 (2022).

18. Yu, B. *et al.* Genome-wide association study of a heart failure related metabolomic profile among African Americans in the Atherosclerosis Risk in Communities (ARIC) study. *Genet. Epidemiol.* **37**, 840–845 (2013).

19. Shah, S. *et al.* Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat. Commun.* **11**, 163 (2020).

20. Smith, N. L. *et al.* Association of genome-wide variation with the risk of incident heart failure in adults of European and African ancestry: a prospective meta-analysis from the cohorts for heart and aging research in genomic epidemiology (CHARGE) consortium. *Circ. Cardiovasc. Genet.* **3**, 256–266 (2010).

21. Morrison, A. C. *et al.* Genomic variation associated with mortality among adults of European and African ancestry with heart failure: the cohorts for heart and aging research in genomic epidemiology consortium. *Circ. Cardiovasc. Genet.* **3**, 248–255 (2010).

22. Li, B. & Ritchie, M. D. From GWAS to Gene: Transcriptome-Wide Association Studies and Other Methods to Functionally Understand GWAS Discoveries. *Front. Genet.* **12**, 713230 (2021).

23. Cao, C. *et al.* Power analysis of transcriptome-wide association study: Implications for practical protocol choice. *PLoS Genet.* **17**, e1009405 (2021).

24. Li, D., Liu, Q. & Schnable, P. S. TWAS results are complementary to and less affected by linkage disequilibrium than GWAS. *Plant Physiol.* **186**, 1800–1811 (2021).

25. Lee, D. S. M. *et al.* Common- and rare-variant genetic architecture of heart failure across the allele frequency spectrum. *medRxiv* (2023) doi:10.1101/2023.07.16.23292724.
26. Khurshid, S. *et al.* Clinical and genetic associations of deep learning-derived cardiac magnetic resonance-based left ventricular mass. *Nat. Commun.* **14**, 1558 (2023).
27. Pirruccello, J. P. *et al.* Genetic analysis of right heart structure and function in 40,000 people. *Nat. Genet.* **54**, 792–803 (2022).
28. Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9**, 1825 (2018).
29. Barbeira, A. N. *et al.* Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biol.* **22**, 49 (2021).
30. Araujo, D. S. *et al.* Multivariate adaptive shrinkage improves cross-population transcriptome prediction and association studies in underrepresented populations. *HGG Adv* **4**, 100216 (2023).
31. Jahng, J. W. S., Song, E. & Sweeney, G. Crosstalk between the heart and peripheral organs in heart failure. *Exp. Mol. Med.* **48**, e217 (2016).
32. Oikonomou, E. K. & Antoniades, C. The role of adipose tissue in cardiovascular health and disease. *Nat. Rev. Cardiol.* **16**, 83–99 (2019).
33. Schefold, J. C., Filippatos, G., Hasenfuss, G., Anker, S. D. & von Haehling, S. Heart failure and kidney dysfunction: epidemiology, mechanisms and management. *Nat. Rev. Nephrol.* **12**, 610–623 (2016).
34. Poelzl, G. *et al.* Liver dysfunction in chronic heart failure: prevalence, characteristics and prognostic significance. *Eur. J. Clin. Invest.* **42**, 153–163 (2012).
35. Xanthopoulos, A., Starling, R. C., Kitai, T. & Triposkiadis, F. Heart Failure and Liver Disease: Cardiohepatic Interactions. *JACC Heart Fail* **7**, 87–97 (2019).
36. Tanai, E. & Frantz, S. Pathophysiology of Heart Failure. *Compr. Physiol.* **6**, 187–214 (2015).
37. Vanburen, P. *et al.* Blood gene expression signatures associate with heart failure outcomes. *Physiol. Genomics* **43**, 392–397 (2011).
38. Li, S.-J. *et al.* Identifying causal genes for migraine by integrating the proteome and transcriptome. *J. Headache Pain* **24**, 1–11 (2023).
39. Wang, J.-J., Zheng, Z. & Zhang, Y. Association of Hematological Biomarkers of Inflammation with 10-Year Major Adverse Cardiovascular Events and All-Cause Mortality in Patients with Metabolic Dysfunction-Associated Steatotic Liver Disease: The ARIC Study. *J. Inflamm. Res.* **17**, 4247–4256 (2024).
40. Zhang, J. *et al.* Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies. *Nat. Genet.* **54**, 593–602 (2022).
41. Mi, S. Protein prediction models - ARIC. *PredictDB* (2022).
42. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, (2013).
43. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, (2016).
44. Akçimen, F. *et al.* Transcriptome-wide association study for restless legs syndrome identifies new susceptibility genes. *Communications Biology* **3**, 1–5 (2020).
45. Szklarczyk, D. *et al.* STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic*

*Acids Res.* **47**, D607–D613 (2018).

46. Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **13**, 2498–2504 (2003).

47. Gong, W., Guo, P., Liu, L., Guan, Q. & Yuan, Z. Integrative Analysis of Transcriptome-Wide Association Study and mRNA Expression Profiles Identifies Candidate Genes Associated With Idiopathic Pulmonary Fibrosis. *Front. Genet.* **11**, (2020).

48. Chin, C. H. *et al.* cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* **8 Suppl 4**, (2014).

49. Bader, G. D. & Hogue, C. W. V. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4**, 1–27 (2003).

50. Sun, P. G., Gao, L. & Han, S. Prediction of Human Disease-Related Gene Clusters by Clustering Analysis. *Int. J. Biol. Sci.* **7**, 61 (2011).

51. Song, J. *et al.* Integrative transcriptome-wide analysis of atopic dermatitis for drug repositioning. *Communications Biology* **5**, 1–13 (2022).

52. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, (2000).

53. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, (2019).

54. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, (2023).

55. Gillespie, M. *et al.* The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* **50**, (2022).

56. Jassal, B. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, (2020).

57. The Gene Ontology Consortium *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25 (2000).

58. Aleksander, S. A. *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224**, iyad031 (2023).

59. Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, (2019).

60. Dutka, M., Zimmer, K., Ćwiertnia, M., Ilczak, T. & Bobiński, R. The role of PCSK9 in heart failure and other cardiovascular diseases-mechanisms of action beyond its effect on LDL cholesterol. *Heart Fail. Rev.* (2024) doi:10.1007/s10741-024-10409-7.

61. Levine, G. N., Keaney, J. F., Jr & Vita, J. A. Cholesterol reduction in cardiovascular disease. Clinical benefits and possible mechanisms. *N. Engl. J. Med.* **332**, 512–521 (1995).

62. Daniels, T. F., Killinger, K. M., Michal, J. J., Wright, R. W., Jr & Jiang, Z. Lipoproteins, cholesterol homeostasis and cardiac health. *Int. J. Biol. Sci.* **5**, 474–488 (2009).

63. Gaffney, D. J. Global properties and functional complexity of human gene regulatory variation. *PLoS Genet.* **9**, e1003501 (2013).

64. Gutierrez-Arcelus, M. *et al.* Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing. *PLoS Genet.* **11**, e1004958 (2015).

65. Sollis, E. *et al.* The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res.* **51**, D977–D985 (2023).

66. Wheeler, D. L. *et al.* Database resources of the National Center for Biotechnology Information: update. *Nucleic Acids Res.* **32**, D35–40 (2004).

67. Gao, Y. *et al.* Genome-wide analysis of DNA methylation and risk of cardiovascular disease

in a Chinese population. *BMC Cardiovasc. Disord.* **21**, 240 (2021).

68. Girerd, N. *et al.* Inflammation and remodeling pathways and risk of cardiovascular events in patients with ischemic heart failure and reduced ejection fraction. *Sci. Rep.* **12**, 8574 (2022).

69. Liu, T. *et al.* Identification of Circular RNA-MicroRNA-Messenger RNA Regulatory Network in Atrial Fibrillation by Integrated Analysis. *Biomed Res. Int.* **2020**, 8037273 (2020).

70. Clemitson, J.-R. *et al.* Genetic dissection of a blood pressure quantitative trait locus on rat chromosome 1 and gene expression analysis identifies SPON1 as a novel candidate hypertension gene. *Circ. Res.* **100**, 992–999 (2007).

71. Jahanshad, N. *et al.* Genome-wide scan of healthy human connectome discovers SPON1 gene variant influencing dementia severity. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 4768–4773 (2013).

72. Fernandez, S. *et al.* Is Associated with Amyloid-β and ε4-Related Cognitive Decline in Cognitively Normal Adults. *J Alzheimers Dis Rep* **5**, 111–120 (2021).

73. Liu, Z. *et al.* APOE influences working memory in non-demented elderly through an interaction with SPON1 rs2618516. *Hum. Brain Mapp.* **39**, 2859–2867 (2018).

74. Tamjidifar, R. *et al.* Prognostic and Diagnostic Values of miR-506 and SPON 1 in Colorectal Cancer with Clinicopathological Considerations. *J. Gastrointest. Cancer* **52**, 125–129 (2021).

75. Miyakawa, R. *et al.* SPON1 is an independent prognostic biomarker for ovarian cancer. *J. Ovarian Res.* **16**, 95 (2023).

76. Shah, A. M. *et al.* Large scale plasma proteomics identifies novel proteins and protein networks associated with heart failure development. *Nat. Commun.* **15**, 528 (2024).

77. Kozel, B. A. *et al.* Williams syndrome predisposes to vascular stiffness modified by antihypertensive use and copy number changes in NCF1. *Hypertension* **63**, 74–79 (2014).

78. Liu, H. *et al.* Ncf1 knockout in smooth muscle cells exacerbates angiotensin II-induced aortic aneurysm and dissection by activating the STING pathway. *Cardiovasc. Res.* **120**, 1081–1096 (2024).

79. Holmdahl, R., Sareila, O., Olsson, L. M., Bäckdahl, L. & Wing, K. Ncf1 polymorphism reveals oxidative regulation of autoimmune chronic inflammation. *Immunol. Rev.* **269**, 228–247 (2016).

80. Da Dalt, L. *et al.* PCSK9 deficiency rewires heart metabolism and drives heart failure with preserved ejection fraction. *Eur. Heart J.* **42**, 3078–3090 (2021).

81. Xu, Q. *et al.* PCSK9: A emerging participant in heart failure. *Biomed. Pharmacother.* **158**, 114106 (2023).

82. Nayor, M. *et al.* Aptamer-Based Proteomic Platform Identifies Novel Protein Predictors of Incident Heart Failure and Echocardiographic Traits. *Circ. Heart Fail.* **13**, e006749 (2020).

83. Hsu, J. *et al.* Genetic Control of Left Atrial Gene Expression Yields Insights into the Genetic Susceptibility for Atrial Fibrillation. *Circ Genom Precis Med* **11**, e002107 (2018).

84. Wang, Q. *et al.* Epistasis regulates genetic control of cardiac hypertrophy. *Res Sq* (2023) doi:10.21203/rs.3.rs-3509208/v1.

85. Wang, X. *et al.* Genetic and phenotypic relationships between coronary atherosclerotic heart disease and electrocardiographic traits. in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (IEEE, 2022). doi:10.1109/bibm55620.2022.9995557.

86. Gao, H., Liu, Y., Sheng, W., Shou, W. & Huang, G. Progresses in genetic testing in congenital heart disease. *Medicine Plus* **1**, 100028 (2024).

87. Shah, S. *et al.* Four genetic loci influencing electrocardiographic indices of left ventricular hypertrophy. *Circ. Cardiovasc. Genet.* **4**, 626–635 (2011).

88. Akhtar, M. M. *et al.* Clinical Phenotypes and Prognosis of Dilated Cardiomyopathy Caused by Truncating Variants in the Gene. *Circ. Heart Fail.* **13**, e006832 (2020).

89. Chauveau, C. *et al.* Recessive TTN truncating mutations define novel forms of core myopathy with heart disease. *Hum. Mol. Genet.* **23**, 980–991 (2014).

90. Tharp, C. A., Haywood, M. E., Sbaizero, O., Taylor, M. R. G. & Mestroni, L. The Giant Protein Titin's Role in Cardiomyopathy: Genetic, Transcriptional, and Post-translational Modifications of TTN and Their Contribution to Cardiac Disease. *Front. Physiol.* **10**, 1436 (2019).

91. Stark, K. *et al.* Genetic association study identifies HSPB7 as a risk gene for idiopathic dilated cardiomyopathy. *PLoS Genet.* **6**, e1001167 (2010).

92. Cappola, T. P. *et al.* Common variants in HSPB7 and FRMD4B associated with advanced heart failure. *Circ. Cardiovasc. Genet.* **3**, 147–154 (2010).

93. Begay, R. L. *et al.* Gene Splice Mutations Cause Dilated Cardiomyopathy. *JACC Basic Transl Sci* **1**, 344–359 (2016).

94. Cui, H. *et al.* Mutation profile of FLNC gene and its prognostic relevance in patients with hypertrophic cardiomyopathy. *Mol Genet Genomic Med* **6**, 1104–1113 (2018).

95. Van Sligtenhorst, I. *et al.* Cardiomyopathy in α-kinase 3 (ALPK3)-deficient mice. *Vet. Pathol.* **49**, 131–141 (2012).

96. Almomani, R. *et al.* Biallelic Truncating Mutations in ALPK3 Cause Severe Pediatric Cardiomyopathy. *J. Am. Coll. Cardiol.* **67**, 515–525 (2016).

97. Cappola, T. P. *et al.* Loss-of-function DNA sequence variant in the CLCNKA chloride channel implicates the cardio-renal axis in interindividual heart failure risk variation. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 2456–2461 (2011).

98. Smith, G. L. *et al.* Renal impairment and outcomes in heart failure: systematic review and meta-analysis. *J. Am. Coll. Cardiol.* **47**, 1987–1996 (2006).

99. Damman, K. *et al.* Worsening renal function and prognosis in heart failure: systematic review and meta-analysis. *J. Card. Fail.* **13**, 599–608 (2007).

100. Lalonde, S. *et al.* Integrative analysis of vascular endothelial cell genomic features identifies AIDA as a coronary artery disease candidate gene. *Genome Biol.* **20**, 133 (2019).

101. Kannel, W. B. *et al.* Profile for estimating risk of heart failure. *Arch. Intern. Med.* **159**, 1197–1204 (1999).

102. Nayak, A., Hicks, A. J. & Morris, A. A. Understanding the Complexity of Heart Failure Risk and Treatment in Black Patients. *Circ. Heart Fail.* **13**, e007264 (2020).