HF-ETIOLOGY: Endotypes from Ethical, Multi-Modal AI driven and Molecular/Phenotypic Data Integration Enabled Discovery

07142026 Meeting | Northwestern, Cleveland Clinic, Weill Cornell

LUO, YUAN | PI
CHENG, FEIXIONG | MPI
HOLMES, KRISTI | MPI
WANG, FEI | MPI

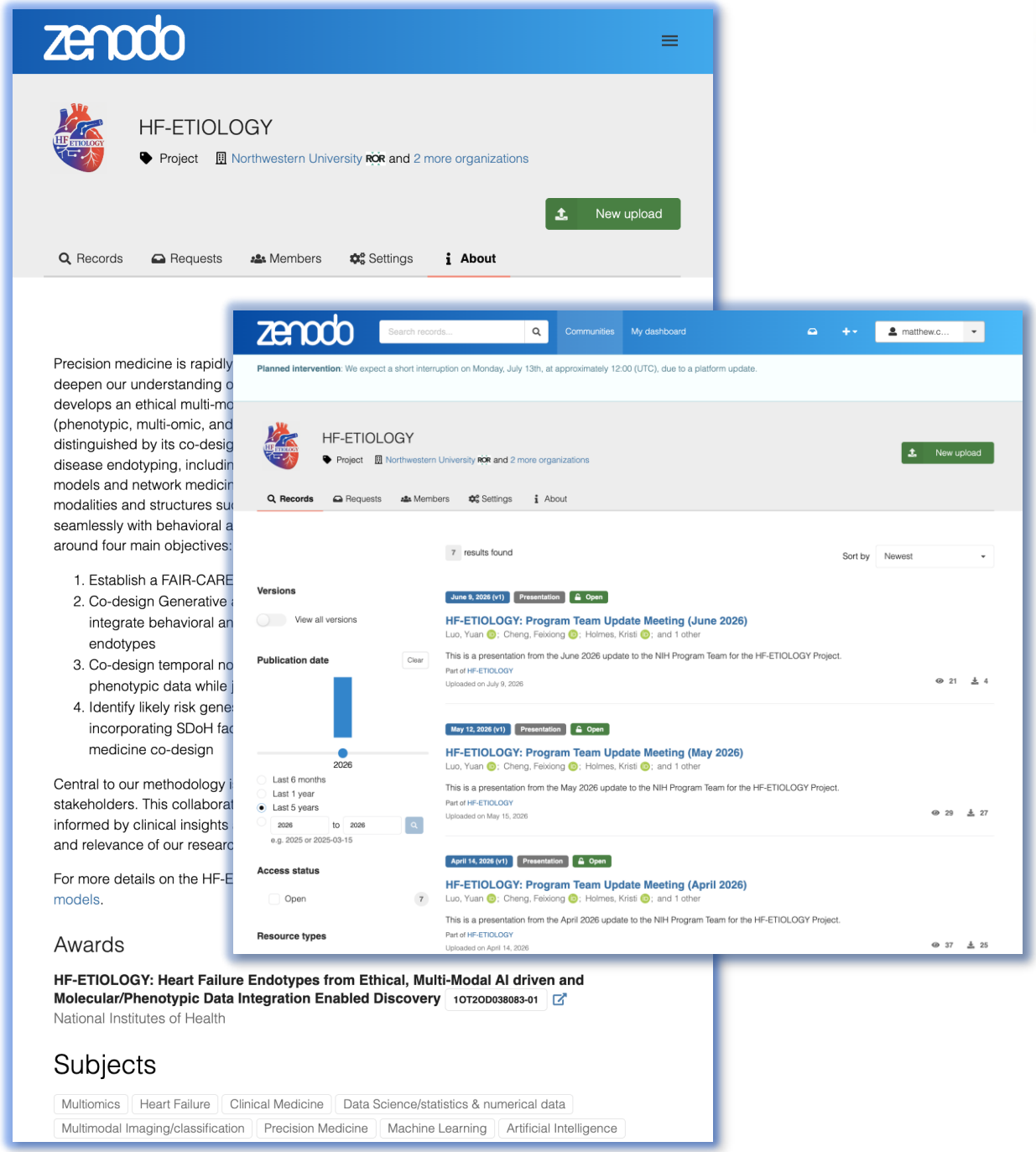
Milestones for Month 9 – Aim 1

- Iterate on the FAIR-CARE framework with stakeholder feedback to co-design HF-ETIOLOGY data and models
- **Exit Criteria (In iteration)**
  - Quarterly reconvene relevant stakeholders to review and update criteria, workflows, and documentation templates
- **Deliverables (In iteration)**
  - Iterated and refined criteria and workflows for HF-ETIOLOGY
  - Adjusted priorities for infrastructure of FAIR-CARE framework

HF-ETIOLOGY Commons

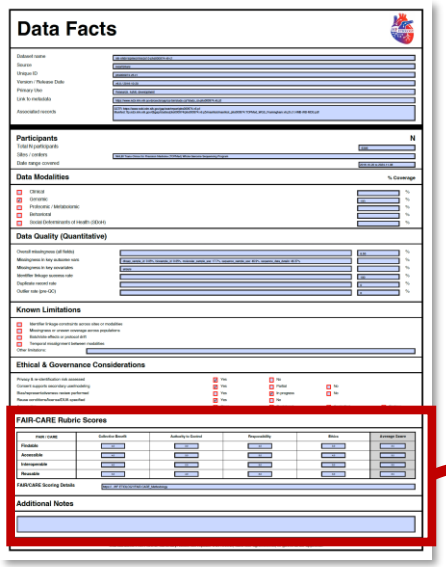
- The **HF-ETIOLOGY Commons** serves as a resource for project members and a hub to share our progress with the community

- Project outputs are open access



Previous Rubric Scoring Strategy

Score Category	Score Range	Interpretation
Low	1.0 - 2.0	Weak evidence or missing implementation
Developing	2.1 - 3.0	Partially documented or indirect evidence
Strong	3.1 - 4.0	Well documented with some gaps
Exemplary	4.1 - 5.0	Publicly documented, machine-actionable, and governance-ready



FAIR-CARE Rubric Scores

FAIR / CARE	Collective Benefit	Authority to Control	Responsibility	Ethics	Average Score
Findable	4.0	3.0	4.0	3.0	3.5
Accessible	4.0	3.0	3.0	4.0	3.5
Interoperable	4.0	2.0	3.0	3.0	3.0
Reusable	5.0	4.0	3.0	3.0	3.5

FAIR/CARE Scoring Details https://.../HF-ETIOLOGY/FAIR-CARE_Methodology

Additional Notes

FAIR-CARE Data Label

Evidence and Maturity Summary

Dataset Identity

facts, not scores

Dataset name	nih-nhlbi-topmed-freeze10-phs000974-v6-c1
Source / program	NHLBI TOPMed / Framingham
Unique ID	phs000974.v6.c1
Version / release	v6.0 / 2016-10-20
Primary use	Research, controlled secondary analysis, AI/ML development with restrictions
Metadata URL	https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000974.v6.p5
Associated records	SSTR: phs000974.v6.p1; Manifest: TOPMed_WGS_Framingham.v6.p5.c1.HMB-IRB-MDS.pdf Used in CVF_analysis.ipynb (https://github.com/luoyuanlab/hf_etiology_data/blob/main/VCF_analysis.ipynb)

Objective Data Facts

measurable fields and access conditions

Total N	6,396	Consent / use terms	HMB-IRB-MDS; secondary use governed by dbGaP/DUA
Sites / centers	Framingham Heart Study / NHLBI TOPMed	Metadata status	Present
Date range	Longitudinal cohort; release-specific	Data dictionary	Partial / controlled
Access tier	Controlled access	DUA/access process	Present

Quantitative Quality Indicators

numeric metrics retained separately from governance maturity

Overall missingness	Identifier linkage	Duplicate record rate	Outlier rate pre-QC
8.55%	100%	0%	0%

Missing-ness in key outcome variables: dbgap_sample_id: 0.05%, biosample_id: 0.05%, molecular_sample_use: 17.7%, sequence_sample_use: 46.9%, sequence_data_details: 46.37%

Maturity Summary

anchored maturity + confidence, not a single compliance score

Domain	Maturity	Confidence	Key evidence / interpretation
FAIR readiness	Mature	High	Persistent accession, landing page, access process; machine-actionability incomplete.
CARE readiness	Developing	Medium	Governance exists, but participant/community authority is not publicly documented.
AI reuse readiness	Developing	Medium	Reusable for controlled research; explicit model-training terms are limited.
Privacy / ethics	Mature	High	Controlled access, consent terms, DUA and IRB-governed reuse.
Evidence completeness	Developing	Medium	Some evidence is public; some remains in controlled documentation or is implicit.

FAIR/CARE Analysis Details: https://.../HF-ETIOLOGY/FAIR-CARE_Methodology

Top Gaps and Recommended Actions

improvement-oriented output

1. Document participant/community authority and downstream governance expectations.
2. Add explicit AI/model-training reuse guidance and prohibited-use language.
3. Publish clearer machine-readable metadata, provenance, and subgroup quality summaries.

New label design and scoring strategy

This label summarizes evidence, maturity, confidence, and recommended actions. It does not replace IRB review, data use agreements, or governance approval.

FAIR-CARE Data Label

Evidence and Maturity Evidence Checklist

How to Interpret This Label

separates facts, evidence, maturity, and confidence

This FAIR-CARE Data Label avoids presenting a single quantitative compliance score. It separates objective dataset facts and quantitative quality indicators from evidence-based maturity judgments. Each maturity judgment should be supported by a cited evidence source and a confidence rating. Use 'Unknown / not documented' when evidence is unavailable rather than forcing a low score.

Maturity Rating Anchors

ordinal, evidence-based ratings

0 - Not documented	No public or supplied evidence that the criterion is addressed.
1 - Minimal	Criterion is weakly addressed or only inferred from limited evidence.
2 - Developing	Criterion is partially addressed; important gaps or ambiguity remain.
3 - Mature	Criterion is substantially addressed with clear evidence and routine implementation.
4 - Advanced / auditable	Criterion is documented, accountable, machine-actionable where relevant, monitored, and regularly updated.
N/A	Criterion does not apply to this dataset, population, modality, or intended use.

Evidence Inventory

yes / partial / no / unknown / N/A + evidence source

Evidence item	Status	Evidence source / note
Persistent identifier or accession	Yes	dbGaP accession phs000974.v6.c1
Public landing page / metadata URL	Yes	https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000974.v6.p5
Version or release documented	Yes	v6.0 / 2016-10-20
Variable-level documentation / data dictionary	Partial	Controlled documentation; not all details public
Consent and reuse terms documented	Yes	HMB-IRB-MDS; secondary use governed by dbGaP/DUA
DUA / controlled access process documented	Yes	dbGaP controlled-access workflow
Provenance and processing history	Partial	Manifest/release metadata available; full harmonization details limited
Representativeness and subgroup coverage	Partial	Cohort context documented; subgroup quality summaries limited
AI/model-training use conditions	Unknown/Partial	Not explicit as a model-training governance field
Participant/community authority	Unknown	No public evidence of ongoing participant/community authority mechanism

Reviewer and Confidence Metadata

supports consistency across projects

Reviewer	MBC
Review date	6/6/2026
Use case assessed	Secondary research / AI-readiness review
Overall evidence confidence	Medium
Known caveats	Assessment based on public and supplied documentation; internal governance records may change ratings.

Method Note

for consistency

For consistency across projects, reviewers should score only after recording evidence. Use the anchors above, record confidence, and compare scores against calibration examples. Composite numeric scores should be optional summaries, not the primary finding.

Recommended workflow: evidence inventory -> anchored maturity rating -> confidence -> recommendations -> optional summary.

HeartShare Omics Data Sets

Dataset	Number of subjects	Subject IDs	SSTR Link	Parent Cohort	Linkable?	Notes	Data Label
CHS parent (and substudies)	5612	100051 - 999909	https://www.ncbi.nlm.nih.gov/gap/sstr/report/phs000287.v7.p1	CHS	Fully	Hidden for privacy	
CHS TOPMED	4877	100051 - 999909	https://www.ncbi.nlm.nih.gov/gap/sstr/report/phs001368.v4.p2	CHS	Fully		
CHS Imaging	5539	100051 - 999909	https://www.ncbi.nlm.nih.gov/gap/sstr/report/phs003639.v1.p1	CHS	Fully		
CHS BioLINCC	Pending	100051 - 999909	Pending - dbGaP registration, but HeartShare has access through BioLINCC access within BDC	CHS	Fully		
JHS parent (and substudies)	5885	1 - 3883	https://www.ncbi.nlm.nih.gov/gap/sstr/report/phs000286.v7.p2	JHS	Fully		
JHS TOPMED	5588	1 - 3883	https://www.ncbi.nlm.nih.gov/gap/sstr/report/phs000964.v5.p1	JHS	Fully		
JHS BioLINCC	3883	1 - 3883	https://www.ncbi.nlm.nih.gov/gap/sstr/report/phs003740.v1.p1	JHS	Fully		
JHS Imaging	3883	1 - 3883	https://www.ncbi.nlm.nih.gov/gap/sstr/report/phs003747.v1.p1	JHS	Fully		
JHS C4R	3883	1 - 3883	https://www.ncbi.nlm.nih.gov/gap/sstr/report/phs002907.v1.p1	JHS	Fully		
ARIC Parent	15162	131000008 - 131199987	https://www.ncbi.nlm.nih.gov/gap/sstr/report/phs000280.v8.p2	ARIC	Mostly		

Next Steps

- Continue…
  - to gather feedback on data label template and content
  - to identify gaps in metadata collection
  - to refine the process for retrieving metadata on data sets
  - to create data labels for datasets in the HeartShare list

- Refine data template as needed

Milestones for Month 9 – Aim 2

- Co-design Generative and Adjustable Prior Bayesian Tensor Factorization (GAP-BTF) to integrate behavioral and SDoH factors with multi-omic network feature learning to identify HF endotypes

- **Exit Criteria (In iteration)**
  - GAP-BTF model co-designed and implemented

- **Deliverables (In iteration)**
  - Initial analysis of HF endotypes with outcome prediction performance (e.g., HF hospitalization, mortality) and interpretability metrics

CHS dataset information

Same with previous test

Study **Cardiovascular Health Study (CHS)**
Outcome **hfdth_status** **(HF mortality, binary)**
Patients with hfdth_status avaliable **4,736**
Patients with phenotype **5,795**
Patients with genomics **4,719**
Patients with all genomics phenotypes and outcomes available **3,836**
Class balance (positives) **283 / 3,836 = 7.37%**

Model Inputs (tensor + phenotype)

Genomic tensor **3,836 pts × 374 pathways × 1,880 genes**
GAP-BTF latent dims **K=T=M**
Clinical features **39 numeric (acei, age, bmi, ef_cat, htn, …)**

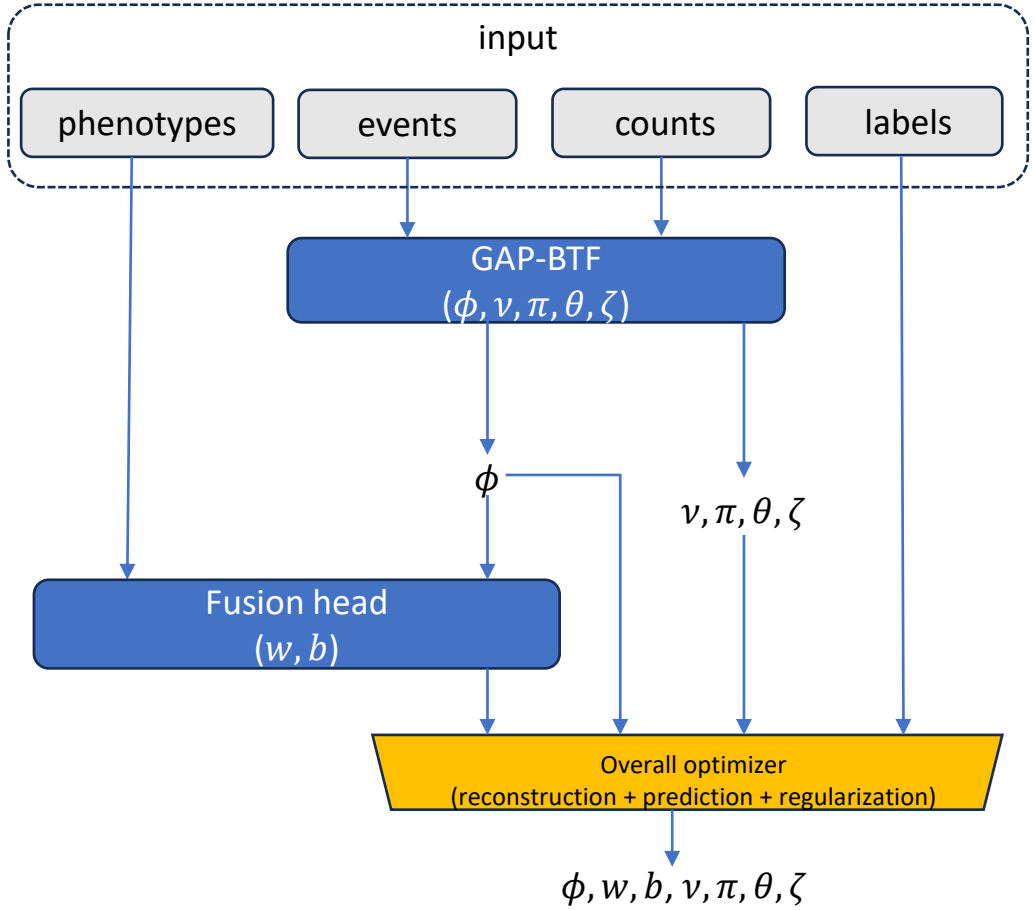
Split	Patients	% of cohort	Positives	Pos. rate
Train	2,301	60.0 %	170	7.39 %
Val	767	20.0 %	56	7.30 %
Test	768	20.0 %	57	7.42 %
Total	**3,836**	**100.0 %**	**283**	**7.37 %**

Splitting protocol

- Stratified by hfdth_status
- 60 % train / 20 % val / 20 % test
- Class balance preserved across splits

GAP-BTF Model architecture update

Previous: patient groups is a learnable parameter

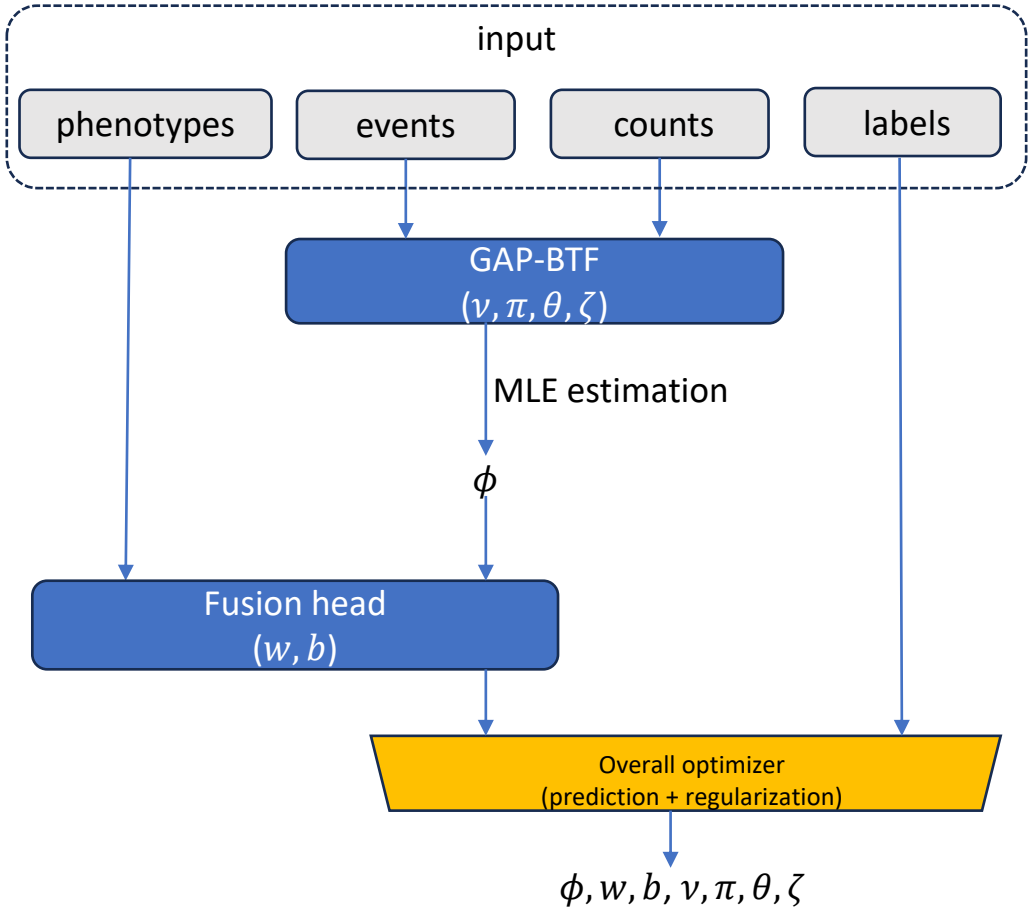


shortcomings:

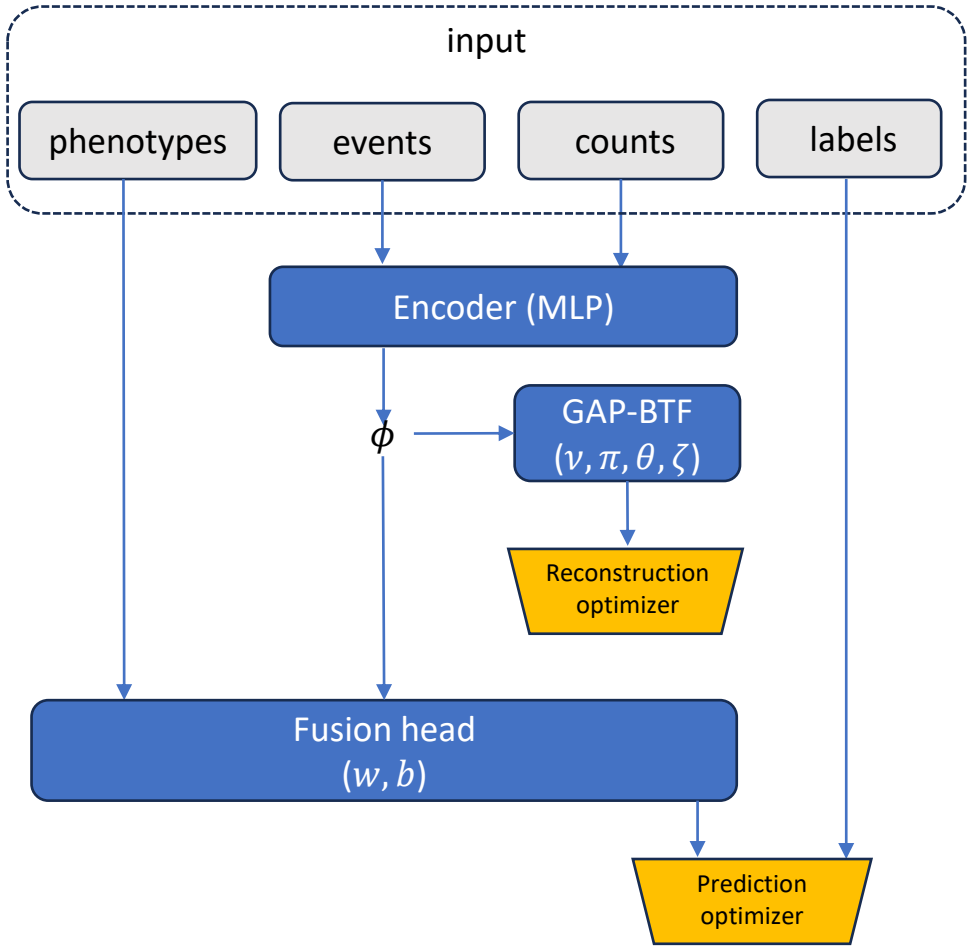
- **Information asymmetry:** In training, ϕ is shaped by the *joint* loss (reconstruction + prediction), so it's label-aware. At inference, ϕ_test is fit by reconstruction only.
- **Non-deterministic predictions**: Inference is an optimization loop, not a forward pass. The result depends on initialization, learning rate, and step count

New solutions: predict patient group ϕ from data
MLE solution: using MLE to get an estimation of ϕ.
Encoder solution: using a learnable encoder to predict ϕ.

MLE Solution



Encoder Solution



Two consistent ways to estimate ϕ, removing the previous train/inference gap: MLE (left) re-estimates ϕ by maximum likelihood; Encoder (right) learns an MLP mapping inputs → ϕ.

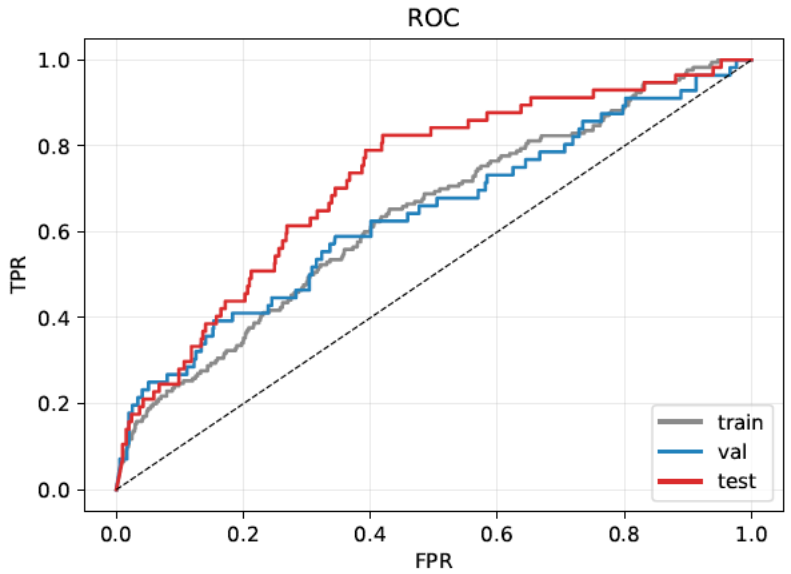
Results

models	AUROC
LR (phenos + omics)	0.6759

		Learnable Phi	MLE solution	Encoder solution
model	KTM	AUROC	AUROC	AUROC
GAP-BTF	10,10,10	0.7154	0.6650	**0.7267**
	20,20,20	0.7238	0.6668	0.7097
	30,30,30	0.7114	0.6633	0.7026
	40,40,40	**0.7410**	0.6580	0.7033
	50,50,50	0.7129	0.6586	0.7173

Encoder solution

KTM = 10, 10, 10



Interpretation

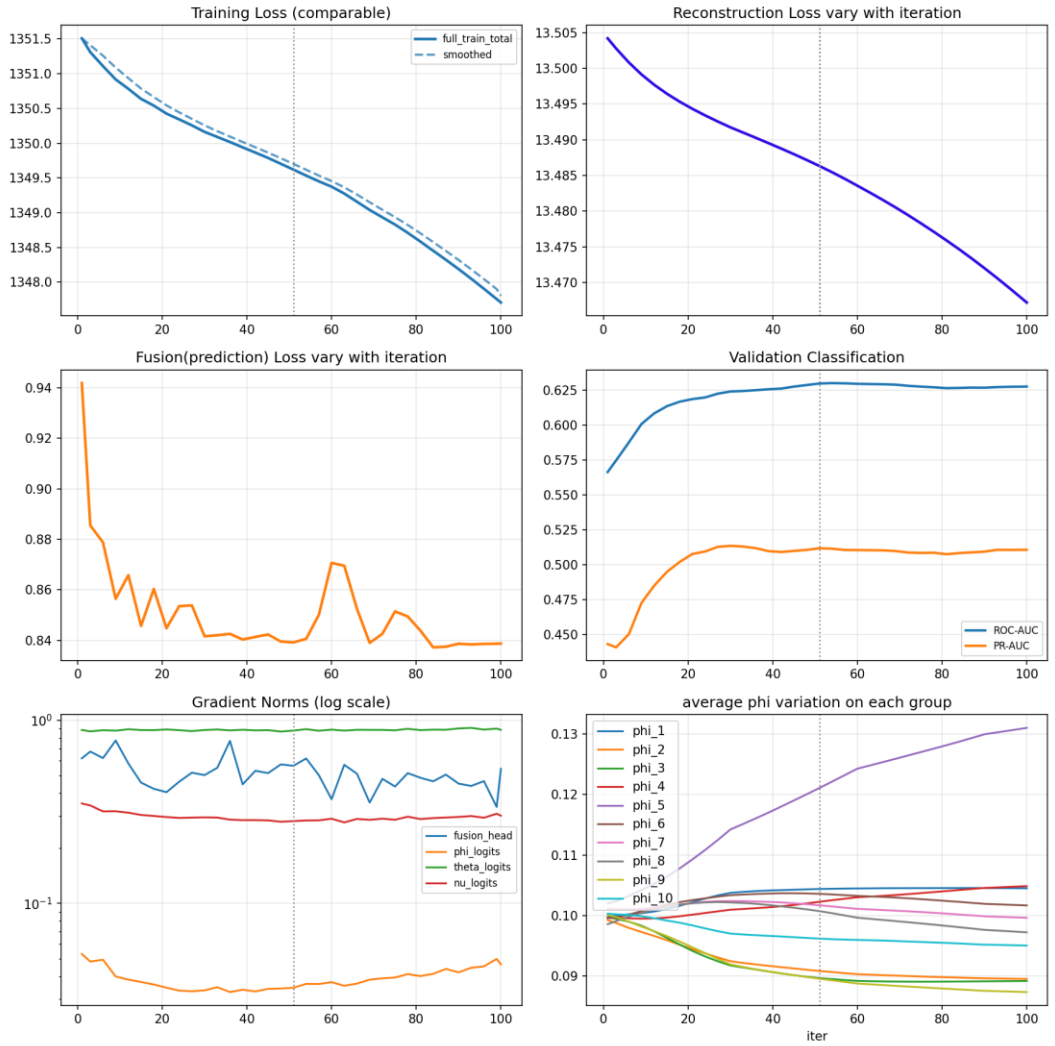
- Two new consistent ϕ solutions (Encoder, MLE) remove the train/inference gap of the learnable ϕ.
- Encoder is the best deployable option: AUROC 0.7267 at KTM=10, close to learnable's best (0.7410)
- Encoder beats MLE on AUROC at every KTM.

Diagnostic analysis

KTM = 10, 10, 10

Interpretation

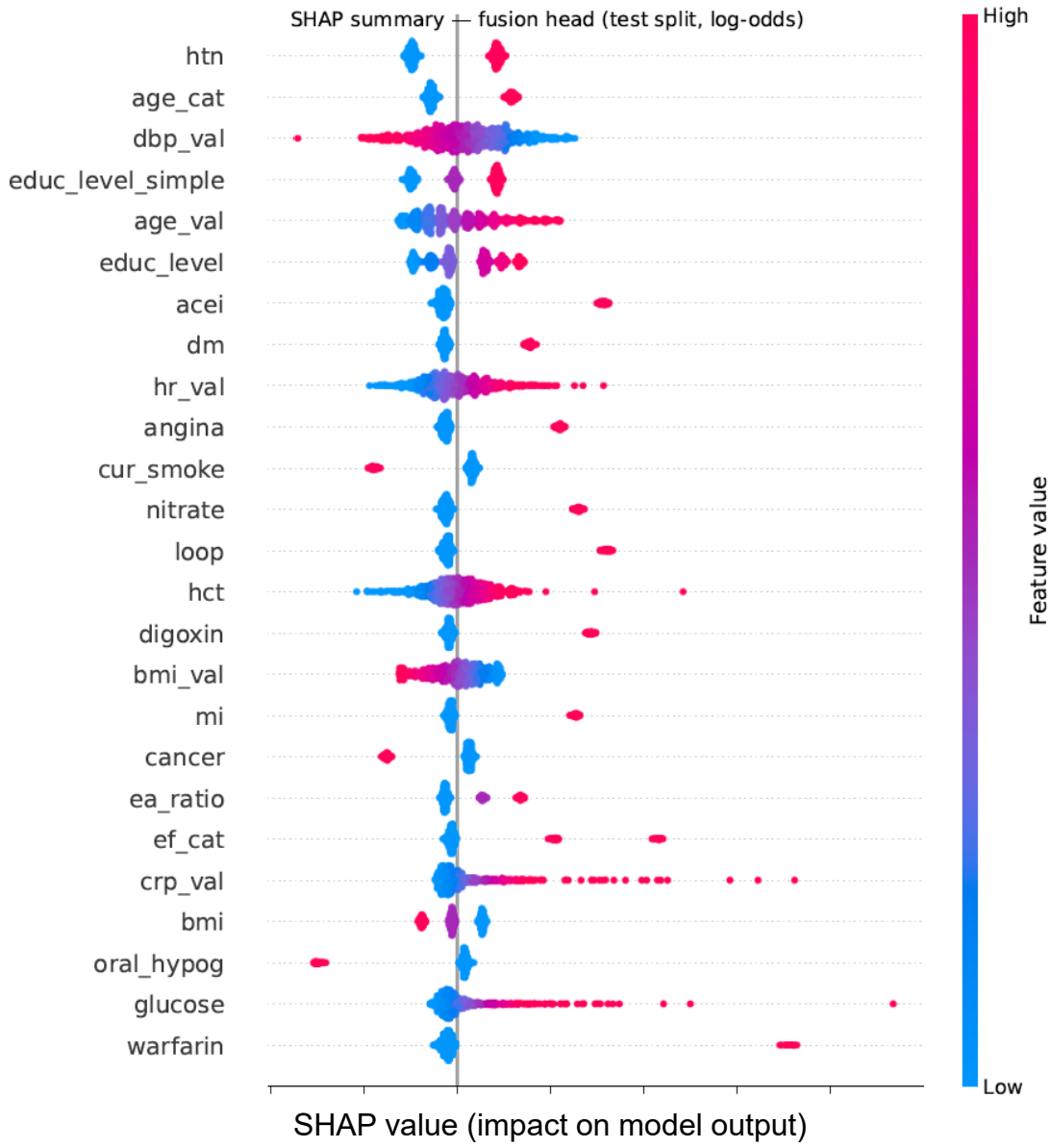
- **Convergence:** training and reconstruction losses and the fusion loss decline steadily. Training is stable.
- **Validation:** ROC-AUC oscillates around 0.55–0.62 across iterations.
- **Gradients:** fusion head, encoder, θ and v norms stay within a stable band on the log scale, no exploding or vanishing gradients.
- **Latent factors (ϕ):** the 10 group means stay in a tight range (~0.09–0.13) and progressively spread apart.



Feature importance analysis

- Explain 768 test patients

KTM = 10, 10, 10



Interpretation

- **Clinically-driven predictions:** the top features are all established risk factors, hypertension (htn), age (age_cat, age_val), diastolic BP, and education level, giving the model strong face validity.
- **SDoH signal present:** education level ranks among the top drivers, capturing the social-determinant contribution the project set out to incorporate.
- **Fewer latent factors on top:** unlike the learnable-ϕ model, the Encoder's top-25 is dominated by clinical variables; the ϕ components contribute more diffusely here.
- **Directionality is sensible:** high feature values (red) for htn, age and diabetes push risk up, matching clinical expectation.

Interpretation

- **Consistent ϕ estimation solved in two ways:** we added an MLE estimator and an MLP Encoder so patient factors ϕ are computed the same way in training and inference, closing last month's train/inference gap.
- **Encoder is the best deployable model:** AUROC 0.7267 at KTM=10, the best result among the two consistent ϕ solutions.
- **MLE underperforms on ranking:** it trails on AUROC (0.658–0.667, near the LR baseline), the Encoder beats it on AUROC at every KTM.
- **Training is healthy:** for the Encoder solution, losses are stable, gradients stay well-behaved, and ϕ groups differentiate, though gains are modest at K=10.
- **Predictions are clinically grounded:** SHAP ranks hypertension, age, blood pressure and education level on top, confirming face validity and capturing the SDoH signal.
- **Class imbalance remains a challenge:** at ~7% positives, accuracy is uninformative. AUROC remains the fair comparison.

Next steps

- **Adopt the Encoder as the default ϕ estimator** and tune K·T·M jointly (beyond the equal-KTM grid). Also try advance attention encoder transformer to get ϕ.
- **Tune the filter rule** to include more genes, pathways and clinical features from the original data.
- **Strengthen the latent factors' predictive role,** the Encoder's ϕ contributes diffusely in SHAP, explore regularization or supervision so factors carry more signal, as they did in the learnable model.
- **Address class imbalance** with class weighting, focal loss, or threshold tuning to improve detection of the positive class on the ~7% HF-mortality outcome.
- **Characterize the learned patient groups:** profile each K group by its top pathways/genes to turn ϕ into clinically interpretable HF endotypes

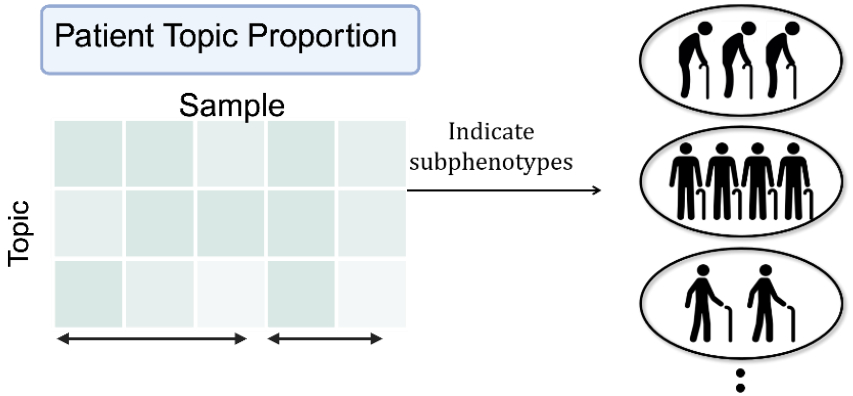
Milestones for Month 9 – Aim 3

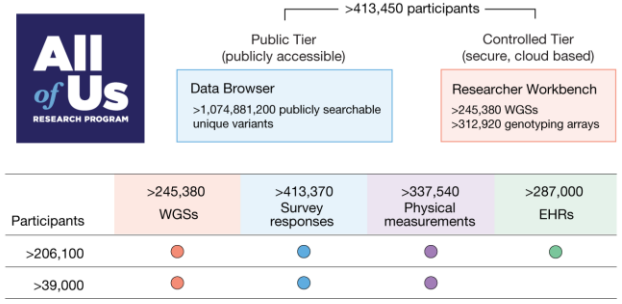
- Co-design temporal non-negative tensor factorization model (TNTF) to integrate longitudinal phenotypic data while jointly modeling behavioral and SDoH factors for HF endotyping

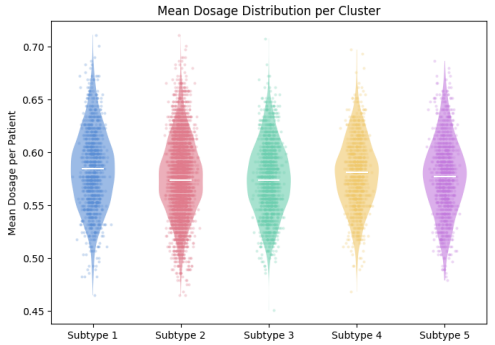
- **Exit Criteria (In iteration)**
  - HF disease network co-designed using the heart-specific human protein–protein interactome data

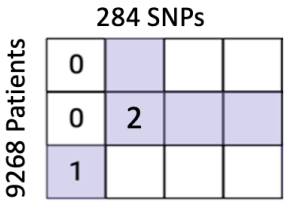
- **Deliverables (In iteration)**
  - Initial validation results from proteomics and pharmacoepidemiologic validation results of top-ranked risk genes/targets

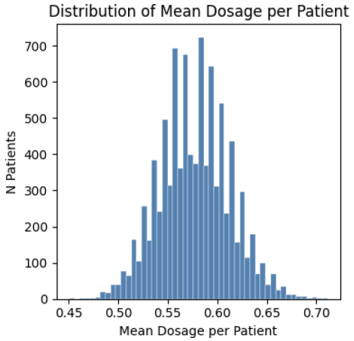
Review of previous progress











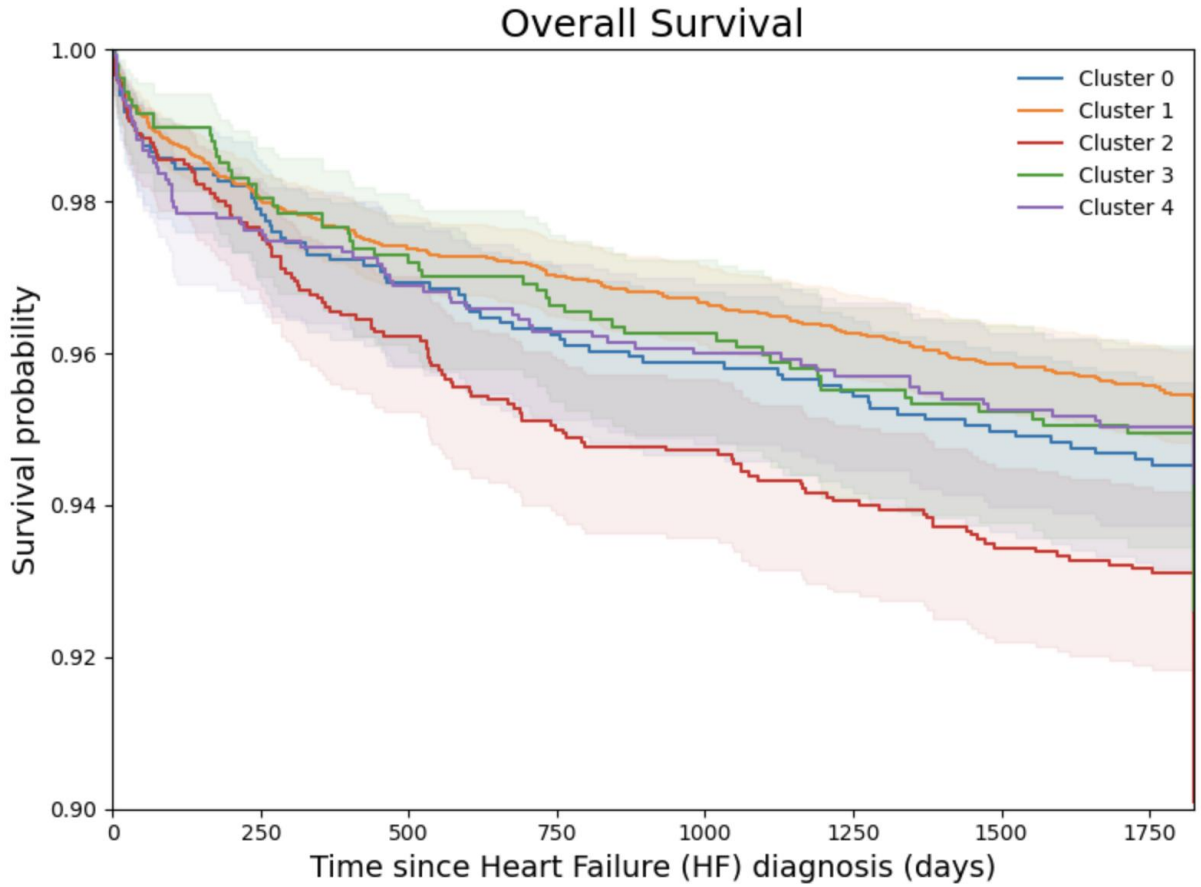
	Subtype 1	Subtype 2	Subtype 3	Subtype 4	Subtype 5	P-value
N	1,334	4,630	1,799	1,070	1,349	
Age, median [Q1, Q3]	70 [60, 78]	64 [54, 72]	62 [54, 70]	68 [60, 75]	59 [50, 68]	<0.001
Female, n (%)	563 (42.2%)	2,543 (54.9%)	938 (52.1%)	277 (25.9%)	680 (50.4%)	<0.001
Race / ethnicity, n (%)						<0.001
White	978 (73.3%)	2,494 (53.9%)	655 (36.4%)	671 (62.7%)	779 (57.7%)	
Black	143 (10.7%)	1,149 (24.8%)	597 (33.2%)	175 (16.4%)	271 (20.1%)	
Asian	12 (0.9%)	42 (0.9%)	11 (0.6%)	9 (0.8%)	5 (0.4%)	
Others	201 (15.1%)	945 (20.4%)	536 (29.8%)	215 (20.1%)	294 (21.8%)	
Dominant Diseases	Cardiac arrhythmia	Musculoskeletal Cardiometabolic syndrome Hypertension	T2D	Coronary artery	Mood disorder	

Progress of this month –Further analysis

- Outcome analysis:
  - **Mortality**: Evaluated 5-year overall survival following HF diagnosis across the five identified subtypes.
  - **Hospitalization**: Evaluated the 5-year hospitalization-free following HF diagnosis across the five identified subtypes.

- Main results:
  - Kaplan–Meier analysis of all-cause mortality (overall survival)
  - Kaplan–Meier analysis of first hospitalization
  - Interpretation of top SNPs in each HF subtype

Kaplan–Meier analysis of mortality

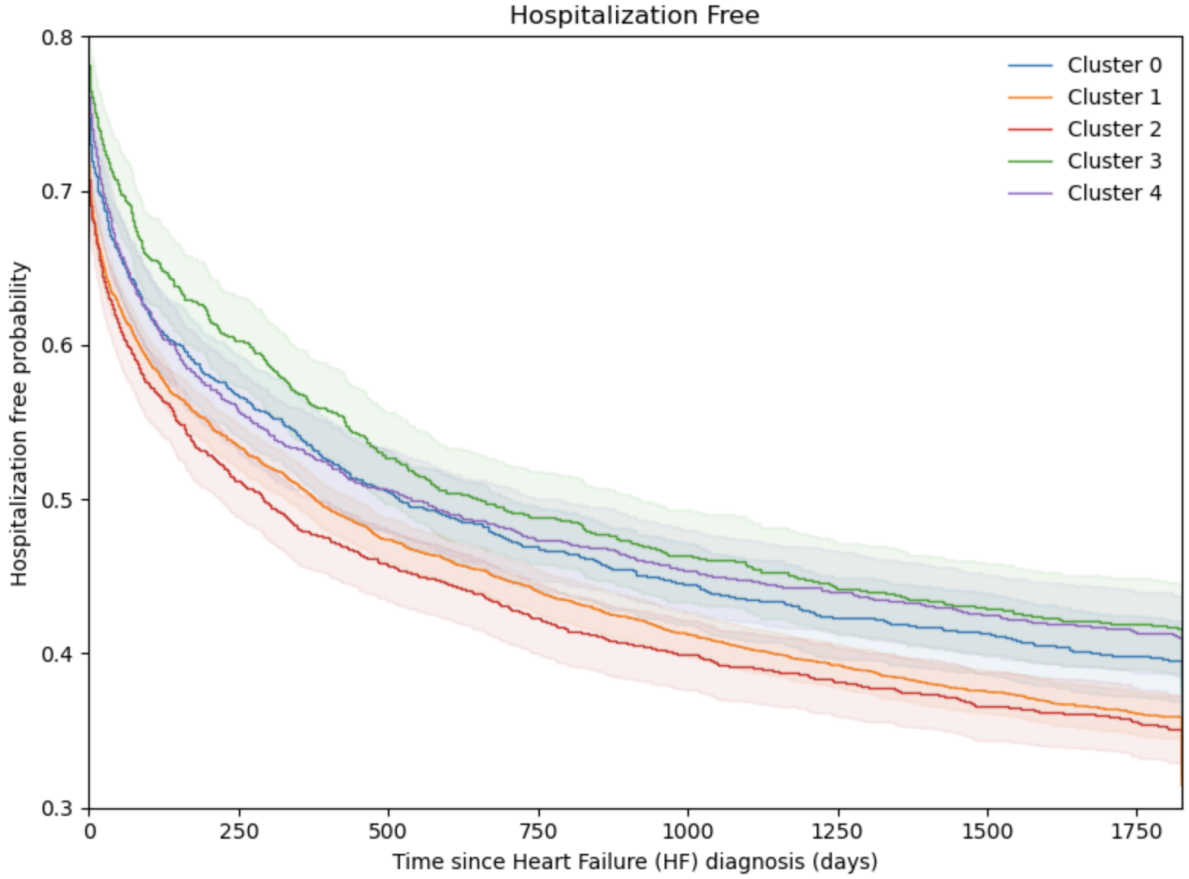


Summary: Overall, the five HF subtypes exhibited distinct all-cause mortality risks following HF diagnosis.

- **Cluster 2** consistently demonstrated the lowest survival probability throughout the 5-year follow-up period, indicating the poorest long-term prognosis.
- **Cluster 1** exhibited the highest survival probability, suggesting the most favorable survival outcome.
- **Cluster 0** was associated with an intermediate-to-high mortality risk, whereas **Clusters 3 and 4** displayed similar survival trajectories, with survival probabilities lower than Cluster 1 but slightly higher than Cluster 0.

The overall ranking of mortality risk was: Cluster 2 > Cluster 0 > Cluster 3 ≈ Cluster 4 > Cluster 1

Kaplan–Meier analysis of hospitalization

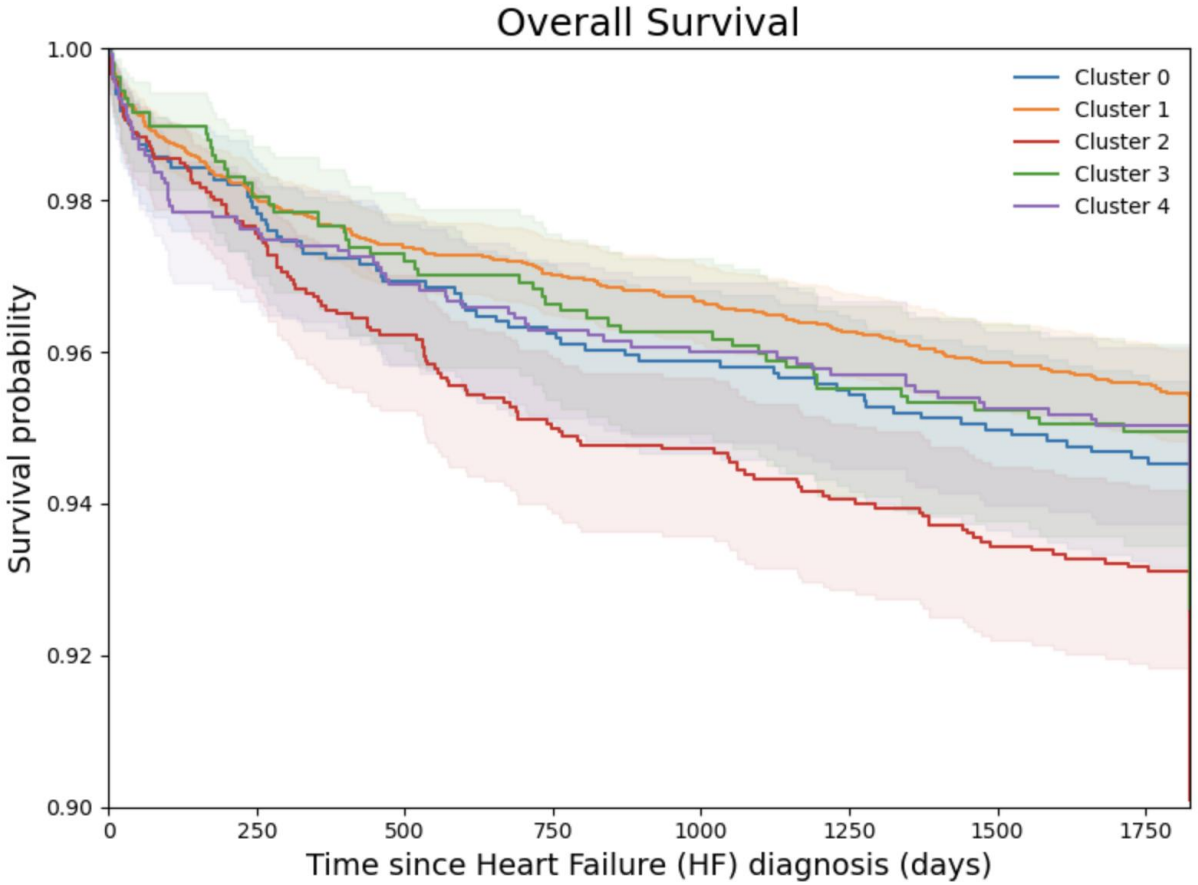


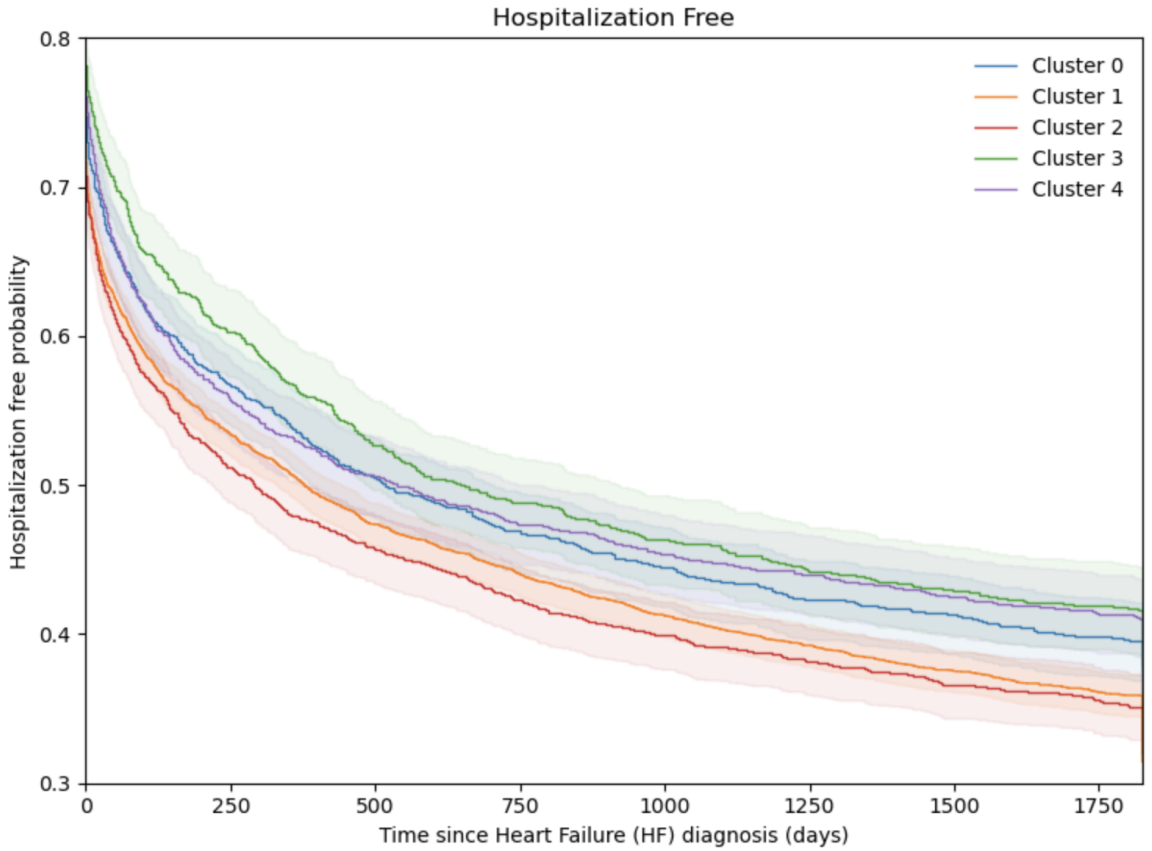
Summary: The five HF subtypes exhibited distinct hospitalization-free survival after HF diagnosis.

- **Cluster 3** had the highest hospitalization-free probability, indicating the lowest risk of first hospitalization.
- **Cluster 2** consistently showed the lowest hospitalization-free probability, suggesting the highest hospitalization risk.
- **Cluster 1** also exhibited relatively poor hospitalization-free survival, whereas **Clusters 0 and 4** demonstrated intermediate and largely overlapping trajectories.

The overall ranking of hospitalization risk was: Cluster 2 > Cluster 1 > Cluster 0 ≈ Cluster 4 > Cluster 3

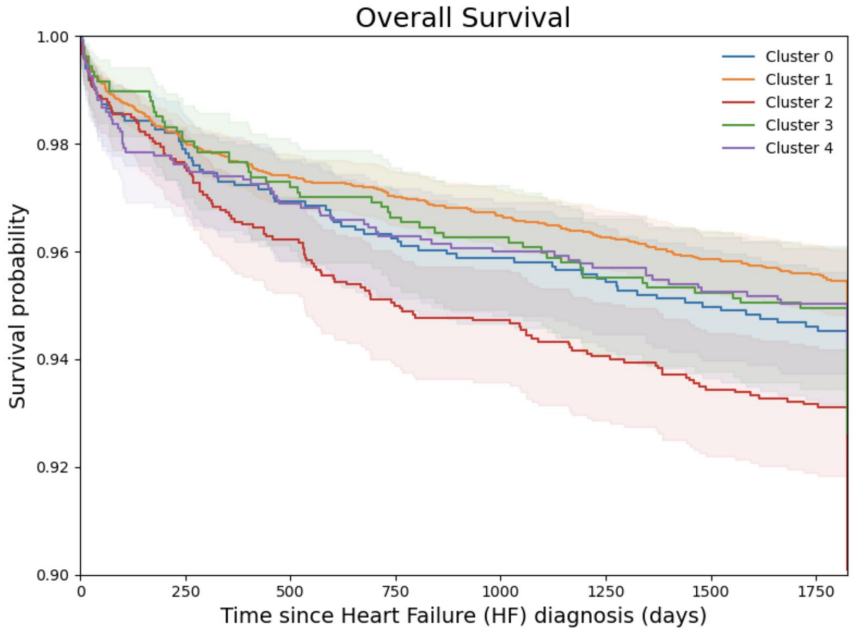
Kaplan–Meier analysis of mortality and hospitalization

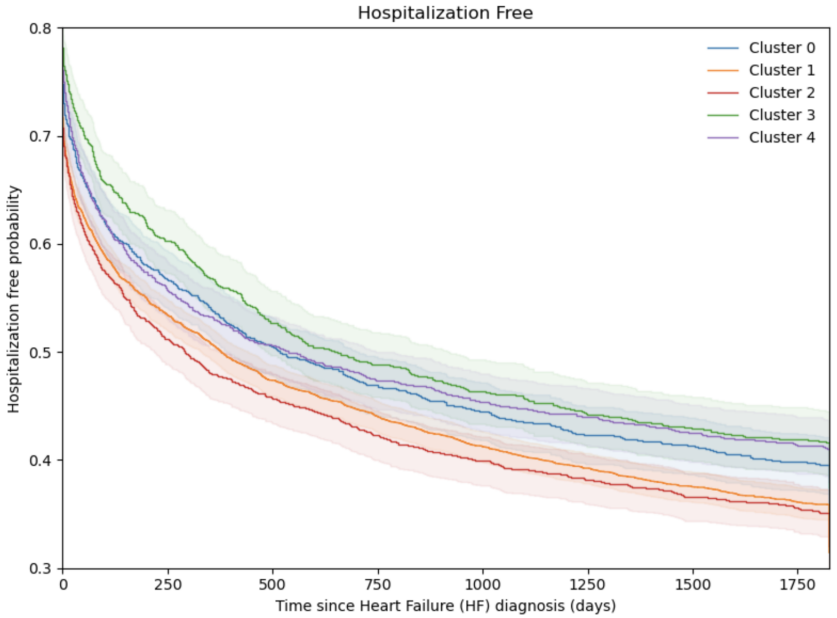


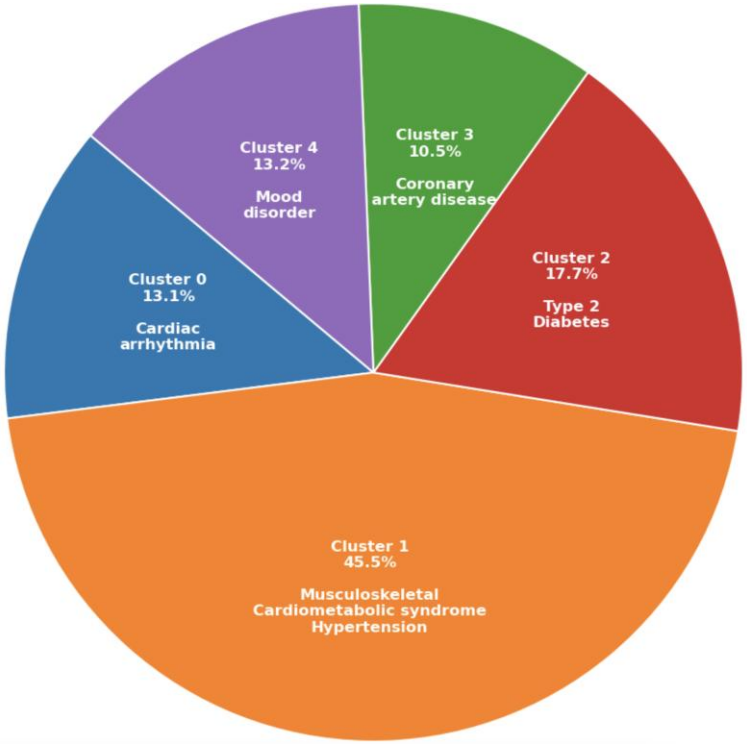


Overall Interpretation: Taken together, these findings demonstrate substantial heterogeneity in long-term clinical outcomes across the five HF subtypes. **Cluster 2** consistently represented the highest-risk subtype for both mortality and hospitalization, whereas **Cluster 1** exhibited relatively favorable survival despite a high hospitalization burden. These results suggest that the identified HF subtypes capture distinct disease trajectories and prognostic profiles, supporting their potential clinical relevance for risk stratification and personalized management.

Kaplan–Meier analysis of mortality and hospitalization







Overall Interpretation: Taken together, these findings demonstrate substantial heterogeneity in long-term clinical outcomes across the five HF subtypes. **In particular, the Type 2 Diabetes–dominant subtype (Cluster 2) consistently represented the highest-risk phenotype for both mortality and hospitalization, whereas the musculoskeletal/cardiometabolic syndrome/hypertension subtype (Cluster 1) demonstrated a unique pattern of frequent hospitalization but relatively favorable survival**.

Interpretation of top SNPs in each cluster

	Top Disease Categories	Top SNPs	Interpretation
Subtype1	Cardiac arrhythmia	'rs548097', 'rs1925951', 'rs660240', 'rs17163345', 'rs9825233', 'rs10925175', 'rs8039305', 'rs6767671', 'rs9398815', 'rs2359171', 'rs6945340', 'rs4833443', 'rs4854342', 'rs73139123', 'rs4499262', 'rs1739833', 'rs1435428', 'rs7232636', 'rs34811474', 'rs59788391', 'rs1490716'	?
Subtype2	Musculoskeletal, Cardiometabolic syndrome, Hypertension	'rs548097', 'rs1925951', 'rs660240', 'rs17163345', 'rs6945340', 'rs6767671', 'rs196321'	?
Subtype3	T2D	'rs660240', **'rs4766578'**, 'rs1408817', 'rs988748', 'rs2269997', 'rs12598405'rs10925175', 'rs740363', 'rs9398815', 'rs11066188', 'rs12150603', 'rs2517953', 'rs3184504', 'rs1384705', 'rs13179413', 'rs11644392', 'rs1484116', 'rs6779426', 'rs12357321', 'rs34811474', 'rs1757223'	rs4766578 is directly associated with type 2 diabetes mellitus and coronary artery disorder. rs660240, rs2269997, and rs1384705 map to coronary artery disorder and myocardial infarction, consistent with T2D cardiovascular complications. rs11066188 is associated with ischemic stroke, a known T2D comorbidity.
Subtype4	Coronary artery	'rs548097', 'rs1925951', 'rs660240', **'rs11257613**', 'rs196321', 'rs1484116', 'rs1861867', 'rs7220955', 'rs8039305', 'rs56094641', 'rs28451064', 'rs988748'	rs11257613 is directly associated with **coronary artery disorder**, and rs4977575 maps to **coronary artery calcification, intermediate coronary syndrome, and congestive heart failure.**
Subtype5	Mood disorder	'rs548097', 'rs1925951', 'rs1713804', 'rs11644392', 'rs2517953', 'rs4977575', 'rs12150603'	?

- Shared SNPs across subtypes (rs548097, rs1925951, rs660240) are associated with heart failure and cardiometabolic traits, **reflecting the common cardiovascular genetic background of this cohort.**
- Subtype 3 (T2D) shows the strongest SNP-disease concordance (rs4766578 directly GWAS-associated with T2D), as T2D is a metabolically distinct comorbidity with well-characterized genetic loci compared to other subtypes.

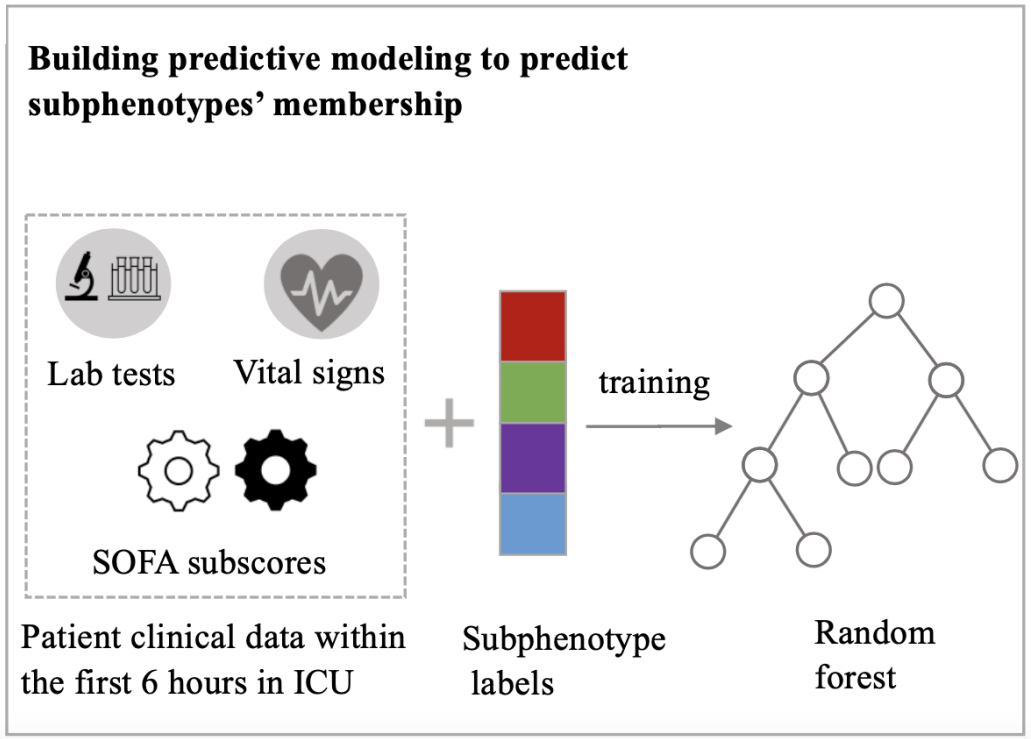
Interpretation of top SNPs in each cluster

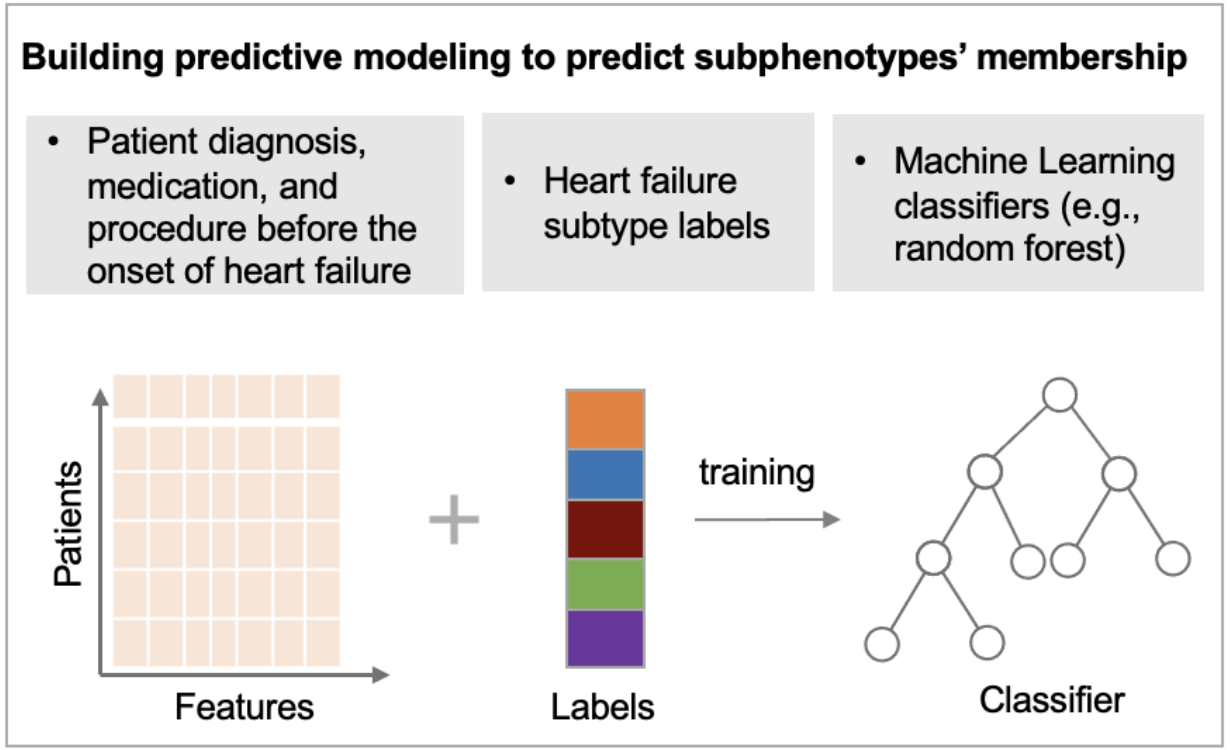
	Top Disease Categories	Top SNPs	Interpretation
Subtype1	Cardiac arrhythmia	'rs548097', 'rs1925951', 'rs660240', 'rs17163345', 'rs9825233', 'rs10925175', **'rs8039305'**, 'rs6767671', 'rs9398815', 'rs2359171', 'rs6945340', 'rs4833443', 'rs4854342', 'rs73139123', 'rs4499262', 'rs1739833', 'rs1435428', 'rs7232636', 'rs34811474', **'rs59788391'**, 'rs1490716'	rs8039305 and rs59788391 have been reported to be associated with cardiac electrophysiological traits and arrhythmia susceptibility, while rs4833443 and rs4854342 are located near genes involved in cardiac conduction and myocardial function. Together, these variants support the predominance of cardiac arrhythmia in this subtype.
Subtype2	Musculoskeletal, Cardiometabolic syndrome, Hypertension	'rs548097', 'rs1925951', 'rs660240', 'rs17163345', 'rs6945340', **'rs6767671'**, **'rs196321'**	rs6767671 has been associated with blood pressure regulation and cardiometabolic traits, whereas rs196321 has been implicated in metabolic and inflammatory pathways. Together with the shared cardiovascular loci, these variants are consistent with the enrichment of hypertension, cardiometabolic syndrome, and musculoskeletal comorbidities observed in this subtype
Subtype3	T2D	'rs660240', **'rs4766578'**, 'rs1408817', 'rs988748', 'rs2269997', 'rs12598405'rs10925175', 'rs740363', 'rs9398815', 'rs11066188', 'rs12150603', 'rs2517953', 'rs3184504', 'rs1384705', 'rs13179413', 'rs11644392', 'rs1484116', 'rs6779426', 'rs12357321', 'rs34811474', 'rs1757223'	rs4766578 is directly associated with type 2 diabetes mellitus and coronary artery disorder. rs660240, rs2269997, and rs1384705 map to coronary artery disorder and myocardial infarction, consistent with T2D cardiovascular complications. rs11066188 is associated with ischemic stroke, a known T2D comorbidity.
Subtype4	Coronary artery	'rs548097', 'rs1925951', 'rs660240', **'rs11257613**', 'rs196321', 'rs1484116', 'rs1861867**'**, 'rs7220955', 'rs8039305', 'rs56094641', 'rs28451064', 'rs988748'	rs11257613 is directly associated with **coronary artery disorder**, and rs4977575 maps to **coronary artery calcification, intermediate coronary syndrome, and congestive heart failure.**
Subtype5	Mood disorder	'rs548097', 'rs1925951', 'rs1713804', **'rs11644392'**, 'rs2517953', 'rs4977575', **'rs12150603**'	rs4977575 is a well-established cardiovascular risk locus associated with coronary artery disease, whereas rs12150603 and rs11644392 have been linked to neurological and neuropsychiatric pathways in previous studies. These findings suggest a potential genetic connection between mood disorders and cardiovascular disease, supporting the distinct clinical characteristics of this subtype.

- Shared SNPs across subtypes (rs548097, rs1925951, rs660240) are associated with heart failure and cardiometabolic traits, **reflecting the common cardiovascular genetic background of this cohort.**
- Subtype 3 (T2D) shows the strongest SNP-disease concordance (rs4766578 directly GWAS-associated with T2D), as T2D is a metabolically distinct comorbidity with well-characterized genetic loci compared to other subtypes.

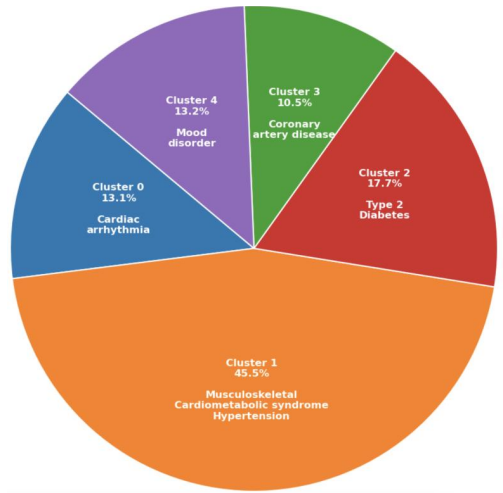
Next steps

Building predictive modeling





Xu, Zhenxing, **Chengsheng Mao**, Chang Su, Hao Zhang, Ilias Siempos, Lisa K. Torres, Di Pan, **Yuan Luo**, Edward J. Schenck, and **Fei Wang**. "Sepsis subphenotyping based on organ dysfunction trajectory." Critical care 26, no. 1 (2022): 197.



Milestones for Month 9 – Aim 4

- Identify likely risk genes/targets and pathways for HF endotypes with machine learning and network medicine approaches

- **Exit Criteria (In iteration)**
  - TNTF model co-designed and implemented to jointly integrate longitudinal phenotypic data with behavioral and SDoH factors

- **Deliverables (In iteration)**
  - Analyzing added value of longitudinal phenotypic trajectories versus snapshot models for HF outcomes

Network Medicine and Therapeutic Target Validation

- Aim 4 activities remain focused on validating prioritized HF/AF risk genes and candidate therapeutics through multi-omics, network proximity, real-world evidence, and experimental studies
- Continued consolidation of prior results, including candidate gene prioritization, multi-omics validation, drug repurposing, and target trial emulation
- Ongoing review of analysis outputs and preparation of the next phase of validation
- No major new results are being presented this month; the Aim 4 team will provide a substantive technical update at the next monthly meeting
- **Next update:** refined candidate prioritization and additional clinical and/or experimental validation results