Video Representation Learning of Cardiac MRI for Genetic Discovery

Matt Sooknah
Calico Life Sciences
mds@calicolabs.com

Sivaramakrishnan Sankarapandian
Calico Life Sciences
sivark@calicolabs.com

Ramprakash Srinivasan
Calico Life Sciences
rams@calicolabs.com

Johannes Riegler
Calico Life Sciences
riegler@calicolabs.com

Jun Xu
Calico Life Sciences
junxu@calicolabs.com

Abstract

In recent years, many studies have utilized cardiac magnetic resonance imaging (cMRI) to define image-derived phenotypes (IDPs) relating to heart structure and function for genome-wide association studies (GWAS). These IDPs are traditionally defined manually from volume, strain, and geometric parameters. Here we introduce an unsupervised learning approach that extracts spatiotemporal representations from cMRI videos in a large human cohort of $\sim$68,000 subjects from the UK BioBank. The resulting representations can be used to predict age and manually crafted IDPs accurately. We further use these representations to define IDPs to capture both known and potential novel genetic associations. Our work suggests that unsupervised learning can be used to extract rich, unbiased information from medical videos with applications to genetic discovery.

1 Introduction

Time-series cMRI is a commonly used diagnostic tool that captures heart dynamics in high spatial and temporal resolution. A long-standing goal in cardiovascular research is to define features in cMRI scans that capture cardiovascular functions (like muscle contraction, heart rhythm, and blood flow) which influence cardiac health, then combine those features with genetic data in large human cohorts to identify genes implicated in cardiovascular disease.

Prior works have developed methods to extract manually defined IDPs like ejection fraction [2] and strain rate [39] from cMRI. Multiple studies have conducted GWAS using these traits, identifying key genetic markers of cardiovascular health [29, 30, 1, 22]. These IDPs are easily defined and interpretable, but only provide a narrow view of the underlying complex structure of the heart. Others have predicted patients' "cardiovascular age" using pre-defined cMRI measurements, and then compared it to chronological age to obtain a "delta-age" phenotype [33]. Compared to manually defined IDPs, "delta-age" captures many signals into a composite measure of cardiovascular health, but still doesn't capture all the underlying biological variation visible in cMRI.

More recent studies have used unsupervised learning to extract novel phenotypes from imaging in large human cohorts, e.g. brain and cardiac MRI [27, 6, 26]. In these cases, a self-supervised model was trained on organ images or 3D meshes, then individual latent feature activations (e.g. from an autoencoder bottleneck) were used as novel IDPs for GWAS. These IDPs can capture subtle signals that supervised methods miss, but no principled ways were proposed to turn the representations into IDPs. In the cardiac MRI case, where we have videos rather than static images, existing

AI for New Drug Modalities at NeurIPS 2024.

representation-based IDPs also do not consider temporal dynamics or strain maps, which provide useful signals for characterizing cardiovascular disease.

Here, we apply unsupervised representation learning on time-series cardiac MRI from the UK BioBank (N∼68,000 subjects) using video masked auto-encoders [42]. We train models on multiple types of cMRI videos as well as their deformation fields (similar to optical flow maps), then fuse the resulting representations to maximally capture the spatiotemporal dynamics. We assess the utility of these representations using prediction of age and cardiac indices as a benchmark. We cluster the representations of each subject and identify clusters enriched for cardiovascular disease. Finally, we derive IDPs using delta-age and subject-to-cluster similarity, then perform GWAS, revealing genes associated with cardiovascular health, including both established and potentially novel hits. The overall approach is outlined in Figure 1.

2 Spatiotemporal Representation Learning on Cardiac Videos

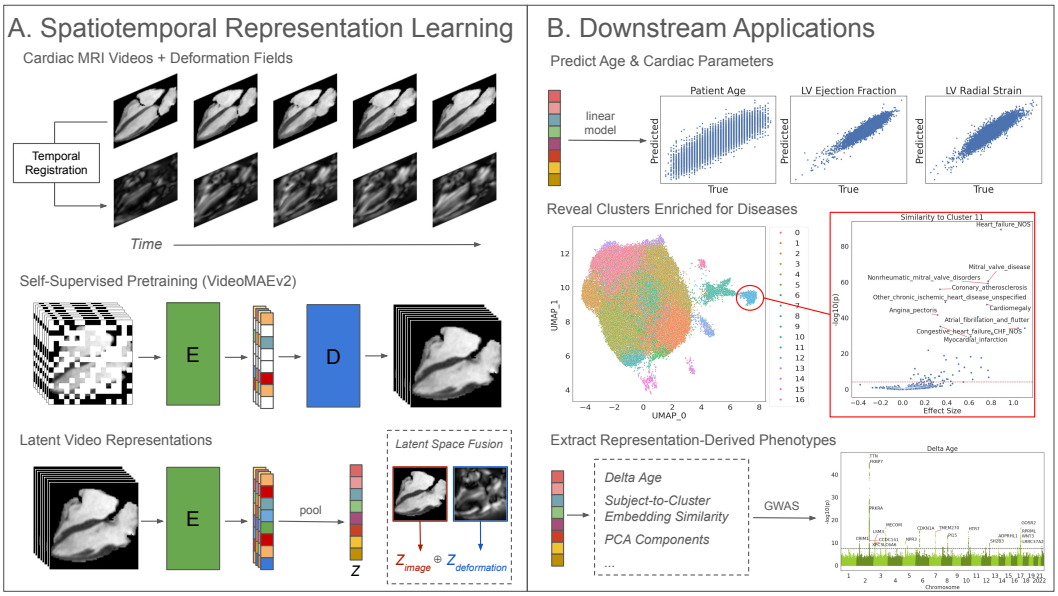

Figure 1: Overview of spatiotemporal modeling for genetic discovery. (A) We obtain cMRI videos for ∼68,000 subjects and calculate deformation fields corresponding to heart motion. We train a VideoMAE model (encoder E, decoder D), then unmasked videos are passed to the frozen encoder to obtain spatiotemporal patch representations, which are average-pooled to get a single video-level representation *Z*. We can augment the representation by combining latent features from multiple input signals (e.g. images plus deformation fields). (B) We use the representations for multiple downstream tasks - predicting age and cardiac function, clustering subjects (which reveals subpopulations associated with cardiovascular disease risk), and extracting representation-derived phenotypes for GWAS (which could enable the discovery of novel genetic targets).

2.1 Cardiac MRI Videos

We use data from the UK BioBank, a large prospective cohort with genotyping, EHR info, and deep phenotyping of 500k subjects from the UK [7]. A subset were imaged using multiple modalities including cardiac MRI [28]; at this time roughly ∼68k subjects have cMRI imaging available. Several types of cMRI imaging were collected; in this study we use the long axis 4-chamber (LA) and short axis (SA) scans. The LA-4ch scans capture a 2D video of a slice of the heart that includes a cross-section of all 4 chambers (left and right ventricles, left and right atria). The SA scans capture a 3D volumetric video that fully covers the left and right ventricles, but excludes the atria. These two scans have orthogonal imaging planes and different trade-offs, with the LA scans covering a larger field of view (FOV) but only at a 2D slice, while the SA scans are pseudo-3D but have a smaller FOV [11]. In our experiments, we assess the value of representations derived from both scan types. All

videos contain a fixed number of frames corresponding to a single heartbeat, with the first frame representing the end-diastole (ED) phase corresponding to maximum filling of the ventricles.

2.2 Temporal Deformation Fields

To identify salient motion features, we generated temporal deformation maps for each video as an additional input to our representation models, inspired by two-stream models that incorporate optical flow signals for video recognition [36]. The deformation maps are obtained by registering each frame of the video to the first frame (end-diastole), using VoxelMorph [3].

Formally, let $\boldsymbol{v_{i,t}} \in \mathcal{R}^{H \times W}$ represent frame *t* of video *i*. We estimate a deformation field that aligns frame 0 to frame *t* via a spatial transformer module [14]. The deformation field consists of a displacement vector $(\Delta x, \Delta y)$ at each pixel coordinate. We take the L2 norm of the displacement vector at each coordinate to flatten the deformation field to a 2D "magnitude image" $\boldsymbol{d_{i,t}} \in \mathcal{R}^{H \times W}$.

To predict the deformation fields, we first trained a VoxelMorph U-Net model following the general strategy and hyperparameters used in [3]. As input, we randomly sample pairs of video frames $(\boldsymbol{v_{i,0}}, \boldsymbol{v_{i,t}})$ and train the model to minimize mean-squared-error with a L2 regularization term on the deformation field magnitude. Given the trained model, we run inference on the entire cohort. At inference time, we compute the deformation magnitude $\boldsymbol{d_{i,t}}$ for each video frame *t*, then concatenate into a single video $\boldsymbol{d_i}$ with the same dimensions as the original video of pixel intensities $\boldsymbol{v_i}$. This deformation video captures dynamics in heart morphology that complement the original pixel intensities.

2.3 Self-Supervised Video Representation Learning

To extract a low-dimensional representation that captures spatiotemporal dynamics in the videos and deformation fields, we train a self-supervised model based on VideoMAEv2, which has shown excellent performance as a pre-training task for downstream video classification [42]. The model tokenizes a video of shape $H \times W \times T$ into a sequence of *N* spatiotemporal patch tokens of shape $H' \times W' \times T'$, embeds them into latent vectors of dimension *D*, then passes them through a sequence of vision transformer layers. During training, some tokens are randomly masked (using tube masking to mask the same patches across time), then a decoder learns to reconstruct the masked video tokens by minimising mean-squared-error loss. During inference, we pass the full (unmasked) video v_i through the encoder to obtain a sequence of spatiotemporal patch embeddings $z_{i,1}, ...z_{i,N}$ which are average-pooled to get a single video-level embedding $z_i \in \mathcal{R}^D$.

We train four VideoMAEv2 models, one each for long axis videos, long axis deformation fields, short axis videos, and short axis deformation fields. We then run inference to obtain four sets of embeddings corresponding to each of the four modalities. Since the model is designed for 2D videos, the short axis models are trained on randomly sampled 2D axial slices from underlying 3D volumes, and the inference is run on all z-slices followed by average-pooling so we always get an output embedding of the same dimension *D*. We can further combine information from multiple representations by concatenating their latent dimensions. For example, we can combine the embeddings of videos and deformation fields from the same modality, or combine across modalities (short and long axis). Note that we use the same datasets for training and inference, since the reconstruction objective using in training is not directly connected to any downstream evaluation tasks; this is a common setup in representation-based target discovery [6].

2.4 Benchmarking Video Representations

To evaluate the predictive power of the learned embeddings, we construct a benchmark task consisting of predicting age and cardiac function parameters via linear regression. Age is a good benchmark due to its strong association with cardiovascular health [23]. Cardiac parameters measured from cMRI are also strong predictors of heart function; for example, left ventricular ejection fraction (LVEF) is used in the diagnosis of heart failure with reduced ejection fraction (HFrEF). Critically, we choose cardiac parameters that are calculated from image features across multiple timepoints of the cardiac cycle. For example, ejection fractions (EF) are calculated by comparing the maximum and minimium volumes over time for a particular cardiac chamber. Likewise strain parameters like circumferential strain (CS) and longitudinal strain (LS) assess the contractility of the myocardial wall over time

across different axes. Thus, we hypothesize that models which can better capture spatiotemporal dynamics will do better at predicting these parameters.

3 Deriving Phenotypes from Representations

To test the utility of our embeddings for target discovery, we explored several ways to extract phenotypes from learned representations, which we term RDPs (representation-derived phenotypes).

3.1 Delta Age

First, we compute cardiovascular delta-age in a manner similar to previous work [33], but using video embeddings rather than manually-crafted IDPs as the input variables. In brief, we first train a linear model to obtain weights $\boldsymbol{w}$ and bias b that predict subjects' chronological age from their embedding:

$$PredictedAge_i = \boldsymbol{w}^T \boldsymbol{z_i} + b$$

We then subtract predicted from chronological age to get a raw delta-age. This raw delta-age value is still correlated with age, so we perform a correction by linear regressing true age from raw delta-age to obtain weight α and bias β, then subtract this prediction to get an unbiased delta-age phenotype:

$$DeltaAge_i = PredictedAge_i - (\alpha \cdot ChronologicalAge_i + \beta)$$

This delta-age phenotype captures the extent to which a subject's predicted age differs from their chronological age, based on features from our spatiotemporal embeddings. Similar formulations of delta-age in the literature have been linked to a variety of health outcomes [33, 37].

3.2 Clustering

Second, we cluster the subjects based on their embeddings to identify subpopulations with different heart characteristics. We hypothesize that some of these clusters will be enriched (either positively or negatively) for cardiovascular diseases. Instead of using a subject's cluster label as a categorical phenotype (which does not account for the complex hierarchical relationships between clusters and subjects), we derive a continuous phenotype based on subject-to-cluster similarity.

Formally, given an assignment of each subject *i* to a cluster $\mathcal{C}_k$ (where $k \in \{1...K\}$), we compute the mean embedding of each cluster:

$$\boldsymbol{c_k} = \frac{1}{|\mathcal{C}_k|} \sum_{i \in \mathcal{C}_k} \boldsymbol{z_i}$$

then compute the cosine similarity of each subject-level embedding to the cluster means, resulting in *K* phenotypes representing similarity of each subject *i* to each cluster *k*:

$$ClusterSimilarity_{k,i} = CosineSimilarity(\boldsymbol{c_k}, \boldsymbol{z_i})$$

This gives us *K* independent phenotypes, each one representing similarity to a prototypical heart embedding of a particular subpopulation inferred from clustering.

4 Experiments

4.1 Dataset Preprocessing

From the raw LA and SA videos, we first segmented the left + right ventricles and atria using a segmentation model described previously [2]. We then resampled the videos to a consistent spatial

resolution, cropped them to a fixed pixel dimension and normalized the pixel intensities to the range $[0, 1]$. We then compute temporal deformation maps for both LA and SA scans (as described in Section 2.2), by training a VoxelMorph model with default settings [3] on a subset of 512 subjects, then inferring on the whole cohort. Finally, we crop the videos and deformation maps using the bounding box of the segmentation mask, resize to 128x128 pixels, and mask the videos to the foreground pixels to emphasize the signal within the heart substructures. The temporal dimension is always 50 frames representing one full cardiac cycle, and is not altered. After postprocessing, we have 64,293 subjects with LA scans and 63,562 with SA scans, but for benchmarking tasks we only use the intersecting subset to maintain a fixed sample size.

4.2 Predicting Age and Cardiac Parameters

We benchmark and compare embeddings by linear probing on biologically meaningful parameters - patient age (which correlates strongly with cardiovascular health) and cardiac function indices such as ejection fractions and strain rates (which relate to cardiac dynamics). We obtain these parameters from published UKBB data fields. For age, use the patient age at the imaging visit, which is available for all 63,562 subjects in our benchmark cohort. For cardiac function parameters, we use published values [2], which are only available for 34,929 subjects in our cohort. We split the subjects into 75% training and 25% test. Note that the size of the training and test sets is larger for age prediction than cardiac parameter prediction, due to the differing number of available labels. For each target variable and input embedding, we train a ridge regression model using 5-fold cross validation (trying regularization strengths between $1...10^4$) to predict the parameter on the train set. The resulting models are run on the test sets and the R^2 values compared across embedding types.

Table 1: Evaluation of age and cardiac parameter prediction (R^2 on the test set) across input data modalities using VideoMAEv2. "SA/LA Concat" denotes combining video and deformation fields within short / long axis views respectively; "All Concat" denotes combining all 4 inputs. LVEF = left ventricular ejection fraction, RVEF = right ventricular ejection fraction, LAEF = left atrial ejection fraction, RAEF = right atrial ejection fraction, LVCO = left ventricular cardiac output, LVRS = left ventricular radial strain, LVCS = left ventricular circumferential strain, LVLS = left ventricular longitudinal strain.

Embedding Inputs	**Age**	**LVEF**	**RVEF**	**LAEF**	**RAEF**	**LVCO**	**LVRS**	**LVCS**	**LVLS**
LA Video	0.648	0.574	0.586	0.673	0.691	0.583	0.618	0.704	0.669
LA Deform	0.632	0.557	0.553	0.641	0.659	0.643	0.596	0.683	0.663
SA Video	0.713	0.822	0.795	0.484	0.388	0.845	0.783	0.892	0.497
SA Deform	0.698	0.745	0.738	0.475	0.404	0.810	0.767	0.863	0.487
LA Concat	0.681	0.583	0.593	0.678	0.705	0.666	0.630	0.712	0.711
SA Concat	0.750	**0.835**	**0.820**	0.504	0.422	**0.868**	**0.812**	**0.904**	0.515
All Concat	**0.773**	0.827	0.810	**0.692**	**0.709**	0.867	0.810	0.900	**0.732**

In Table 1, we evaluate representations from each modality (long axis, short axis), signal type (image, deformation) as well as hybrid representations from concatenating the latent dimensions. We use a fixed set of hyperparameters based on default settings from the VideoMAEv2 paper: ViT-base model architecture, patch size 8, sampling rate 3, and mask ratio 90%.

We find that the SA-derived representations do better on indices that are traditionally derived from SA scans (LVEF, RVEF, LVCO, LVRS, LVCS) and vice versa for LA (LAEF, RAEF, LVLS). Concatenating raw videos and deformation fields within each modality is always beneficial for performance. Finally, concatenating across views (SA, LA) is only beneficial for some metrics but not others. With this in mind, in our downstream analysis we concatenate the long axis image + deformations into one representation, and likewise for short axis, but do not mix short and long axis views.

4.3 Model Ablations

In Table 2, we conduct ablation studies by training multiple VideoMAEv2 models with different hyperparameters, focusing on one input modality (long-axis videos). For comparison, we also

Table 2: Evaluation of age and cardiac parameter prediction (R^2 on the test set) across models and hyperparameter settings, using long axis video embeddings as input. I-MAE = Image-MAE, V-MAE = VideoMAE. ResNet50 was trained on ImageNet classification; all other models were trained on cardiac videos in a self-supervised manner. See Table 1 for abbreviations. The row in *italics* denotes the model settings that were used for our other analyses; note that these settings don't correspond to the absolute best performance, but the difference with best-performing settings is marginal.

Model Arch	**Model Size**	**Patch Size**	**Sample Rate**	**Mask Ratio**	**Age**	**LVEF**	**LVCO**	**LVCS**	**LVLS**
ResNet50	-	-	-	-	0.352	0.229	0.310	0.346	0.281
I-MAE	base	8	-	75	0.506	0.374	0.494	0.495	0.446
V-MAEv2	small	8	3	90	0.591	0.549	0.498	0.673	0.640
V-MAEv2	base	8	2	90	0.641	0.571	0.584	0.698	0.669
V-MAEv2	base	8	3	75	0.651	0.575	0.586	0.706	**0.679**
V-MAEv2	*base*	*8*	*3*	*90*	*0.648*	*0.574*	*0.583*	*0.704*	*0.669*
V-MAEv2	base	16	3	90	0.605	0.552	0.484	0.679	0.635
V-MAEv2	base	8	6	90	0.635	0.566	0.572	0.692	0.669
V-MAEv2	base	8	3	95	0.626	0.567	0.576	0.696	0.663
V-MAEv2	large	8	3	90	**0.657**	**0.581**	**0.600**	**0.707**	0.677

train a conventional ImageMAE model [13] on the same inputs, treating each timeframe as an independent image. The ImageMAE uses a comparable architecture to our VideoMAE "base" model (same encoder and decoder dimensions, depth and number of heads). As an additional baseline not trained on cardiac videos, we include ResNet50 [12] pretrained on ImageNet classification [9], with weights obtained from the *torchvision* package, using the activations of the last layer as an embedding. For these static image models, each 2D frame is encoded separately and the embeddings are average-pooled across time.

We expect the ResNet50 model to extract general texture features, but nothing specific to cardiac MRI (since it was trained on natural images). Conversely, we expect ImageMAE to capture cardiac-specific texture (since it is trained on our dataset), but to lack explicit modeling of temporal dependencies. Accordingly, we see that ResNet50 performs worst, and static MAE performs better but does not meet the performance of any VideoMAE model (even ViT-small) across almost all benchmark tasks.

For VideoMAE models, going from a "small" to "base" ViT architecture yields substantial improvement, while going from "base" to "large" gives marginal improvement at the cost of much longer training time. Likewise increasing patch size from 8 to 16 leads to a performance drop. Sample rate denotes how many video frames are skipped when sampling; reducing from 6 to 3 improves performance slightly, but further reducing to 2 does not. Finally, increasing the mask ratio to 95% slightly harms performance and reducing it to 75% slightly improves performance.

4.4 Clustering Analysis

In Figure 2, we visualize both short and long axis embeddings from the whole cohort using UMAP [17], overlaying the labels from Leiden clustering [40].

Prior to clustering, we first remove confounding effects from the embeddings by regressing out the effects of age, sex and BMI using linear regression, as implemented in the *scanpy* package [45]. Then we run nearest neighbor connectivity (using cosine distance metric) and Leiden clustering. The resulting number of clusters K is determined automatically.

We find that there are some clear outlier clusters in the long axis embeddings, while the short axis clusters are less separated. We hypothesize that the long axis embeddings (which derive from a single cross-sectional slice of both atria and ventricles) may encode more population-level variation than the short axis embeddings (which are averaged over many slices, some of which only cover a small portion of ventricular morphology, and don't cover atria at all). We therefore focus on long axis embeddings in the subsequent analysis.

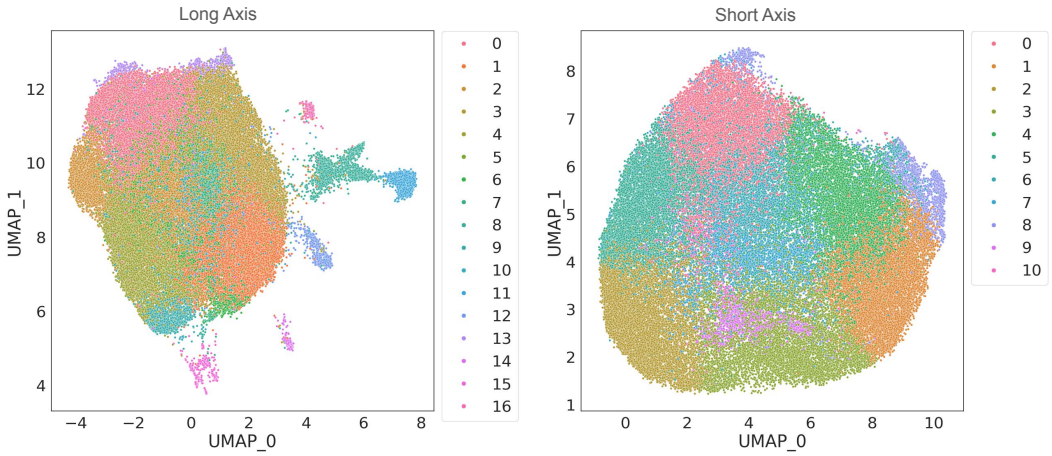

Figure 2: UMAP and clustering of long axis and short axis embeddings. Long axis shows more separation of outlier clusters, which we subsequently interrogate with PheWAS and GWAS.

4.5 Phenome-Wide Association Studies

We conduct phenome-wide association studies (PheWAS) on the derived RDPs to better understand their associations with traditional phenotypes. For simplicity, we restrict the set of traditional phenotypes to ICD-10 codes reported in the UK BioBank, which roughly correspond to diagnoses of medical conditions. Following the general approach of the *PHESANT* package [20], we first preprocess each RDP by regressing out the effect of sex, age, BMI and imaging center, then applying a rank-inverse normal transform so that the RDP follows a normal distribution. Then, for each combination of RDP and ICD-10 code, we perform logistic regression to obtain an effect size and p-value indicating the strength of the association between the RDP and ICD-10 code.

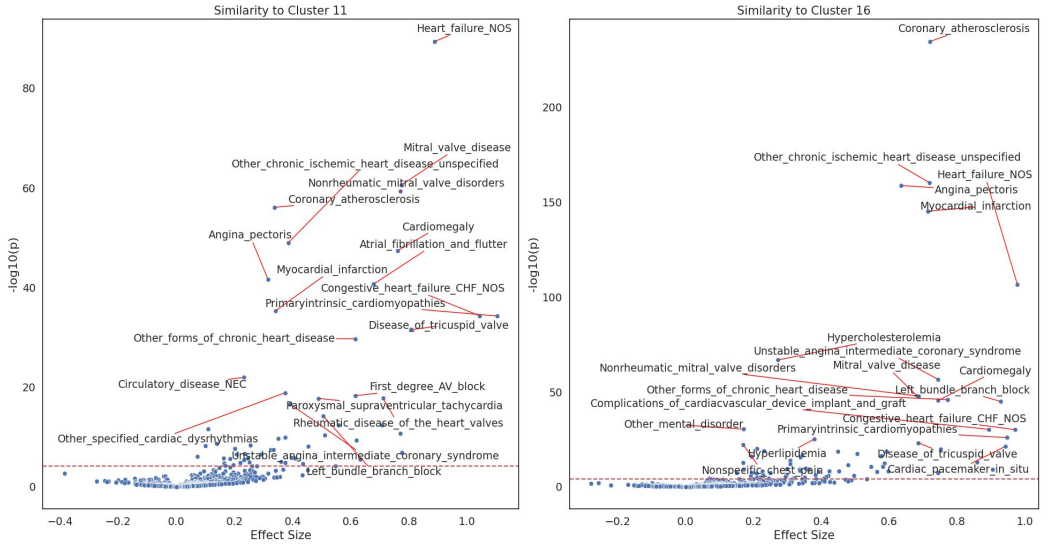

Figure 3: Example volcano plots from ICD-10 PheWAS to identify diseases associated with embedding similarity RDPs for two clusters identified as outliers in the long-axis embeddings (see Figure 2). The RDPs show strong association with cardiovascular disease, suggesting that the representations extract biologically meaningful signal. Dashed line corresponds to Bonferroni-corrected significance threshold.

Figure 3 shows example PheWAS results from long-axis cluster RDPs (Section 3.2) corresponding to two outlier clusters seen in Figure 2. Several clusters are highly enriched for phenotypes related to

cardiometabolic disease, including heart failure, myocardial infarction, atrial fibrillation, coronary atherosclerosis, mitral valve disorders, high cholesterol, high blood pressure, left bundle branch block, angina and type 2 diabetes. Moreover there are distinct enrichments in multiple clusters, suggesting multiple subtypes for diseases like heart failure. Full results are reported in the appendix.

4.6 Genome-Wide Association Studies

We run genome-wide association studies (GWAS) on each RDP to identify genetic associations with latent spatiotemporal features in cMRI videos, using the UKBB genotyping data described in [7]. We follow the genotype imputation, filtering and QC procedure described in [15]. We use the GWAS implementation in *regenie* [16], adjusting for sex, age, age^2, imaging center, genotyping array and top 10 genetic principal components. After applying subject-level filtering from [15], we are left with N=62,605 subjects. We filter variants using simple peak detection with a minimum genomic distance of 30 between neighboring peaks. As this is a proof-of-concept analysis, we do not perform LD-score regression or fine-mapping, but acknowledge their importance for further validation of the genetic associations. We then use variant-to-gene annotations from OpenTargets [25] to identify the most likely gene associated with each variant. We report all hits with a p-value below the base genome-wide significance threshold for the number of SNPs being tested. Since we are testing multiple phenotypes, we also report a Bonferroni-adjusted p-value threshold, but note that this may be overly conservative given potential correlations between phenotypes.

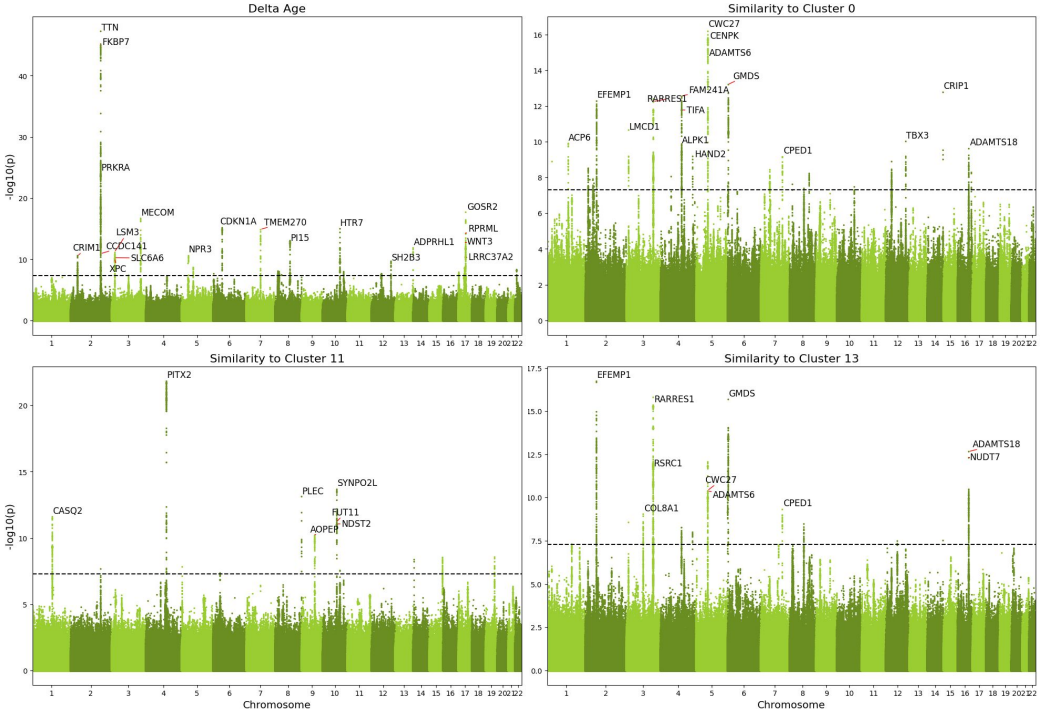

Figure 4: Manhattan plots of delta-age and cluster-mean similarities for 3 example clusters derived from long axis embeddings. For clarity, we only label peak hits for each gene where $-log_{10}(p) >= 9$. Dashed line corresponds to genome-wide significance threshold.

Here, we focus on results from the long-axis RDPs (based on combined video + deformation embeddings). A larger list of hits from both long and short axis phenotypes are listed in the appendix. Some example Manhattan plots are shown in Figure 4. In total we find 146 significant genes using the base genome-wide significance threshold of $p <= 5.0 \times 10^{-8}$. However, note that we are testing 18 phenotypes (1 delta-age + 17 clustering); with the Bonferroni-adjusted value of $p <= 2.8 \times 10^{-9}$, there are 74 significant genes. We examine some of the more significant and interesting hits below.

4.6.1 GWAS Results for Delta Age

Several top delta-age hits are related to TTN, which controls myocardial contractility and is strongly associated with dilated cardiomyopathy [43]. It has also been associated with atrial fibrillation [32] and left ventricular parameters [29]. Nearby hits that are assigned to FKBP7 and PRKRA by OpenTargets, but also linked to TTN, have been linked to musculoskeletal wound healing phenotypes [5] and hypertrophic cardiomyopathy [19]. GOSR2, CDKN1A, ADPRHL1 and LSM3 have all been linked to phenotypes like QRS duration [31], PR interval [24] and electrocardiogram morphology [41]. MECOM is associated with hypertension and blood pressure phenotypes [4]. HTR7 has no reported association with cardiovascular disease, but encodes a 5-HT serotonin receptor which is able to control cardiac contraction and may have a role in arrhythmias [21]. We also reproduce several of the top genes reported in prior studies of delta-age [33], including TTN and PI15. We also find a variant assigned to TMEM270 which is proximal to ELN (another gene reported in that study).

4.6.2 GWAS Results Derived by Clustering

Our clustering hits capture established associations with cardiovascular disease, like PITX2 (atrial fibrillation [32]) and SYNPO2L (atrial fibrillation [8] and heart failure [34]). Likewise variants in FADS1 and TMEM258 have been linked to high-density lipoprotein, red cell distribution width, and blood lipid levels in general [4]. CCDC141 is proximal to TTN and has been associated with pulse pressure [38]. ADAMTS6 has been reported in studies of cardiac conduction [31] and is an emerging gene of interest [18]. Variants assigned to CWC27 and CENPK have less cardiovascular evidence but are proximal to ADAMTS6. ADAMTS18 may play a role in maintaining haemostatic balance and could be connected to atrial fibrillation and thrombus [44]. EFEMP1 has been associated with maximum left atrial volume, a potential marker of diastolic dysfunction [39], and the encoded protein is also linked to adverse outcomes in heart failure [10]. Finally, GMDS and RARRES1 do not have much evidence linking to cardiovascular disease, suggesting a need for further exploration.

5 Discussion

Representation learning on large biomedical imaging datasets enables the discovery of new biologically relevant features in an unsupervised and unbiased manner. When combined with genetic data, this approach shows great promise for target identification. Timeseries cMRI captures spatiotemporal dynamics, which are essential for characterizing cardiovascular diseases due to their inherently temporal nature. Though previous work used representations from static short axis heart meshes to identify targets [6], and other works have trained a VideoMAE on long axis videos for cardiac function prediction [35], we are the first to our knowledge to use VideoMAE representations for target discovery, as well as first to incorporate complete cMRI videos combined with their deformation fields, from multiple cardiac views, to better encode the underlying dynamics of heart function.

Our benchmarking experiments suggest that video representations can capture age and cardiac parameters with high coefficient of determination. The ability to predict measures of cardiac dynamics (like ejection fraction) suggests that the representations encode spatiotemporal information. Deformation field magnitude is overall almost as good as pixel-space information when used as an input to the representation models, and there is a small but consistent benefit to concatenating the representations. Further, VideoMAE performs better than regular MAE or a pretrained ResNet, suggesting that self-supervised video modeling outperforms static image modeling. Future work could explore more sophisticated methods for fusing pixel-space and deformation fields, as well as improved modeling of temporal dynamics in the representations (rather than averaging over time frames).

Clustering the representations reveals enrichment of cardiovascular phenotypes in specific clusters. Our analysis is limited to ICD-10 codes, which are based on billing codes and generally under-estimate the true disease prevalence, so exploration of other phenotypes would be valuable. Clustering based on image-derived representations could enable subtyping of diseases and discovering at-risk patients.

Our GWAS on RDPs (delta-age and subject-to-cluster similarity) captures genes with known associations to cardiomyopathy (TTN), atrial fibrillation (PITX2, SYNPO2L), cardiac conduction (e.g. GOSR2, CDKN1A) and other cardiovascular phenotypes. We also find genes with less established connection to cardiovascular disease (e.g. HTR7, RARRES1, GMDS) which suggest further study.

While prediction benchmarks, PheWAS and GWAS are a useful starting point for validating our representations, further work is needed to define proper benchmarks for evaluating the expressiveness of representation models trained on large human cohorts.

Acknowledgements

This research was conducted using data from UK BioBank, application number 18448.

References

[1] N. Aung et al. Genome-Wide Analysis of Left Ventricular Image-Derived Phenotypes Identifies Fourteen Loci Associated With Cardiac Morphogenesis and Heart Failure Development. *Circulation*, 140(16):1318–1330, Oct. 2019. doi: 10.1161/CIRCULATIONAHA.119.041161. URL https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.119.041161. Publisher: American Heart Association.

[2] W. Bai et al. A population-based phenome-wide association study of cardiac and aortic structure and function. *Nature Medicine*, 26(10):1654–1662, Oct. 2020. ISSN 1546-170X. doi: 10.1038/s41591-020-1009-y. URL https://www.nature.com/articles/s41591-020-1009-y. Publisher: Nature Publishing Group.

[3] G. Balakrishnan et al. An Unsupervised Learning Model for Deformable Medical Image Registration. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9252–9260, June 2018. doi: 10.1109/CVPR.2018.00964. URL https://ieeexplore.ieee.org/document/8579062. ISSN: 2575-7075.

[4] A. R. Barton et al. Whole-exome imputation within UK Biobank powers rare coding variant association and fine-mapping analyses. *Nature genetics*, 53(8):1260–1269, Aug. 2021. ISSN 1546-1718. doi: 10.1038/s41588-021-00892-1. URL https://europepmc.org/articles/PMC8349845.

[5] P. Baumert et al. Polygenic mechanisms underpinning the response to exercise-induced muscle damage in humans: In vivo and in vitro evidence. *Journal of Cellular Physiology*, 237(7):2862–2876, 2022. ISSN 1097-4652. doi: 10.1002/jcp.30723. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jcp.30723.

[6] R. Bonazzola et al. Unsupervised ensemble-based phenotyping enhances discoverability of genes related to left-ventricular morphology. *Nature Machine Intelligence*, 6(3):291–306, Mar. 2024. ISSN 2522-5839. doi: 10.1038/s42256-024-00801-1. URL https://www.nature.com/articles/s42256-024-00801-1. Publisher: Nature Publishing Group.

[7] C. Bycroft et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 562(7726):203–209, Oct. 2018. ISSN 1476-4687. doi: 10.1038/s41586-018-0579-z. URL https://www.nature.com/articles/s41586-018-0579-z. Publisher: Nature Publishing Group.

[8] A. G. Clausen et al. Loss-of-Function Variants in the SYNPO2L Gene Are Associated With Atrial Fibrillation. *Frontiers in Cardiovascular Medicine*, 8:650667, Mar. 2021. ISSN 2297-055X. doi: 10.3389/fcvm.2021.650667. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7985167/.

[9] J. Deng et al. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, June 2009. doi: 10.1109/CVPR.2009.5206848. URL https://ieeexplore.ieee.org/document/5206848. ISSN: 1063-6919.

[10] M. Dib et al. Proteomic Associations of Adverse Outcomes in Human Heart Failure. *Journal of the American Heart Association*, 13(5):e031154, Mar. 2024. doi: 10.1161/JAHA.123.031154. URL https://www.ahajournals.org/doi/full/10.1161/JAHA.123.031154. Publisher: Wiley.

[11] D. T. Ginat et al. Cardiac Imaging: Part 1, MR Pulse Sequences, Imaging Planes, and Basic Anatomy. *American Journal of Roentgenology*, 197(4):808–815, 2011. doi: 10.2214/AJR.10.7231. URL https://doi.org/10.2214/AJR.10.7231.

[12] K. He et al. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, June 2016. doi: 10.1109/CVPR.2016.90. URL https://ieeexplore.ieee.org/document/7780459. ISSN: 1063-6919.

[13] K. He et al. Masked Autoencoders Are Scalable Vision Learners. pages 16000–16009, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/He_Masked_Autoencoders_Are_Scalable_Vision_Learners_CVPR_2022_paper.

[14] M. Jaderberg et al. Spatial Transformer Networks. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper_files/paper/2015/hash/33ceb07bf4eeb3da587e268d663aba1a-Abstract.html.

[15] Y. Liu et al. Genetic architecture of 11 organ traits derived from abdominal MRI using deep learning. *eLife*, 10:e65554, June 2021. ISSN 2050-084X. doi: 10.7554/eLife.65554. URL https://doi.org/10.7554/eLife.65554. Publisher: eLife Sciences Publications, Ltd.

[16] J. Mbatchou et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nature Genetics*, 53(7):1097–1103, July 2021. ISSN 1546-1718. doi: 10.1038/s41588-021-00870-7. URL https://www.nature.com/articles/s41588-021-00870-7. Publisher: Nature Publishing Group.

[17] L. McInnes et al. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, Sept. 2020. URL http://arxiv.org/abs/1802.03426. arXiv:1802.03426 [cs, stat].

[18] T. J. Mead. ADAMTS6: Emerging roles in cardiovascular, musculoskeletal and cancer biology. *Frontiers in Molecular Biosciences*, 9, Oct. 2022. ISSN 2296-889X. doi: 10.3389/fmolb.2022.1023511. URL https://www.frontiersin.org/journals/molecular-biosciences/articles/10.3389/fmolb.2022.1023511/full. Publisher: Frontiers.

[19] R. Mendes de Almeida et al. Whole gene sequencing identifies deep-intronic variants with potential functional impact in patients with hypertrophic cardiomyopathy. *PLoS ONE*, 12(8):e0182946, Aug. 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0182946. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5552324/.

[20] L. A. Millard et al. Software Application Profile: PHESANT: a tool for performing automated phenome scans in UK Biobank. *International Journal of Epidemiology*, 47(1):29–35, Oct. 2017. ISSN 0300-5771. doi: 10.1093/ije/dyx204. URL https://doi.org/10.1093/ije/dyx204.

[21] J. Neumann et al. Cardiac Roles of Serotonin (5-HT) and 5-HT-Receptors in Health and Disease. *International Journal of Molecular Sciences*, 24(5):4765, Jan. 2023. ISSN 1422-0067. doi: 10.3390/ijms24054765. URL https://www.mdpi.com/1422-0067/24/5/4765. Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.

[22] C. Ning et al. Genome-wide association analysis of left ventricular imaging-derived phenotypes identifies 72 risk loci and yields genetic insights into hypertrophic cardiomyopathy. *Nature Communications*, 14(1):7900, Nov. 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-43771-5. URL https://www.nature.com/articles/s41467-023-43771-5. Publisher: Nature Publishing Group.

[23] B. J. North et al. The Intersection Between Aging and Cardiovascular Disease. *Circulation Research*, 110(8):1097–1108, 2012. doi: 10.1161/CIRCRESAHA.111.246876. URL https://www.ahajournals.org/doi/abs/10.1161/CIRCRESAHA.111.246876.

[24] I. Ntalla et al. Multi-ancestry GWAS of the electrocardiographic PR interval identifies 202 loci underlying cardiac conduction. *Nature communications*, 11(1):2542, May 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-15706-x. URL https://europepmc.org/articles/PMC7242331.

[25] D. Ochoa et al. The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. *Nucleic Acids Research*, 51(D1):D1353–D1359, Nov. 2022. ISSN 0305-1048. doi: 10.1093/nar/gkac1046. URL https://doi.org/10.1093/nar/gkac1046.

[26] S. Ometto et al. Unsupervised cardiac MRI phenotyping with 3D diffusion autoencoders reveals novel genetic insights, Nov. 2024. URL https://www.medrxiv.org/content/10.1101/2024.11.04.24316700v1. Pages: 2024.11.04.24316700.

[27] K. Patel et al. Unsupervised deep representation learning enables phenotype discovery for genetic association studies of brain imaging. *Communications Biology*, 7(1):414, Apr. 2024. ISSN 2399-3642. doi: 10.1038/s42003-024-06096-7. URL https://doi.org/10.1038/s42003-024-06096-7.

[28] S. E. Petersen et al. Imaging in population science: cardiovascular magnetic resonance in 100,000 participants of UK Biobank - rationale, challenges and approaches. *Journal of Cardiovascular Magnetic Resonance*, 15(1):46, May 2013. ISSN 1532-429X. doi: 10.1186/1532-429X-15-46. URL https://doi.org/10.1186/1532-429X-15-46.

[29] J. P. Pirruccello et al. Analysis of cardiac magnetic resonance imaging in 36,000 individuals yields genetic insights into dilated cardiomyopathy. *Nature Communications*, 11(1):2254, May 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-15823-7. URL https://www.nature.com/articles/s41467-020-15823-7. Publisher: Nature Publishing Group.

[30] J. P. Pirruccello et al. Genetic analysis of right heart structure and function in 40,000 people. *Nature Genetics*, 54(6):792–803, June 2022. ISSN 1546-1718. doi: 10.1038/s41588-022-01090-3. URL https://www.nature.com/articles/s41588-022-01090-3. Publisher: Nature Publishing Group.

[31] B. P. Prins et al. Exome-chip meta-analysis identifies novel loci associated with cardiac conduction, including ADAMTS6. *Genome biology*, 19(1):87, July 2018. ISSN 1474-760X. doi: 10.1186/s13059-018-1457-6. URL https://europepmc.org/articles/PMC6048820.

[32] C. Roselli et al. Multi-ethnic genome-wide association study for atrial fibrillation. *Nature genetics*, 50(9): 1225–1233, June 2018. ISSN 1546-1718. doi: 10.1038/s41588-018-0133-9. URL https://europepmc.org/articles/PMC6136836.

[33] M. Shah et al. Environmental and genetic predictors of human cardiovascular ageing. *Nature Communications*, 14(1):4941, Aug. 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-40566-6. URL https://www.nature.com/articles/s41467-023-40566-6. Publisher: Nature Publishing Group.

[34] S. Shah et al. Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nature Communications*, 11(1):163, Jan. 2020. ISSN 2041-1723. doi: 10.1038/s41467-019-13690-5. URL https://www.nature.com/articles/s41467-019-13690-5. Publisher: Nature Publishing Group.

[35] C. Shen et al. Spatiotemporal Representation Learning for Short and Long Medical Image Time Series. In M. G. Linguraru, Q. Dou, A. Feragen, S. Giannarou, B. Glocker, K. Lekadir, and J. A. Schnabel, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, pages 656–666, Cham, 2024. Springer Nature Switzerland. ISBN 978-3-031-72120-5. doi: 10.1007/978-3-031-72120-5_61.

[36] K. Simonyan et al. Two-Stream Convolutional Networks for Action Recognition in Videos. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL https://proceedings.neurips.cc/paper/2014/hash/00ec53c4682d36f5c4359f4ae7bd7ba1-Abstract.html.

[37] S. M. Smith et al. Estimation of brain age delta from brain imaging. *NeuroImage*, 200:528–539, 2019. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2019.06.017. URL https://www.sciencedirect.com/science/article/pii/S1053811919305026.

[38] P. Surendran et al. Discovery of rare variants associated with blood pressure regulation through meta-analysis of 1.3 million individuals. *Nature Genetics*, 52(12):1314–1332, Dec. 2020. ISSN 1546-1718. doi: 10.1038/s41588-020-00713-x. URL https://www.nature.com/articles/s41588-020-00713-x. Publisher: Nature Publishing Group.

[39] M. Thanaj et al. Genetic and environmental determinants of diastolic heart function. *Nature Cardiovascular Research*, 1(4):361–371, Apr. 2022. ISSN 2731-0590. doi: 10.1038/s44161-022-00048-2. URL https://www.nature.com/articles/s44161-022-00048-2. Publisher: Nature Publishing Group.

[40] V. A. Traag et al. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, 9(1):5233, Mar. 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-41695-z. URL https://www.nature.com/articles/s41598-019-41695-z. Publisher: Nature Publishing Group.

[41] N. Verweij et al. The Genetic Makeup of the Electrocardiogram. *Cell systems*, 11(3):229–238.e5, Sept. 2020. ISSN 2405-4720. doi: 10.1016/j.cels.2020.08.005. URL https://europepmc.org/articles/PMC7530085.

[42] L. Wang et al. VideoMAE V2: Scaling Video Masked Autoencoders with Dual Masking. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14549–14560, June 2023. doi: 10.1109/CVPR52729.2023.01398. URL https://ieeexplore.ieee.org/document/10203656. ISSN: 2575-7075.

[43] J. S. Ware et al. Role of titin in cardiomyopathy: from DNA variants to patient stratification. *Nature Reviews Cardiology*, 15(4):241–252, Apr. 2018. ISSN 1759-5010. doi: 10.1038/nrcardio.2017.190. URL https://www.nature.com/articles/nrcardio.2017.190. Publisher: Nature Publishing Group.

[44] J. Wei et al. ADAMTS-18: A metalloproteinase with multiple functions. *Frontiers in bioscience (Landmark edition)*, 19:1456–1467, June 2014. ISSN 1093-4715. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4410865/.

[45] F. A. Wolf et al. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology*, 19(1): 15, Feb. 2018. ISSN 1474-760X. doi: 10.1186/s13059-017-1382-0. URL https://doi.org/10.1186/s13059-017-1382-0.

A Appendix / supplemental material

A.1 Comparison to Prior Work on Age Prediction

As another benchmark, we compare our age prediction model to prior work that predicted cardiovascular age from manually defined cMRI phenotypes [33]. Their analysis is performed on a subset of 5063 "healthy" participants from the UK BioBank imaging cohort. They define "healthy" as the absence of cardiovascular, metabolic or respiratory disease (as defined by presence of disease codes or self-reported illness) with BMI under 30. They then split the data into 80% train / 20% test and trained a gradient boosting model from a series of manually extracted IDPs. Although we do not have access to the list of subjects they used for the train/test split, we followed their data filtering procedure to define a comparable set of 5063 "healthy" participants split into 80% train / 20% test, then trained a ridge regression model from our combined long + short axis embeddings (as described in Section 4.2). Results are reported in Table S1; although the test sets are not exactly comparable, our model shows significantly better performance, further validating our approach.

Table S1: Comparison of age prediction from cMRI features on a healthy subcohort. MAE = mean absolute error.

Age Prediction Model	**Test R^2**	**Test MAE**
Shah et al. 2023	0.49	4.21
Ours	**0.63**	**3.65**

A.2 PheWAS Hits

Tables S2 and S3 list the top ICD-10 disease code hits for RDPs derived from long axis and short axis (respectively). For each ICD-10 code, we count the number of RDPs (delta age + K cluster similarities) with a significance of $p <= 0.05/635$ (applying a Bonferroni correction for the number of ICD-10 codes are being tested). We report the most significant RDP association and the total number of significant associations.

A.3 GWAS Hits

Tables S4 and S5 list the top gene hits for RDPs derived from long axis and short axis (respectively). Similiar to the PheWAS results, for each gene we count the number of RDPs (delta age + K cluster similarities) with a significance of $p <= 5.0 \times 10^{-8}$ (applying a Bonferroni correction based on the number of variants being tested). We report the lead variant, most significant RDP association, and the total number of significant associations.

Table S2: Top 50 aggregated PheWAS hits for representation-derived phenotypes from long-axis scans, in descending order of significance.

ICD-10 Code Name	Max -log10(p)	Most Sig RDP	# Sig RDPs
Coronary atherosclerosis	234.591	cluster16	13
Other chronic ischemic heart disease unspecified	160.224	cluster16	12
Angina pectoris	158.675	cluster16	12
Myocardial infarction	145.149	cluster16	11
Left bundle branch block	144.128	cluster12	7
Heart failure NOS	106.491	cluster16	16
Hypercholesterolemia	66.711	cluster16	12
Mitral valve disease	60.536	cluster11	13
Nonrheumatic mitral valve disorders	59.261	cluster11	13
Unstable angina intermediate coronary syndrome	56.568	cluster16	9
Other forms of chronic heart disease	53.874	cluster12	10
Cardiomegaly	47.382	cluster11	13
Atrial fibrillation and flutter	40.643	cluster11	8
Primaryintrinsic cardiomyopathies	35.020	cluster12	7
Congestive heart failure CHF NOS	34.297	cluster11	11
Disease of tricuspid valve	31.507	cluster11	10
Other mental disorder	30.315	cluster16	8
Complications of cardiacvascular device implant and graft	30.195	cluster16	4
Diaphragmatic hernia	28.810	cluster6	4
Hyperlipidemia	25.090	cluster16	7
Nonrheumatic aortic valve disorders	24.460	cluster12	4
Nonspecific chest pain	22.158	cluster16	4
Congenital anomalies of great vessels	21.951	cluster12	3
Circulatory disease NEC	21.933	cluster11	6
Cardiac pacemaker in situ	21.198	cluster16	6
Tobacco use disorder	20.796	cluster6	9
Obesity	20.770	cluster13	6
Type 2 diabetes	20.078	cluster16	10
Rheumatic disease of the heart valves	19.691	cluster16	8
Other specified cardiac dysrhythmias	18.839	cluster11	9
Other acute and subacute forms of ischemic heart disease	18.694	cluster16	3
Asthma	18.274	cluster2	8
First degree AV block	18.247	cluster11	5
Paroxysmal supraventricular tachycardia	17.720	cluster11	4
Chronic renal failure CKD	17.522	cluster16	6
Aortic valve disease	16.727	cluster12	5
Acute renal failure	16.475	cluster16	7
Chronic airway obstruction	15.344	cluster8	11
Cardiac arrest	13.046	cluster16	5
Atrioventricular block complete	12.909	cluster12	5
Peripheral vascular disease unspecified	12.360	cluster16	3
Hemoptysis	11.935	cluster16	5
GERD	11.431	cluster6	4
Aneurysm and dissection of heart	11.031	cluster12	6
Reflux esophagitis	10.888	cluster7	4
Occlusion and stenosis of precerebral arteries	10.626	cluster16	3
Pericarditis	10.425	cluster16	3
Premature beats	10.296	cluster11	3
Occlusion of cerebral arteries	9.644	cluster11	6
Other anemias	9.524	cluster12	2

Table S3: Top 50 aggregated PheWAS hits for representation-derived phenotypes from short-axis scans, in descending order of significance.

ICD-10 Code Name	Max -log10(p)	Most Sig RDP	# Sig RDPs
Coronary atherosclerosis	61.241	cluster10	7
Heart failure NOS	53.007	cluster9	5
Other chronic ischemic heart disease unspecified	48.865	cluster10	6
Myocardial infarction	43.924	cluster10	5
Left bundle branch block	42.007	cluster10	2
Hypercholesterolemia	34.905	cluster10	10
Type 2 diabetes	34.259	delta_age	8
Angina pectoris	33.614	cluster10	8
Mitral valve disease	32.618	cluster9	4
Nonrheumatic mitral valve disorders	32.040	cluster9	4
Other mental disorder	23.308	delta_age	6
Tobacco use disorder	23.015	delta_age	7
Congestive heart failure CHF NOS	22.467	cluster9	3
Primaryintrinsic cardiomyopathies	21.931	cluster9	2
Other forms of chronic heart disease	21.915	cluster10	5
Chronic airway obstruction	21.380	delta_age	7
Disease of tricuspid valve	20.297	cluster9	2
Obesity	19.619	cluster2	4
Cardiomegaly	18.697	cluster9	5
First degree AV block	18.216	cluster9	2
Diverticulosis	13.833	delta_age	6
Diaphragmatic hernia	13.735	delta_age	3
Right bundle branch block	13.671	cluster9	2
Circulatory disease NEC	12.938	cluster9	3
Atrial fibrillation and flutter	12.455	cluster9	1
Hyposmolality andor hyponatremia	11.922	delta_age	1
Nonspecific chest pain	11.711	delta_age	3
Other specified cardiac dysrhythmias	11.208	cluster9	2
Unstable angina intermediate coronary syndrome	10.714	cluster10	4
Rheumatic disease of the heart valves	10.366	cluster9	3
Emphysema	9.841	delta_age	5
Hyperlipidemia	9.619	cluster10	4
Acute renal failure	9.388	cluster10	2
Congenital anomalies of great vessels	8.775	cluster6	3
Complications of cardiacvascular device implant and graft	8.700	cluster10	2
Hypopotassemia	8.644	cluster10	1
Type 1 diabetes	8.445	delta_age	1
Nonrheumatic aortic valve disorders	7.939	cluster6	3
Septicemia	7.737	delta_age	1
Cardiac pacemaker in situ	7.366	cluster10	1
GERD	7.328	delta_age	2
Atrioventricular block complete	7.285	cluster9	1
Bundle branch block	7.257	cluster9	1
Hemoptysis	7.187	delta_age	2
Obstructive chronic bronchitis	7.006	delta_age	6
Cardiac arrest	6.890	cluster9	1
Sepsis	6.869	delta_age	1
Alcoholrelated disorders	6.452	delta_age	2
Premature beats	6.446	cluster9	1
Other acute and subacute forms of ischemic heart disease	6.233	cluster10	2

Table S4: Top 50 aggregated GWAS hits for representation-derived phenotypes from long-axis scans, in descending order of significance.

Gene Name	Lead Variant	Max -log10(p)	# Significant RDPs	Most Significant RDP
TTN	rs2042995	47.328	4	delta_age
FKBP7	rs1001238	45.229	3	delta_age
PRKRA	rs2253324	24.008	1	delta_age
PITX2	rs4611994	21.857	4	cluster11
GOSR2	rs17608766	17.731	1	delta_age
EFEMP1	rs59985551	16.749	3	cluster13
MECOM	rs9850919	16.685	1	delta_age
CWC27	rs2278353	16.201	6	cluster0
CCDC141	rs17362588	15.792	4	cluster4
CENPK	rs1309553	15.781	4	cluster0
GMDS	rs767102318	15.7078	5	cluster1
RARRES1	rs12637678	15.358	4	cluster13
CDKN1A	rs113578873	15.1916	1	delta_age
ADAMTS6	rs4700662	15.0322	4	cluster1
HTR7	rs10748555	15.0027	2	delta_age
TMEM270	rs370616120	14.8768	1	delta_age
FADS1	rs174566	14.4253	1	cluster3
RPRML	rs145153053	14.2427	1	delta_age
TMEM258	rs174533	13.9727	1	cluster3
SYNPO2L	rs60632610	13.7004	1	cluster11
PLEC	rs11786896	13.1577	4	cluster11
PI15	rs2732010	13.0876	1	delta_age
FADS2	rs97384	12.917	1	cluster3
CRIP1	rs55633823	12.7888	2	cluster0
WNT3	rs75230966	12.7247	1	delta_age
ADAMTS18	rs17689197	12.6626	3	cluster13
CCDC91	rs143823594	12.6286	4	cluster12
FAM241A	rs11376680	12.5597	3	cluster0
NUDT7	rs62043885	12.3228	3	cluster13
CPED1	rs3801387	12.2871	4	cluster1
ADPRHL1	rs76382172	11.8723	1	delta_age
TIFA	rs326847	11.7791	3	cluster0
RSRC1	rs73164066	11.6457	3	cluster13
CASQ2	rs10157905	11.6026	1	cluster11
HHIP	rs1512288	11.5656	2	cluster3
FUT11	rs11000771	11.2002	1	cluster11
LSM3	rs13061705	11.1621	1	delta_age
ZFPM1	rs12595858	11.0858	2	cluster4
NDST2	rs4746151	11.0666	1	cluster11
COL8A1	rs114796243	10.8133	3	cluster5
LMCD1	rs165177	10.6812	3	cluster0
TBX3	rs5801092	10.6752	3	cluster1
NPR3	rs13154066	10.6182	1	delta_age
CMSS1	rs12488245	10.5618	2	cluster6
CRIM1	rs4670549	10.5224	1	delta_age
TRIOBP	rs62236745	10.4527	1	cluster16
C7orf25	rs13234512	10.3517	2	cluster1
LRRC37A2	rs3874943	10.3187	1	delta_age
SLC6A6	rs10865722	10.2497	1	delta_age
KCNJ2	rs2285569	10.235	1	cluster6

Table S5: Top 50 aggregated GWAS hits for representation-derived phenotypes from short-axis scans, in descending order of significance.

Gene Name	Lead Variant	Max -log10(p)	Num Significant RDPs	Most Significant RDP
TTN	rs2042995	31.6562	1	delta_age
FKBP7	rs2042996	31.3298	1	delta_age
LIMK1	rs113395463	23.1973	1	delta_age
PI15	rs6472877	19.593	1	delta_age
NUDT7	rs62043885	16.1109	3	cluster8
ADAMTS18	rs17689197	15.9459	3	cluster8
RARRES1	rs1714518	14.5289	1	cluster8
GOSR2	rs17608766	14.0096	1	delta_age
WRNIP1	rs4355649	13.1031	1	cluster3
ADPRHL1	rs76382172	12.9088	1	delta_age
PITX2	rs6843082	12.4039	1	cluster9
TMEM270	rs79344387	12.1143	1	delta_age
NPR3	rs13154066	12.0771	1	delta_age
RPRML	rs117953218	11.9735	1	delta_age
MECOM	rs34585560	10.8959	1	delta_age
FBXO32	rs7006122	10.6356	1	delta_age
LRRC37A	rs2532351	10.621	1	delta_age
WNT3	rs8069437	10.5871	1	delta_age
FAM241A	rs1903403	10.5165	4	cluster8
SMG6	rs903160	10.4007	1	delta_age
SESTD1	rs10930844	10.3843	1	delta_age
RSRC1	rs73164066	10.3262	1	cluster8
TRIB2	rs6727552	10.2328	4	cluster3
HTR7	rs10748555	9.9881	1	delta_age
LSM3	rs11715111	9.87709	1	delta_age
KANSL1	rs7225002	9.8478	1	delta_age
CWC27	rs540834152	9.76052	5	cluster5
MAPT	rs5820605	9.73052	1	delta_age
TIFA	rs326847	9.59493	2	cluster8
CENPK	rs1309558	9.4822	3	cluster5
HDGFL1	rs13211463	9.4656	1	delta_age
CCDC141	rs373251994	9.37475	1	delta_age
CDCA2	rs11985475	9.02453	1	delta_age
TRIOBP	rs62236745	8.85557	1	cluster10
GMDS	rs2761236	8.83456	5	cluster8
GORAB	rs59452262	8.79816	1	cluster1
SRR	rs11371517	8.71519	1	delta_age
MON1B	rs62046461	8.66593	1	cluster8
NMB	rs8033343	8.62444	1	cluster6
SLC6A6	rs10865722	8.60522	1	delta_age
CPS1	rs1047891	8.59287	1	delta_age
CFTR	rs10235008	8.58544	1	cluster10
WNT2	rs73211959	8.49328	1	cluster10
CBR4	rs12644874	8.42726	2	cluster0
WT1	rs10835891	8.41244	1	cluster5
PNPT1	rs77045491	8.29785	1	cluster4
EFEMP1	rs1346786	8.29432	2	cluster2
SOX5	rs137913153	8.28827	1	cluster9
MLF1	rs10154978	8.26661	1	cluster8
ZEB2	rs12996668	8.2329	1	cluster10

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and intro focus on three claims: (1) we train a representation learning model on cardiac MRI videos, (2) we use the representations to predict age and cardiac parameters accurately, and (3) we can define IDPs to capture genetic associations. (1) is addressed by methods description in Section 2, (2) is addressed by experiments in Sections 4.2-4.3 and (3) is addressed by methods in Section 3 and experiments in Section 4.6.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: In the discussion, we acknowledge that future work needs to be done to explore other spatiotemporal modeling methods and benchmarking strategies. Moreover, we acknowledge that further study of GWAS hits is needed. We include ablations in Section 4.3 to test our assumptions around modeling choices.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best

judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification: We do not have theoretical results or proofs. We have formulas describing our representation derived phenotypes (Section 3), but these do not depend on theoretical assumptions.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: In our experiments we describe the dataset, preprocessing, model hyperparameters, packages and settings used for downstream analysis.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.

(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: Data from UK BioBank is regulated and cannot be shared publicly. Our code is not released but we link to the relevant papers for the model architectures we use, and mention packages used for downstream analysis.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: We describe hyperparameters (or indicate default settings for existing methods that have them) and describe our data splits.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: We report p-values for the PheWAS and GWAS results (and provide references to the methods used, which include a description of assumptions and how these p-values are calculated). We also note wherever multiple hypotheses are tested and how significance thresholds are adjusted accordingly.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [No]

   Justification: We do not provide this info but we provide references for the underlying methods that were used for any compute-intensive steps (data preprocessing, VoxelMorph, VideoMAE, PheWAS, GWAS); the references go into more detail on compute requirements.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: As part of an approved UK BioBank research application, we conform to all requirements for analysis of data from human subjects research (data privacy, consent, etc). Beyond that, we do not foresee any risks that require mitigation.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.

- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We introduce a method for identifying targets for treatment of diseases, which we believe could have a positive impact, which is mentioned in the paper. We do not believe there are negative societal impacts specific to our work (other than the general risks of all biological research).

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: We are not releasing any data or models.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: We acknowledge our UK BioBank research application and provide references for any models used.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: We do not release any new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: We analyze data from human subjects (UK BioBank), but did not acquire the data ourselves. More details from UK BioBank (ethics, protocol, etc) can be found here.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [Yes]

    Justification: IRB approval is through our UK BioBank research application (#18448).

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.