

## Deep learning enabled analysis of cardiac sphericity

Milos Vukadinovic<sup>1</sup>, Alan C. Kwan<sup>1</sup>, Victoria Yuan<sup>2</sup>, Michael Salerno<sup>3</sup>, Daniel C. Lee<sup>4</sup>,

Christine M. Albert<sup>1</sup>, Susan Cheng<sup>1</sup>, Debiao Li<sup>5</sup>, David Ouyang<sup>1,6,\*</sup>, Shoa L. Clarke<sup>3,\*</sup>

1. Department of Cardiology, Smidt Heart Institute, Cedars-Sinai Medical Center, Los Angeles, CA
  2. School of Medicine, University of California Los Angeles, Los Angeles, CA
  3. Division of Cardiovascular Medicine, Department of Medicine, Stanford University, Stanford, CA
  4. Department of Medicine and Radiology, Northwestern Medicine, Chicago, IL
  5. Biomedical Imaging Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA
  6. Division of Artificial Intelligence in Medicine, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA
- \*. Co-senior author

Contact: [David.ouyang@cshs.org](mailto:David.ouyang@cshs.org) and [Shoa@stanford.edu](mailto:Shoa@stanford.edu)

## Abstract

Quantification of chamber size and systolic function is a fundamental component of cardiac imaging, as these measurements provide a basis for establishing both diagnosis and appropriate treatment for a spectrum of cardiomyopathies. However, the human heart is a complex structure with significant uncharacterized phenotypic variation beyond traditional metrics of size and function. Characterizing variation in cardiac shape and morphology can add to our ability to understand and classify cardiovascular risk and pathophysiology. We describe deep learning enabled measurement of left ventricle (LV) sphericity using cardiac magnetic resonance imaging data from the UK Biobank and show that among adults with normal LV volumes and systolic function, increased sphericity is associated with increased risk for incident atrial fibrillation (HR 1.31 per SD, 95% CI 1.23-1.38), cardiomyopathy (HR 1.62 per SD, 95% CI 1.29-2.02), and heart failure (HR 1.24, 95% CI 1.11-1.39), independent of traditional risk factors including age, sex, hypertension, and body mass index. Using genome-wide association studies, we identify four loci associated with sphericity at genome-wide significance. These loci harbor known and suspected cardiomyopathy genes. Through genetic correlation and Mendelian randomization, we provide evidence that sphericity may represent a subclinical manifestation of non-ischemic cardiomyopathy.

## Introduction

The heart is a complex three-dimensional structure exposed to continuous dynamic stress. Genetic risk factors, environmental stressors, and age-associated changes impact cardiac morphology and function over the lifespan in a multitude of ways. Dilation of cardiac chambers and/or decline in systolic function are key indicators of disease, and conventional imaging assessments aim to quantify such changes. In turn, there has been significant interest in understanding the genetic basis for variation in such phenotypes<sup>1,2</sup>.

The study of cardiac morphology may provide additional clinical and genetic insights. Specifically, large biobanks with cardiac imaging data now offer an opportunity to define and analyze variation in cardiac morphology that is incompletely quantified by traditional measurements but present in normal populations<sup>3</sup>. However, precise phenotyping using imaging data is challenged by variability in image acquisition and human measurement<sup>4</sup>. Further, manual phenotyping is time-consuming and not feasible for testing novel phenotypes at larger scale. These challenges can be addressed by recent advances in deep learning and computer vision which allow for high throughput automated measurements of cardiac structures<sup>5-8</sup>, and may identify previously uncharacterized clinically-relevant phenotypic variation<sup>3,9,10</sup>.

Cardiomyopathies of different etiologies often result in a similar end-stage phenotype of a more round, spherical ventricle. Among patients with cardiac disease, left ventricle (LV) sphericity index has been previously associated with adverse outcomes and progression of disease, most commonly in echocardiography studies<sup>11-16</sup>. We hypothesized that within the spectrum of normal LV chamber size and systolic function, there exists variation in LV sphericity, and this variation may be a marker of cardiac risk with genetic underpinnings. To test this hypothesis, we applied automated deep-learning segmentation to cardiac magnetic resonance imaging (MRI) data

available in the UK Biobank. We show that among patients with normal LV chamber size and systolic function, the sphericity index predicts incident cardiovascular diseases, including cardiomyopathy, atrial fibrillation (AF), and heart failure. Genetic analysis suggests a shared architecture between sphericity and non-ischemic cardiomyopathy (NICM), with NICM as a possible causal factor for LV sphericity among individuals with normal LV size and function.

## Results

### Measurement of the sphericity index using a convolutional neural network

In order to examine the clinical relevance and genetic influences of LV shape in the setting of normal size and function, we defined our main study cohort ( $n = 38,897$ ) as UK Biobank participants with cardiac MRIs demonstrating normal LV end-diastolic volume, normal LV end-systolic volume, and normal LV ejection fraction (**Supplementary Figure 1**). **Table 1** shows the study cohort baseline characteristics. We applied a fully convolutional neural network for semantic segmentation of cardiac chambers to cardiac MRIs from the UK Biobank. During end-diastole in each segmented a 4-chamber image, we defined the smallest rectangle that fully encompassed the LV blood pool and measured the LV long axis length and short axis length. We defined the sphericity index as the ratio of the short axis length to the long axis length (**Figure 1a**). **Figure 1b** provides representative examples of the observed sphericity index in our study cohort. Sphericity index was similarly normally distributed in both men and women (**Figure 1c**) and showed a slight trend of increasing with age (**Figure 1d**).

### Phenome-wide association study of left ventricular sphericity

We broadly assessed for clinical associations with LV sphericity using a phenome-wide association study (PheWAS) approach (**Figure 2**), and we identified AF as the top association by P value (Pearson  $r$  0.05,  $P$   $9.5 \times 10^{-14}$ ). In comparison, LV long axis length had no association with AF ( $r$  -0.007,  $P$  0.3), while LV short axis length showed a similar association with AF ( $r$  0.05,  $P$   $2.6 \times 10^{-14}$ ). Notably however, AF only ranked 8<sup>th</sup> by P value for short axis length. Other associations, including pulse rate ( $r$  -0.15,  $P$   $2.4 \times 10^{-129}$ ), systolic blood pressure ( $r$  0.09,  $P$   $2.3 \times 10^{-43}$ ), and obesity ( $r$  0.08,  $P$   $3.2 \times 10^{-38}$ ) were substantially stronger for PheWAS of LV short axis

length (**Supplementary Figure 2**). These results show that while metrics of shape may have overlapping clinical associations with metrics of size, analyses of shape may help highlight some associations by attenuating non-specific associations related to anthropometrics.

### **Left ventricular sphericity predicts incident disease**

We next assessed whether the sphericity index predicts incident disease. Given prior studies<sup>12–17</sup> and the results of our PheWAS analysis, we chose four outcomes available through first-occurrence data from the UK Biobank: AF, cardiomyopathy, heart failure, and cardiac arrest. We performed Cox analysis of incident outcomes with adjustments for age at MRI, sex, BMI, pulse rate, and baseline hypertension. We found that the sphericity index predicted incident AF, cardiomyopathy, and heart failure but not cardiac arrest (**Figure 3**). The hazard ratios (HR) associated with a 1-SD increase in sphericity index were 1.31 (95% CI 1.23-1.38,  $P < 2 \times 10^{-16}$ ), 1.62 (95% CI 1.29-2.02,  $P = 2.4 \times 10^{-5}$ ), and 1.24 (95% CI 1.11-1.39,  $P = 1.8 \times 10^{-4}$ ) for AF, cardiomyopathy, and heart failure respectively.

### **Genome-wide association study of left ventricular sphericity**

We then performed a genome-wide association study (GWAS) of common genetic variants and LV sphericity index and compared the results to the GWAS of the LV long and short axis lengths (**Figure 4**). We identified four loci associated with LV sphericity index at genome-wide significance ( $P \leq 5 \times 10^{-8}$ ): *PLN*, *ANGPT1*, *PDZRN3*, and *HLA DR/DQ* (**Table 2**). *PLN* is gene well known for its causal role in dilated and arrhythmogenic cardiomyopathies<sup>18,19</sup>. *ANGPT1* has not been associated with cardiomyopathy to date, but mouse experiments have demonstrated a critical role in mammalian heart development<sup>20,21</sup>. The *PDZRN3* locus has been previously associated with

electrocardiographic characteristics<sup>22</sup> and LV fractal structure in prior GWAS<sup>3</sup>, and regulation of post-natal expression has been implicated in murine cardiac maturation and geometry<sup>23</sup>. The *HLA DR/DQ* locus has long been implicated in dilated cardiomyopathy<sup>24,25</sup>, and recent GWAS that included both normal and abnormal cardiac MRIs have identified this locus for both left and right chamber size metrics<sup>2,26</sup>.

Comparing the GWAS of sphericity to LV short and long axis lengths highlights the value of studying shape-related phenotypes. Given the inherent interplay between sphericity and axis lengths, we expected to see overlap in genetic associations. Indeed, the genetic correlations with sphericity index were 0.64 (P  $2.7 \times 10^{-25}$ ) and -0.63 (P  $6.8 \times 10^{-27}$ ) for short axis and long axis lengths respectively. In terms of genome-wide significant loci, *PLN* was identified by both sphericity and short axis length but not long axis length. Notably, the significance of the association was substantially stronger for sphericity. Conversely, *ANGPT1* and *PDRZRN3* were identified by long axis length but not short axis length. Interestingly, the *TTN* gene was significant for short axis length ( $4.5 \times 10^{-8}$ ) and suggestive for long axis length (P  $1.5 \times 10^{-7}$ ), but the association with sphericity was much weaker (P  $2.1 \times 10^{-5}$ ). This pattern might suggest that *TTN* impacts overall heart size more so than shape.

### **Relationship between sphericity and cardiomyopathy**

We explored the relationship between LV sphericity and non-ischemic cardiomyopathy (NICM) by first measuring the genetic correlation. Given the findings from our Cox analysis, we also assessed AF as a comparison. For each outcome, we used recent large GWAS<sup>27,28</sup>, and correlations were assessed by LD score regression, a technique that is robust to potential sample overlap. We

found a significant correlation between sphericity index and cardiomyopathy ( $rg$  0.42,  $P$  0.01) but no genetic correlation between sphericity and AF ( $rg$  0.04,  $P$  0.5).

We then assessed for evidence of causal relationships using bidirectional two-sample Mendelian randomization by inverse variance weighted meta-analysis. To do so, we used summary statistics for NICM and AF from the FinnGen project<sup>29</sup>, a cohort independent of the UK Biobank. When assessing sphericity index as an exposure, we found no evidence for causality with either NICM or AF. We then assessed sphericity index as an outcome. When using AF as an exposure, we found no evidence for causality with sphericity (**Supplementary Table**). However, when using NICM as an exposure, we found a significant causal relationship for sphericity (beta 0.002,  $P$  0.03). In sensitivity analyses, the relationship remained significant by weighted medians (beta 0.001,  $P$  0.049) but not by MR Egger (beta 0.0007,  $P$  0.6). Together, these findings suggest that cardiomyopathy may be causal for increased sphericity even among individuals with normal measurements of LV systolic function and volumes.



## Discussion

We have shown that the LV sphericity index predicts risk for incident AF, cardiomyopathy, and heart failure in healthy adults. Our genetic analyses suggest that LV sphericity has overlapping genetic architecture with clinical NICM, and the study of phenotypes related to organ shape may provide complimentary insight to the study of organ size. Overall, our data support the hypothesis that spherical LVs in healthy patients may reflect an early subclinical manifestation of cardiomyopathy.

We chose to study LV sphericity as a simple and interpretable metric of LV shape, motivated by clinical experience and prior studies that suggest cardiac sphericity may be a marker of disease. Prior studies were limited by their focus on sphericity after the onset of clinical disease<sup>11-16</sup>, highlighting the relationship between shape and impaired contractility as spherical remodeling occurs secondary to disease or injury. We have added to these prior works by applying deep learning to conduct the first large-scale study of sphericity, while also assessing its clinical relevance and genetic architecture among subjects with normal LV chamber size and systolic function.

The main limitation of this study is the use of a single cohort. The UK Biobank is known to have a healthy cohort bias and does not represent a random sample of the UK population. Replication in additional cohorts will be necessary to validate the use of sphericity for clinical risk prediction. Additionally, in order to limit inflation due to population substructure, our GWAS included only a subset of unrelated participants of similar European ancestry. Lastly, sphericity index may not fully capture phenotypic variation within the left ventricle, as additional features of variable trabeculation, angle, and thickness simultaneously influence cardiac function.

In conclusion, this study demonstrates the utility of using deep learning and advanced imaging analysis to define and study non-traditional cardiac imaging risk biomarkers using the large-scale data that is increasingly available in biobanks. While conventional imaging metrics have significant diagnostic and prognostic value, some of these measurements have been adopted out of convenience or tradition. By representing a specific multi-dimensional remodeling phenotype, sphericity has emerged as a distinct morphologic trait with features not adequately captured by conventional measurements. We expect that the search space of potential imaging measurements is vast, and we have only begun to scratch at the surface of disease associations.

## Methods

### Cohort

The UK Biobank (UKB) is a population-based cohort that links genetic and phenotypic data for approximately 500,000 adult participants from the United Kingdom<sup>30 31</sup>. Our analyses focused on ~48,000 participants who have undergone cardiac MRI<sup>32</sup>. A subset of this group that met quality control filters and that had normal MRI measurements (LV end-diastolic volume 88-218 mL, LV end-systolic volume 31-97 mL, and LV ejection fraction 48-70%)<sup>33</sup> was used as the main cohort (**Supplementary Figure 1**). For the genome-wide association study (GWAS), we selected subjects from the main cohort who passed additional GWAS filters.

### Extraction of Sphericity Index Trait using Machine Learning

We downloaded cardiac MRI containing steady-state free precession image sequences of horizontal long-axis view. Each sequence is given as  $210 \times 208 \times 50$  matrix in DICOM format, with in-plane resolution  $1.8 \times 1.8 \text{ mm}^2$ . In the pre-processing phase, we converted DICOM images and manual annotations (.XML) to NIFTI format, which saves the sequence as a single 3d image for the sake of better file management.

We used a fully convolutional neural network<sup>5,34</sup> for automated LV segmentation of MRI images. Trained weights were obtained from a prior work and inference was run on each video of each patient with cardiac MRIs in the UKB. For each individual, the semantic segmentation was manually evaluated and extreme outliers were excluded, as well as outliers in summary statistics provided from the UKB. We performed a segmentation quality check in two phases. We both heuristically checked (identifying when the segmentation was not continuous or implausibly small) and visually inspected segmentations to excluded patients with suboptimal segmentations.

For all images that passed quality check we created an automated bounding box around the segmentation of left ventricular blood pool (**Figure 1a**). The longer side of the rectangle was taken as LV long axis length (LVL), and the shorter side was taken as LV short axis length (LVS). Sphericity index is calculated as quotient of LVS and LVL ( $SI = \frac{LVS}{LVL}$ ). Higher value of sphericity index indicates more spherical appearance, and lower value indicates an elongated appearance of left ventricular chamber.

### **PheWAS**

Diagnostic data was downloaded as ICD10 codes<sup>5</sup>, processed, and saved as diagnostic groups organized by phecodes using a mapping file provided by PheWAS Catalog<sup>7</sup>. Phenome-wide association studies (PheWAS)<sup>35</sup> were performed by measuring the Pearson *r* correlation between sphericity index and each phecode. In addition, three quantitative measurements were assessed: pulse rate, systolic blood pressure, and diastolic blood pressure. A total of 966 correlations were assessed. Significant associations were identified as those with a P value below the Bonferroni-corrected threshold of  $5 \times 10^{-5}$ .

### **Cox Survival Analysis**

We performed a time-to-event analysis to assess the association of sphericity index with incident atrial fibrillation, cardiomyopathy, heart failure, and cardiac arrest, using the UKB first occurrence data. These outcomes are provided by the UKB as part of a curated set of outcomes derived from health record diagnostic codes. We defined days without incident as the time from the first MRI visit to the first occurrence of the given outcome, death, or last follow-up (March 31st, 2021). For outcomes of atrial fibrillation, heart failure, and cardiomyopathy, participants with the prevalent

disease at the time of the first MRI visit were excluded. We measured the association between the sphericity index and each outcome using a multivariable Cox proportional hazards model adjusted for age at MRI, sex, body mass index (BMI), pulse rate, and hypertension. For this analysis, the sphericity index was normalized to mean equaling 0 and standard deviation (SD) equaling 1, and the hazard ratio (HR) is reported as per 1-SD. We additionally stratified the cohort into the lower 20th, middle 60th, and upper 20th percent of the sphericity index and plotted adjusted cumulative incidence curves by strata.

### **Genome-wide association study**

We used the UKB imputed genotype calls in BGEN v1.2 format. Samples were genotyped using the UK BiLEVE or UK Biobank Axiom arrays. Imputation was performed using the Haplotype Reference Consortium panel and the UK10K+1000 Genomes panel<sup>30</sup>. We used the QC files provided by UKB to create a GWAS cohort consisting of subjects who did not withdraw, were of inferred European ancestry, and were unrelated. Subjects with a genotype call rate  $< 0.98$  were also removed. We considered variants with a minor allele frequency (MAF)  $\geq 0.01$ , and we required genotyped variant to have a call rate  $\geq 0.95$  and imputed variants to have an INFO score  $\geq 0.3$ . Variants with a Hardy-Weinberg equilibrium P value  $< 1 \times 10^{-20}$  were excluded. GWAS was done on a Spark 3.1.1 cluster, using library Hail 0.2 with Python version 3.6. The GWAS was adjusted for age at MRI and sex. We used the conventional P value of  $5 \times 10^{-8}$  as the threshold for defining genome-wide significance.

### **Genetic correlation**

Genetic correlations were estimated using *ldsc version 1.01*<sup>36</sup>. For GWAS performed in this study

(sphericity index, short axis length, long axis length), we used well imputed variants by filtering for  $\text{INFO} \geq 0.9$ . For NICM and AF, we used summary statistics from previously published GWAS<sup>27,28</sup>, and we limited to well imputed variants by intersecting with HapMap3 variants.

### **Mendelian randomization**

Two-sample Mendelian randomization was conducted using the 2SampleMR version 0.5.6 R package. We used a bidirectional approach with LV sphericity index as the exposure for the outcomes of NICM and AF and with LV sphericity as an outcome with NICM and AF as exposures. We used FinnGen<sup>29</sup> release 6 GWAS summary statistics for NICM (finngen\_R6\_I9\_NONISCHCARDMYOP\_STRICT) and AF (finngen\_R6\_I9\_AF) to assure non-overlapping samples. The FinnGen NICM strict phenotype excludes hypertrophic cardiomyopathy. For each exposure, the instrument variable was created by identifying independent significant or suggestive single nucleotide polymorphisms (SNPs) that could be harmonized to the outcome of interest unambiguously. Independent SNPs were defined using a clumping distance of 10 megabases and an  $R^2$  threshold of 0.001. Both sphericity index and NICM had a limited number of genome-wide significant variants after excluding palindromic SNPs and indels, and we therefore included SNPs with a suggestive P value of  $<10^{-6}$ . This resulted in an 11-SNP instrument variable for sphericity index and a 7-SNP instrument variable for NICM. For AF, we only considered genome-wide significant SNPs ( $\leq 5 \times 10^{-8}$ ), resulting in a 34-SNP instrument variable. We used inverse-variance weighted meta-analysis as our primary approach, and we used median weighted meta-analysis and MR Egger as sensitivity analyses.

## References

1. Vasan, R. S. *et al.* Genetic variants associated with cardiac structure and function: a meta-analysis and replication of genome-wide association data. *JAMA* **302**, 168–178 (2009).
2. Pirruccello, J. P. *et al.* Analysis of cardiac magnetic resonance imaging in 36,000 individuals yields genetic insights into dilated cardiomyopathy. *Nat. Commun.* **11**, (2020).
3. Meyer, H. V. *et al.* Genetic and functional insights into the fractal structure of the heart. *Nature* **584**, 589–594 (2020).
4. Augusto, J. B. *et al.* Diagnosis and risk stratification in hypertrophic cardiomyopathy using machine learning wall thickness measurement: a comparison with human test-retest performance. *Lancet Digit Health* **3**, e20–e28 (2021).
5. Bai, W. *et al.* Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J. Cardiovasc. Magn. Reson.* **20**, 65 (2018).
6. Ouyang, D. *et al.* Video-based AI for beat-to-beat assessment of cardiac function. *Nature* **580**, 252–256 (2020).
7. Bai, W. *et al.* A population-based phenome-wide association study of cardiac and aortic structure and function. *Nat. Med.* **26**, 1654–1662 (2020).
8. Duffy, G. *et al.* High-Throughput Precision Phenotyping of Left Ventricular Hypertrophy with Cardiovascular Deep Learning. *arXiv [eess.IV]* (2021).
9. Fries, J. A. *et al.* Weakly supervised classification of aortic valve malformations using unlabeled cardiac MRI sequences. *Nat. Commun.* **10**, 3111 (2019).
10. Yu, M. *et al.* Computational estimates of annular diameter reveal genetic determinants of mitral valve function and disease. *JCI Insight* **7**, (2022).

11. St John Sutton, M. *et al.* Cardiovascular death and left ventricular remodeling two years after myocardial infarction: baseline predictors and impact of long-term use of captopril: information from the Survival and Ventricular Enlargement (SAVE) trial. *Circulation* **96**, 3294–3299 (1997).
12. Wong, S. P. *et al.* Relation of left ventricular sphericity to 10-year survival after acute myocardial infarction. *Am. J. Cardiol.* **94**, 1270–1275 (2004).
13. Di Donato, M. *et al.* Left ventricular geometry in normal and post-anterior myocardial infarction patients: sphericity index and “new” conicity index comparisons. *Eur. J. Cardiothorac. Surg.* **29 Suppl 1**, S225-30 (2006).
14. Levine, Y. C. *et al.* Left ventricular sphericity independently predicts appropriate implantable cardioverter-defibrillator therapy. *Heart Rhythm* **13**, 490–497 (2016).
15. Nakamori, S., Ismail, H., Ngo, L. H., Manning, W. J. & Nezafat, R. Left ventricular geometry predicts ventricular tachyarrhythmia in patients with left ventricular systolic dysfunction: a comprehensive cardiovascular magnetic resonance study. *J. Cardiovasc. Magn. Reson.* **19**, 79 (2017).
16. Khanna, S. *et al.* Left Ventricular Sphericity Index is a reproducible bedside echocardiographic measure of geometric change between acute phase Takotsubo’s syndrome and acute anterior myocardial infarction. *Int J Cardiol Heart Vasc* **29**, 100547 (2020).
17. Ambale-Venkatesh, B. *et al.* Left ventricular shape predicts different types of cardiovascular events in the general population. *Heart* **103**, 499–507 (2017).
18. Schmitt, J. P. *et al.* Dilated cardiomyopathy and heart failure caused by a mutation in phospholamban. *Science* **299**, 1410–1413 (2003).



19. van der Zwaag, P. A. *et al.* Phospholamban R14del mutation in patients diagnosed with dilated cardiomyopathy or arrhythmogenic right ventricular cardiomyopathy: evidence supporting the concept of arrhythmogenic cardiomyopathy. *Eur. J. Heart Fail.* **14**, 1199–1207 (2012).
20. Jeansson, M. *et al.* Angiotensin-1 is essential in mouse vasculature during development and in response to injury. *J. Clin. Invest.* **121**, 2278–2289 (2011).
21. Kim, K. H., Nakaoka, Y., Augustin, H. G. & Koh, G. Y. Myocardial Angiotensin-1 Controls Atrial Chamber Morphogenesis by Spatiotemporal Degradation of Cardiac Jelly. *Cell Rep.* **23**, 2455–2466 (2018).
22. Ntalla, I. *et al.* Multi-ancestry GWAS of the electrocardiographic PR interval identifies 202 loci underlying cardiac conduction. *Nat. Commun.* **11**, 2542 (2020).
23. Pernot, M. *et al.* Decrease of *Pdzrn3* is required for heart maturation and protects against heart failure. *Sci. Rep.* **12**, 8 (2022).
24. Carlquist, J. F., Menlove, R. L., Murray, M. B., O’Connell, J. B. & Anderson, J. L. HLA class II (DR and DQ) antigen associations in idiopathic dilated cardiomyopathy. Validation study and meta-analysis of published HLA association studies. *Circulation* **83**, 515–522 (1991).
25. Pankuweit, S. *et al.* The HLA class II allele DQB1 0309 is associated with dilated cardiomyopathy. *Gene* **531**, 180–183 (2013).
26. Pirruccello, J. P. *et al.* Genetic analysis of right heart structure and function in 40,000 people. *bioRxiv* (2021) doi:10.1101/2021.02.05.429046.
27. Aragam, K. G. *et al.* Phenotypic Refinement of Heart Failure in a National Biobank Facilitates Genetic Discovery. *Circulation* (2018) doi:10.1161/CIRCULATIONAHA.118.035774.

28. Roselli, C. *et al.* Multi-ethnic genome-wide association study for atrial fibrillation. *Nat. Genet.* **50**, 1225–1233 (2018).
29. Kurki, M. I. *et al.* FinnGen: Unique genetic insights from combining isolated population and national health register data. *bioRxiv* (2022) doi:10.1101/2022.03.03.22271360.
30. Littlejohns, T. J. *et al.* The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nat. Commun.* **11**, 2624 (2020).
31. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
32. Petersen, S. E. *et al.* UK Biobank’s cardiovascular magnetic resonance protocol. *J. Cardiovasc. Magn. Reson.* **18**, 1–7 (2016).
33. Petersen, S. E. *et al.* Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (CMR) in Caucasians from the UK Biobank population cohort. *J. Cardiovasc. Magn. Reson.* **19**, 18 (2017).
34. Long, J., Shelhamer, E. & Darrell, T. Fully convolutional networks for semantic segmentation. 3431–3440 (2015).
35. Ye, Z. *et al.* Phenome-wide association studies (PheWASs) for functional variants. *Eur. J. Hum. Genet.* **23**, 523–529 (2015).
36. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).

## Tables

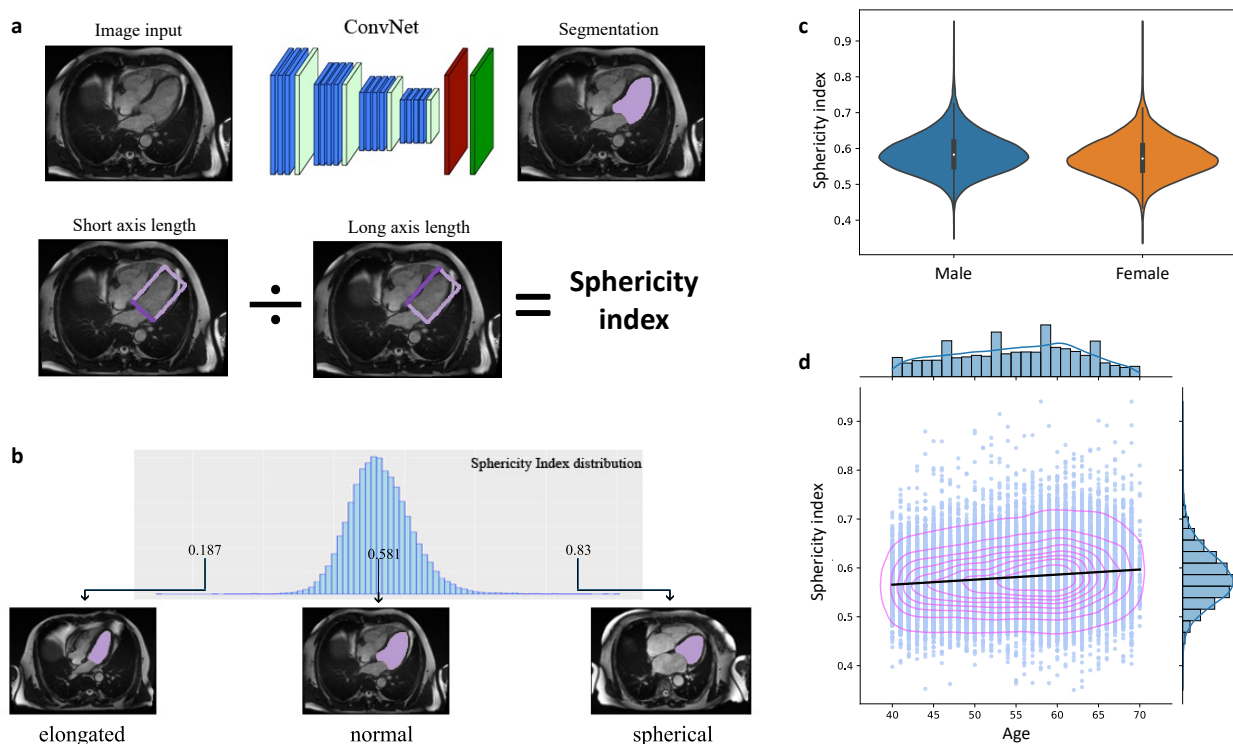
**Table 1. Cohort baseline characteristics**

<b>Characteristic</b>	<b>Mean or n</b>
N	38897
Age at MRI	54.9 ± 7.6
Male	17983 (46.2%)
Self-identified White British	32965 (84.7%)
Body mass index (kg/m <sup>2</sup> )	26.5 ± 4.2
Hypertension	2388 (6.1%)
Pulse rate	67.7 ± 10.7
LV ejection fraction (%)	57.0 (4.47)
LV end diastolic volume (mL)	135 ± 26.4
LV end systolic volume (mL)	58.4 ± 13.4

**Table 2. Lead variants and candidate genes for genome-wide significant loci of each left ventricle (LV) trait.**

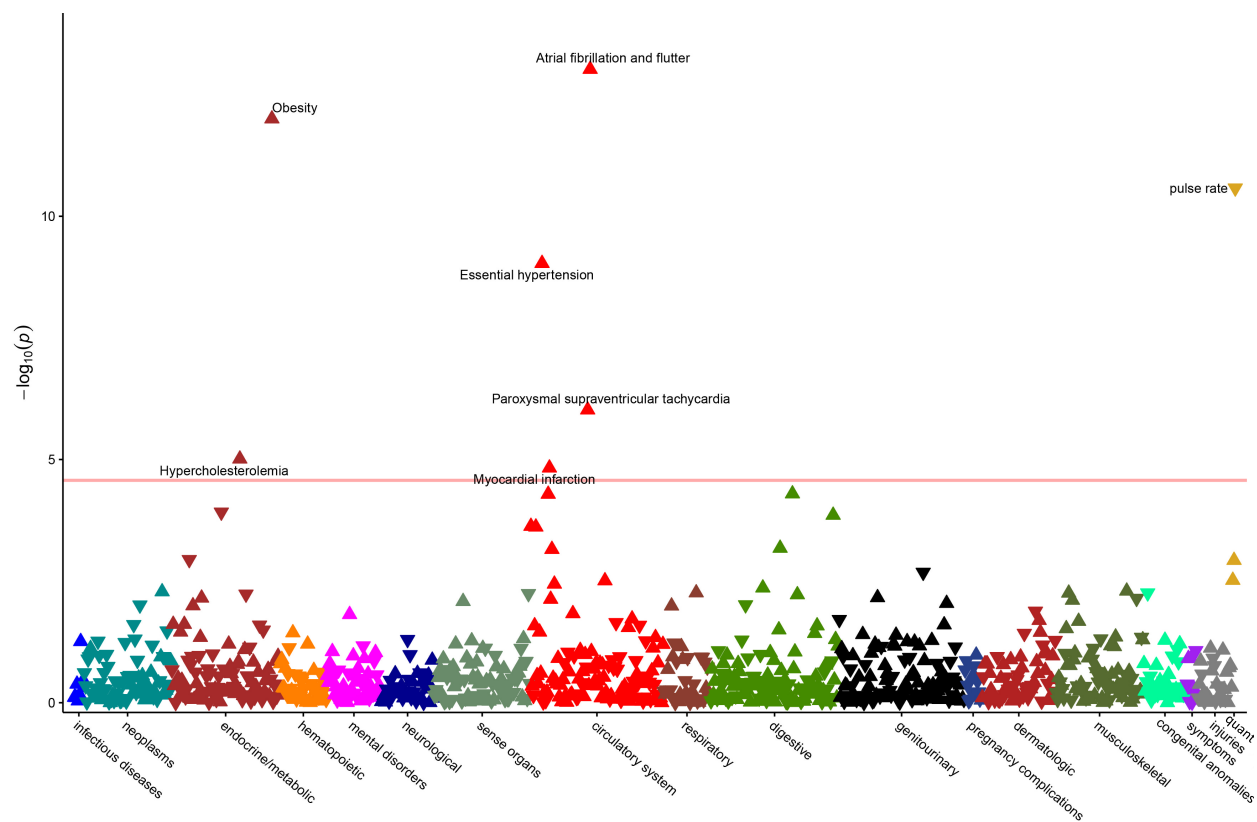
LV trait	RSID	Chr	hg19 position	Ref	Alt	Beta	P	Candidate gene
Sphericity index	rs11756438	6	118993632	C	A	0.004	1.16x10 <sup>-15</sup>	<i>PLN</i>
Sphericity index	rs1461990	8	108087628	C	G	0.003	4.90x10 <sup>-10</sup>	<i>ANGPT1</i>
Sphericity index	none	3	73578036	AACACAC	A	0.003	2.28x10 <sup>-8</sup>	<i>PDZRN3</i>
Sphericity index	rs199682224	6	32486364	G	T	0.006	3.58x10 <sup>-8</sup>	<i>HLA-DRB5</i>
Short axis length	rs57912492	6	118702621	C	T	0.24	2.43x10 <sup>-10</sup>	<i>PLN</i>
Short axis length	rs462797	16	22832777	G	A	0.46	2.93x10 <sup>-8</sup>	<i>HS3ST2</i>
Short axis length	rs111692972	2	179669931	C	T	-0.44	4.48x10 <sup>-8</sup>	<i>TTN</i>
Long axis length	none	3	73578036	AACACAC	A	-0.37	4.31x10 <sup>-9</sup>	<i>PDZRN3</i>
Long axis length	rs72007904	8	108085680	A	AACTATTC	-0.36	7.33x10 <sup>-9</sup>	<i>ANGPT1</i>

## Figures



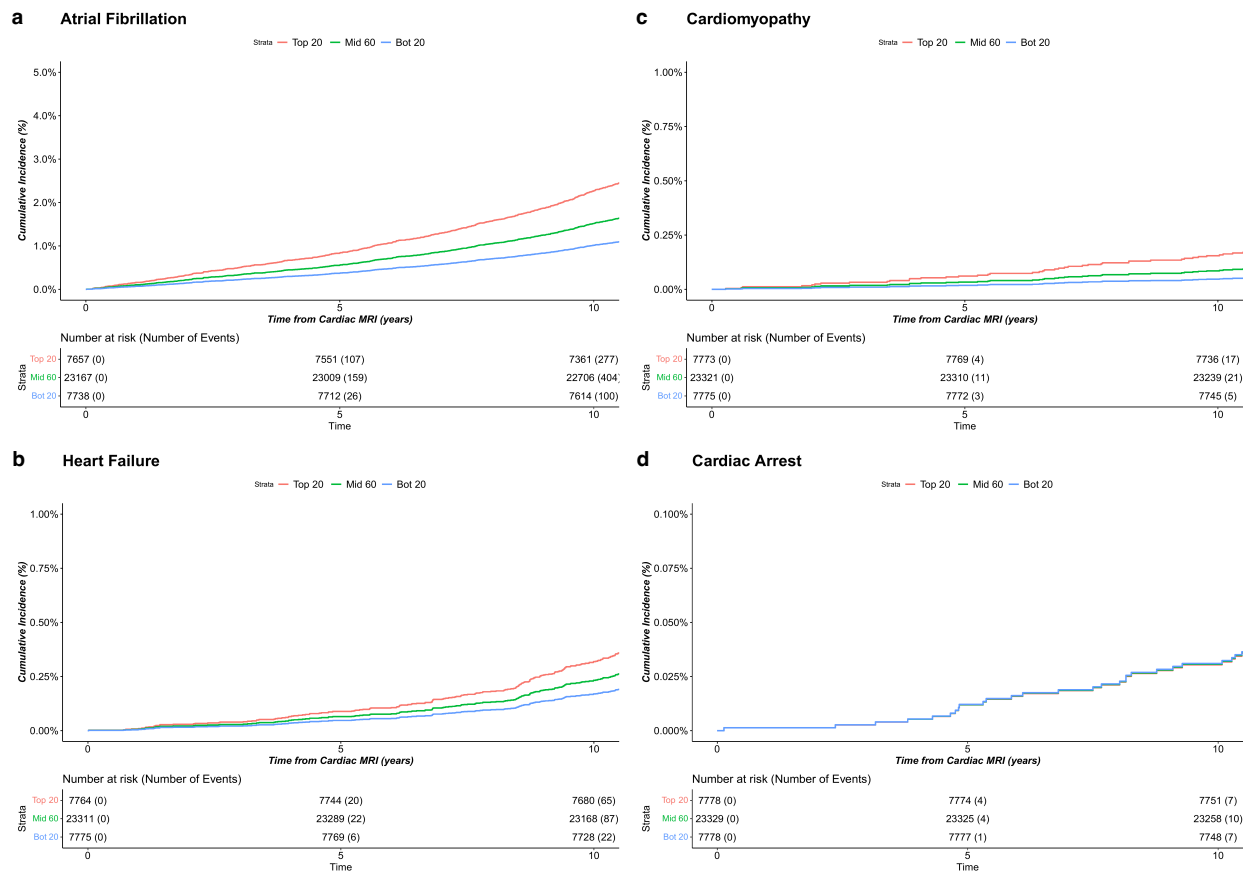
**Figure 1. Deep learning enabled measurement of left ventricular sphericity index.**

**a**, Automated image segmentation of cardiac chambers was performed by applying a fully convolutional neural net to 4-chamber view cardiac MRI images. Sphericity index was defined by the ratio of the short and long axis lengths of a bounding box of the left ventricle. **b**, Distribution of the sphericity index in the study population and representative examples. **c**, The sphericity index is similarly normally distributed in male and female subjects. **d**, The sphericity index showed a small positive correlation with age. Concentric contour lines reflect increasing density of data point.



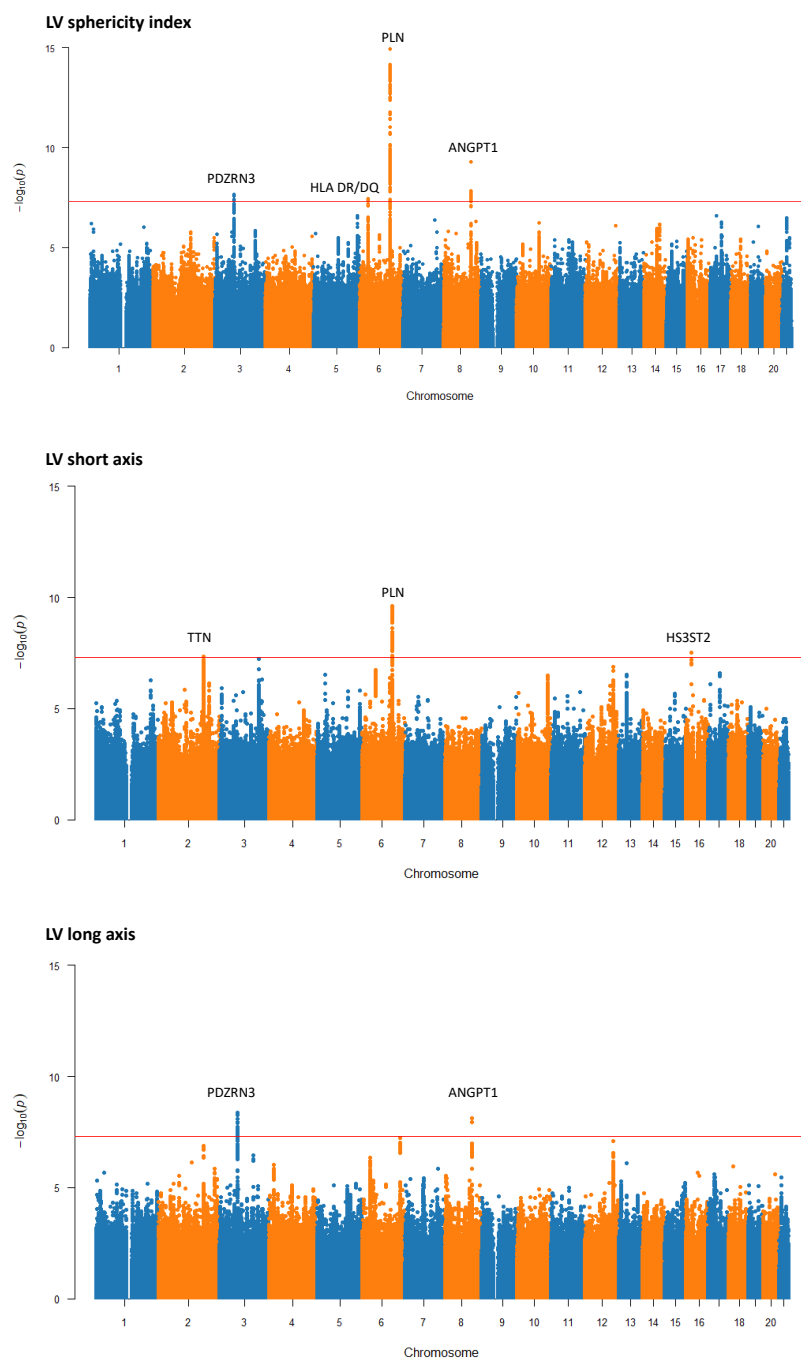
**Figure 2. Phenome-wide association study of left ventricular sphericity index.**

Upward pointing triangles represent positive correlations with sphericity index, and downward pointing triangles represent negative correlations. The horizontal red line reflects the threshold of Bonferroni significance for 966 tests ( $5 \times 10^{-5}$ ). Associations meeting Bonferroni significance are labeled.



**Figure 3. Risk of incident disease stratified by quintiles of sphericity index.**

Time-to-event analysis was performed from time of MRI to first occurrence of (a) atrial fibrillation, (b) cardiomyopathy, (c) heart failure, and (d) cardiac arrest. Models were adjusted for age at MRI, sex, body mass index, hypertension, and pulse rate. Plots show cumulative incidence of disease stratified by sphericity index quintiles, comparing the top 20%, middle 60%, and bottom 20%.



**Figure 4. Manhattan plots for genome-wide associations studies of left ventricular sphericity index, short axis length, and long axis length.**

The red horizontal line represents a genome-wide significant P value of  $5 \times 10^{-8}$ . Regions that reach genome-wide significant are labeled with candidate loci.