

## Impact of Measurement Imprecision on Genetic Association Studies of Cardiac

### Function

Milos Vukadinovic<sup>1,2,4</sup>, Gauri Renjith<sup>1</sup>, Victoria Yuan<sup>1,2</sup>, Alan Kwan<sup>1,4</sup>, Susan C. Cheng<sup>1</sup>, Debiao Li<sup>4</sup>, Shoa L. Clarke<sup>5,\*</sup>, David Ouyang<sup>1,6,\*</sup>

1. Department of Cardiology, Smidt Heart Institute, Cedars-Sinai Medical Center, Los Angeles, CA
2. Department of Bioengineering, University of California Los Angeles, Los Angeles, CA
3. Department of Medicine, University of California Los Angeles, Los Angeles, CA
4. Biomedical Imaging Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA
5. Department of Medicine, Division of Cardiovascular Medicine, Stanford University, Stanford, CA
6. Division of Artificial Intelligence in Medicine, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA

\*. Co-senior author

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

### Keywords

Precision phenotyping, Genome-wide association study, UK Biobank, Left ventricular ejection fraction, Cardiac magnetic resonance imaging

## **Abstract**

### **Background**

Recent studies have leveraged quantitative traits from imaging to amplify the power of genome-wide association studies (GWAS) to gain further insights into the biology of diseases and traits. However, measurement imprecision is intrinsic to phenotyping and can impact downstream genetic analyses.

### **Methods**

Left ventricular ejection fraction (LVEF), an important but imprecise quantitative imaging measurement, was examined to assess the impact of precision of phenotype measurement on genetic studies. Multiple approaches to obtain LVEF, as well as simulated measurement noise, were evaluated with their impact on downstream genetic analyses.

### **Results**

Even within the same population, small changes in the measurement of LVEF drastically impacted downstream genetic analyses. Introducing measurement noise as little as 7.9% can eliminate all significant genetic associations in an GWAS with almost forty thousand individuals. An increase of 1% in mean absolute error (MAE) in LVEF had an equivalent impact on GWAS power as a decrease of 10% in the cohort sample size, suggesting optimizing phenotyping precision is a cost-effective way to improve power of genetic studies.

### **Conclusions**

Improving the precision of phenotyping is important for maximizing the yield of genome-wide association studies.

## **Clinical perspective**

### **What Is New?**

- Measurement imprecision in cardiac imaging phenotypes can substantially impact downstream genetic association studies, explaining much of the difference in identified genetic variants between echocardiography and cardiac magnetic resonance imaging.
- Using the example of left ventricular ejection fraction as an important but imprecise clinical measurement, the analysis suggests that the measurement variation within the range of clinician interpretation reduced genome-wide association studies' power to detect genetic risk factors as much as decreasing the study population size by 20%.

### **What Are the Clinical Implications?**

- More precise measurements can result in a better understanding of the genetics of cardiac phenotypes and accelerate the development of precision medicine.
- Rather than simply increasing population size, improving measurement precision allows for cost-effective discovery of genetic associations.

## Introduction

Cardiovascular disease is the leading cause of death in the world, and significant work has been undertaken to understand the mechanisms of disease and develop preventive measures. By studying the human genome, insights have been obtained to understand pathways and mechanisms of function and disease risk, and in recent studies, researchers have moved beyond binary labels of disease diagnosis to quantitative phenotypes to obtain greater power in assessing the relationship between genotype and phenotype<sup>1-4</sup>. From quantitative laboratory biomarkers elucidating the relationship between hypercholesterolemia and coronary artery disease<sup>5</sup> to imaging characteristics in population cohorts<sup>4</sup> revealing the genetic determinants of cardiovascular development<sup>6,7</sup>, quantitative assessments of health provide additional signal compared to conventional binary labels of disease.

Despite its relative frequency, critical public health importance, and often penetrant inheritance, heart failure has relatively few known genetic risk factors. Early classic genetic studies were not able to identify many genetic associations with measurements determined by echocardiography<sup>8</sup>. Different cohorts resulted in heterogeneous findings in the genetic determinants of cardiomyopathy. More recent studies with measurements from cardiac magnetic resonance imaging (MRI) have been able to find additional loci of relevance and reaffirm previously suspected variants<sup>8,9</sup>, suggesting both larger sample sizes as well as improvements in phenotyping precision, can improve our understanding of the human disease. However, as populations increase, the scaling of phenotyping becomes more costly, and automated approaches are used for most quantitative measurements, despite only moderate correlation compared to manual measurements ( $R^2 = 0.318$  for cardiac MRI LVEF)<sup>9</sup>.

While quantitative traits have more power than often binary labels of disease, quantitative traits can have noisy measurements. For example, left ventricular ejection fraction (LVEF) as measured by echocardiography can have measurement variation up to 7 - 10%<sup>10,11</sup>, impacting downstream analyses. While cardiac MRI is frequently thought of as a reference standard, given its improved image quality and high signal-to-noise ratio, automated approaches similarly have measurement variation of between 3 – 5%<sup>9</sup>. We use LVEF, an important metric of cardiac function, as an example of an important but noisy measurement to explore the impact of measurement variability on downstream genetic association studies. We compare various methods to obtain the same phenotypic measurement as well as introduce simulated noise in the phenotype measurement to evaluate the relative impact of measurement imprecision and sample size on downstream genetic studies.

## Methods

### Cohort

The UK Biobank is a population-based cohort that links genetic and phenotypic data for approximately 500,000 adult participants from the United Kingdom<sup>12,13</sup>. We focused on 39,624 participants who had InlineVF measured LVEF<sup>14</sup>, cardiac MRI, and genetic data available. Before running Genome-Wide Association Studies this cohort was passed through additional quality check filters (**Supplementary Figure 1**).

### Multiple Approaches to Measure LVEF

Multiple methods of calculating LVEF from the same underlying imaging data were used to assess the impact of phenotyping precision on downstream analyses. First, the UKB provides automated LVEF measurements derived from MRI using Inline VF software<sup>9</sup>, however, this is presented without manual quality control. To compare alternative automated approaches, we also derived LVEF from MRIs using the deep learning segmentation approach suggested by Bai et al<sup>10</sup>. From the short-axis view videos, segmentation was performed and we calculated the LV volume for each frame using Simpson's method.

To simulate reader variability, additional experiments were performed introducing Gaussian noise with a mean of 0 and a standard deviation (sd) ranging from [1,10]. We generated multiple phenotypic measurements from the same underlying imaging data, gradually incrementing Gaussian noise, and performed GWAS on each to investigate how measurement error/imprecision affects genetic associations.

Additionally, we further compared results with two final approaches to assess LVEF. When visually assessing LVEF, clinicians often round the value to the nearest 5%, thus we generated a

set of phenotype labels by rounding LVEF values to the nearest multiple of 5. For the final comparison, we generated binary LVEF labels by categorizing values as normal or abnormal, with normal values ranging from 52-72 for males and 54-74 for females.

### **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>12</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 variants 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. After variant filtering, we were left with 9774199 filtered variants. GWAS was done on a Spark 3.1.1 cluster, using the 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.

### **Assessing Association Power's Relationship with Cohort Size**

Apart from noise in phenotype measurements, we also evaluate the effect of cohort decrease on GWAS results. We generated 6 different phenotype files where, starting from the original LVEF cohort (39,624), we keep 90% (35,661), 80% (31,699), 70% (27,736), 60% (23,774), 50% (19812), and 40% (15850) of the samples. Cohort decrease was performed before GWAS QC,

and for each step the selection of samples to be excluded was random. Inspecting the effect of cohort decrease helps us define the relationship between the number of LVEF samples and GWAS power.

### SNP-based accuracy

We use an accuracy metric to determine the amount of overlap in significant SNPs between the baseline GWAS results and noise-modified GWAS results. First, we remove all non-significant SNPs by excluding SNPs with a p-value less than  $5 \times 10^{-8}$ , which is the Bonferroni corrected p-value threshold. Then, we consider significant SNPs found in both the base results and noise-modified results as true positives (TP), the SNPs found only in the noise-modified results as false positives (FP), and the SNPs not found in the noise-modified results but found in the base results as false negatives (FN). We then calculate  $SNP_{accuracy} = \frac{TP}{TP+FP+FN}$ .

### GWAS Sensitivity

Sensitivity determines the amount of overlap in significant loci between the baseline GWAS results and noise-modified GWAS results. Specifically, given that  $peaks_{base}$  is the number of significant loci in base GWAS, and  $peaks_{correct}$  is the number of significant loci that persisted in noise GWAS then  $Sensitivity = \frac{peaks_{correct}}{peaks_{base}}$ . The number of loci and their position can be determined by manual inspection, but we also developed an automatic method. Our automatic method uses a hierarchical clustering algorithm to determine the number and the position of loci from both GWAS, which we then use to compute  $peaks_{base}$  and  $peaks_{correct}$ .



## Results

### Quantitative phenotypes improve power of association studies

The study cohort for all analyses consisted of 39,624 adult unrelated subjects of European ancestry (**Table 1**). As a baseline, we first conducted a GWAS of the LVEF phenotype released with the UKBB cardiac MRI data. We identified 5 loci at genome-wide significance on chromosomes 1, 6, 8, 10, and 19 near genes ZBTB17, CDKN1A, CTSB, BAG3, and AP1M1 (**Figure 1**). In comparison, for an LVEF phenotype binarized to simply abnormal or normal, multiple previously detected loci lost genome-wide significance (including loci for CTSB and AP1M1). Similarly, recognizing the inherent variation present in measuring LVEF, we additionally compared the results if the LVEF was bucketed to 5% bins and showed such imprecision decreased statistical power in all SNPs in the association study compared to the continuous LVEF baseline phenotype.

### Phenotype noise degrades power of association studies

To investigate the effect of measurement imprecision on GWAS power, we performed a series of association studies while introducing noise in the range of known clinician variation (**Figure 2**). Simulated variation to the LVEF measurement naturally increases in mean absolute error. Noise with a gaussian standard deviation of 5 results in a mean absolute error of 3.97% and  $R^2$  of 0.65 (**Supplementary Table 1**), and results in the loss of genome-wide significance for the AP1M1 loci on chromosome 19. As we increase phenotypic noise in the range of clinical variation, power gradually declines and the noise equivalent to 7.92% MAE results in a complete loss of genomic-wide significance (**Table 2**). Given echocardiography is known to have a clinician-to-

clinician variation of the same or greater MAE<sup>10</sup>, such measurement imprecision could contribute to the limited hits in historical echocardiography-derived GWAS<sup>8</sup>.

### **Comparison of Impact of Phenotype Noise vs Cohort Size**

Given the summary statistics from 16 different GWAS, we modeled the relationship between noise and GWAS power (**Figure 3, Figure 4**). There is a linear relationship between the increase in MAE and the decrease in GWAS power. We calculated that an increase of 1% in MAE causes the loci sensitivity to decrease by 13% ( $p=5.5e-6$ ) and the SNP accuracy by 14% ( $p=6.6e-5$ ). Experiments with other methods of introducing noise in assessing LVEF similarly show a decrease in genetic association with more imprecise measurements (**Supplementary Figure 2 and Supplementary Table 2**). A similar effect occurs with reductions in cohort size, as a 1% decrease in cohort size results in a 1.3% decrease in loci sensitivity ( $p=0.01$ ) and a 1.9% decrease in SNP-based accuracy ( $p=0.0007$ ). We found that a 1% MAE increase has the same effect on loci sensitivity as a 10.3% cohort decrease and the same effect as a 7.2% cohort decrease on SNP accuracy.

### **Improving phenotyping augments downstream genetic analyses**

Cardiac MRI provides clinicians and researchers with a plethora of high-resolution imaging, with even the abbreviated 20-min UK Biobank cardiac MRI protocol resulting in 9 sequences with over 30,000 images per study<sup>11</sup>. With so many images and patients, the released UKBB measurements were generated using a fully automated workflow (with Siemens inLineVF) without quality inspection and bias correction. When compared with manual clinician evaluation, the automated measurements of LVEF result in a mean absolute error of 3.4%, R2 of 0.348, and ICC of 0.521 for LVEF. Using a previously published deep learning segmentation model<sup>16</sup>, we

independently derived LV segmentation-based calculated LVEF and found a mean absolute error of 6.1%, R2 of 0.335, and ICC of 0.431 for LVEF compared to the automated measurements from UKBB (**Figure 5**). However, with these deep learning segmentation derived LVEF measurements, the same cohort identified more loci of interest with significant experimental data backing its relevance. In particular, loci on chromosomes 2, 5, and 8 near genes TTN, DNAJC18, and ZNF572 were not previously identified using the released UKBB LVEF measurements but able to be picked up with our quality-controlled measurements.

## Discussion

In this study, we assessed the impact of measurement imprecision on genetic associations with LVEF and found substantially impaired power in downstream GWAS analysis with even slight increases in imprecision. Even modest phenotyping variation significantly impacted downstream genetic associations, often to a greater extent than changes in cohort size. As measurement variation is present in many clinical measurements and large cohorts increasingly rely on automated measurements, efforts to improve the precision of measurements can potentially be a cost-effective way to maximize the yield of genetic association studies.

Cardiac function as measured by LVEF is an important clinical measurement that defines disease and identifies patients who are eligible for life-prolonging therapeutics. In echocardiography, human test-retest evaluation of LVEF can range between 7-10%, with slight changes in annotation as well as timing that can significantly impact calculations<sup>10,16</sup>. Few variability studies have been undertaken in cardiac MRI, although improved degrees of manual measurement variability have been found<sup>9,17</sup>. Our analysis suggests that a substantial gain in signal comes from the improvement of measurements precision, such that its impact is often greater in relative size than sample size. Efforts to improve phenotyping precision that can significantly affect the power and accuracy of downstream analyses.

Noise in measurements can appear in both semi-automated and fully automated workflows<sup>17</sup>, and by improving the precision of measuring LVEF, we also improve the accuracy and robustness of downstream GWAS results. The relatively large improvement in yield of genetic association with more precise phenotyping was substantial in comparison to the marginal benefit of increasing the cohort size. As more genetic analyses are undertaken with automated measurements or assessments<sup>4,7,18,19</sup>, an additional evaluation must be taken to assess the

variability and quality of the phenotyping. Such insights ideally will be confirmed with parallel measurements of similar phenotypes or multiple approaches to obtain the same phenotype

In summary, genetic association studies on imaging phenotypes allow researchers to discover many associations that help understand the underlying biology of the disease and structure<sup>20</sup>. For LVEF, even advanced imaging has variability in measurements that can substantially impact downstream association studies. The impact of such variability is even more profound than significant changes in cohort size, suggesting improvement in imaging precision and precise phenotyping in general has significant additional value in improving the power of genetic association studies.

**Acknowledgments** not applicable

**Sources of Funding** This work was supported in part by the National Institutes of Health NIH K99 HL157421.

**Disclosures**

None.

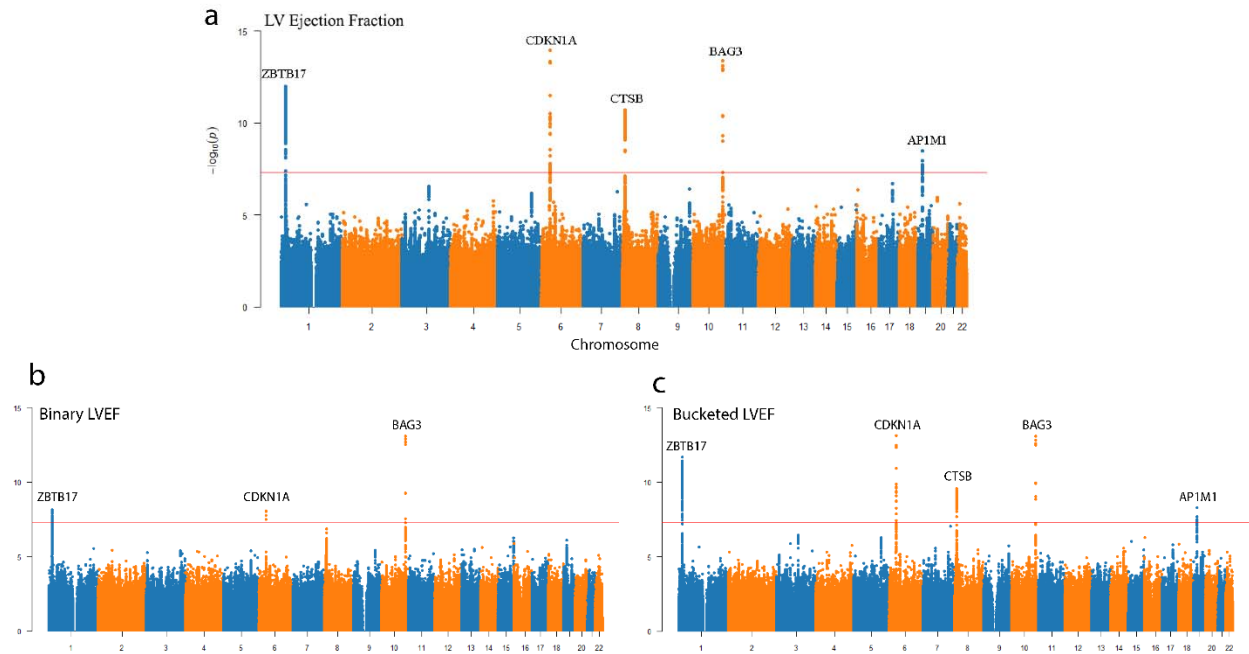
## References

1. Pirruccello JP, Achille PD, Nauffal V, Nekoui M, Friedman SN, Klarqvist MDR, Chaffin MD, Khurshid S, Roselli C, Batra P, Ng K, Lubitz SA, Ho JE, Lindsay ME, Philippakis AA, et al. Genetic analysis of right heart structure and function in 40,000 people. *bioRxiv*. 2021.
2. Pirruccello JP, Bick A, Wang M, Chaffin M, Friedman S, Yao J, Guo X, Venkatesh BA, Taylor KD, Post WS, Rich S, Lima JAC, Rotter JI, Philippakis A, Lubitz SA, et al. Analysis of cardiac magnetic resonance imaging in 36,000 individuals yields genetic insights into dilated cardiomyopathy. *Nature communications*. 2020;11(1):2254.
3. Agrawal S, Wang M, Klarqvist MDR, Smith K, Shin J, Dashti H, Diamant N, Choi SH, Jurgens SJ, Ellinor PT, Philippakis A, Claussnitzer M, Ng K, Udler MS, Batra P, et al. Inherited basis of visceral, abdominal subcutaneous and gluteofemoral fat depots. *Nature communications*. 2022;13(1):3771.
4. Haas ME, Pirruccello JP, Friedman SN, Wang M, Emdin CA, Ajmera VH, Simon TG, Homburger JR, Guo X, Budoff M, Corey KE, Zhou AY, Philippakis A, Ellinor PT, Loomba R, et al. Machine learning enables new insights into genetic contributions to liver fat accumulation. *Cell genomics*. 2021;1(3).
5. Liu DJ, Peloso GM, Yu H, Butterworth AS, Wang X, Mahajan A, Saleheen D, Emdin C, Alam D, Alves AC, Amouyel P, Di Angelantonio E, Arveiler D, Assimes TL, Auer PL, et al. Exome-wide association study of plasma lipids in >300,000 individuals. *Nature genetics*. 2017;49(12):1758–1766.
6. Bai W, Sinclair M, Tarroni G, Oktay O, Rajchl M, Vaillant G, Lee AM, Aung N, Lukaschuk E, Sanghvi MM, Zemrak F, Fung K, Paiva JM, Carapella V, Kim YJ, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *Journal of cardiovascular magnetic resonance: official journal of the Society for Cardiovascular Magnetic Resonance*. 2018;20(1):65.
7. Meyer HV, Dawes TJW, Serrani M, Bai W, Tokarczuk P, Cai J, de Marvao A, Henry A, Lumbers RT, Gierten J, Thumberger T, Wittbrodt J, Ware JS, Rueckert D, Matthews PM, et al. Genetic and functional insights into the fractal structure of the heart. *Nature*. 2020;584(7822):589–594.
8. Vasan RS, Glazer NL, Felix JF, Lieb W, Wild PS, Felix SB, Watzinger N, Larson MG, Smith NL, Dehghan A, Grosshennig A, Schillert A, Teumer A, Schmidt R, Kathiresan S, et al. Genetic variants associated with cardiac structure and function: a meta-analysis and replication of genome-wide association data. *JAMA: the journal of the American Medical Association*. 2009;302(2):168–178.
9. Suinesiaputra A, Sanghvi MM, Aung N, Paiva JM, Zemrak F, Fung K, Lukaschuk E, Lee AM, Carapella V, Kim YJ, Francis J, Piechnik SK, Neubauer S, Greiser A, Jolly M-P, et al. Fully-automated left ventricular mass and volume MRI analysis in the UK Biobank population cohort: evaluation of initial results. *The international journal of cardiovascular imaging*. 2018;34(2):281–291.
10. Farsalinos KE, Daraban AM, Ünlü S, Thomas JD, Badano LP, Voigt J-U. Head-to-Head Comparison of Global Longitudinal Strain Measurements among Nine Different Vendors: The EACVI/ASE Inter-Vendor Comparison Study. *Journal of the American Society of Echocardiography: official publication of the American Society of Echocardiography*.

2015;28(10):1171–1181, e2.

11. O’Dell WG. Accuracy of Left Ventricular Cavity Volume and Ejection Fraction for Conventional Estimation Methods and 3D Surface Fitting. *Journal of the American Heart Association*. 2019;8(6):e009124.
12. Littlejohns TJ, Holliday J, Gibson LM, Garratt S, Oesingmann N, Alfaro-Almagro F, Bell JD, Boulwood C, Collins R, Conroy MC, Crabtree N, Doherty N, Frangi AF, Harvey NC, Leeson P, et al. The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nature communications*. 2020;11(1):2624.
13. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O’Connell J, Cortes A, Welsh S, Young A, Effingham M, McVean G, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203–209.
14. Petersen SE, Matthews PM, Francis JM, Robson MD, Zemrak F, Boubertakh R, Young AA, Hudson S, Weale P, Garratt S, Collins R, Piechnik S, Neubauer S. UK Biobank’s cardiovascular magnetic resonance protocol. *Journal of cardiovascular magnetic resonance: official journal of the Society for Cardiovascular Magnetic Resonance*. 2016;18:8.
15. Petersen SE, Matthews PM, Francis JM, Robson MD, Zemrak F, Boubertakh R, Young AA, Hudson S, Weale P, Garratt S, Collins R, Piechnik S, Neubauer S. UK Biobank’s cardiovascular magnetic resonance protocol. *Journal of cardiovascular magnetic resonance: official journal of the Society for Cardiovascular Magnetic Resonance*. 2016;18(1):1–7.
16. Yuan N, Jain I, Rattehalli N, He B, Pollick C, Liang D, Heidenreich P, Zou J, Cheng S, Ouyang D. Systematic Quantification of Sources of Variation in Ejection Fraction Calculation Using Deep Learning. *JACC. Cardiovascular imaging*. 2021;14(11):2260–2262.
17. Augusto JB, Davies RH, Bhuva AN, Knott KD, Seraphim A, Alfarih M, Lau C, Hughes RK, Lopes LR, Shiwani H, Treibel TA, Gerber BL, Hamilton-Craig C, Ntusi NAB, Pontone G, et al. Diagnosis and risk stratification in hypertrophic cardiomyopathy using machine learning wall thickness measurement: a comparison with human test-retest performance. *The Lancet. Digital health*. 2021;3(1):e20–e28.
18. Zekavat SM, Raghu VK, Trinder M, Ye Y, Koyama S, Honigberg MC, Yu Z, Pampana A, Urbut S, Haidermota S, O’Regan DP, Zhao H, Ellinor PT, Segrè AV, Elze T, et al. Deep Learning of the Retina Enables Phenome- and Genome-Wide Analyses of the Microvasculature. *Circulation*. 2022;145(2):134–150.
19. Kosaraju A, Goyal A, Grigorova Y, Makaryus AN. Left Ventricular Ejection Fraction. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing; 2022.
20. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, Martin HC, Lappalainen T, Posthuma D. Genome-wide association studies. *Nature Reviews Methods Primers*. 2021;1(1):1–21.

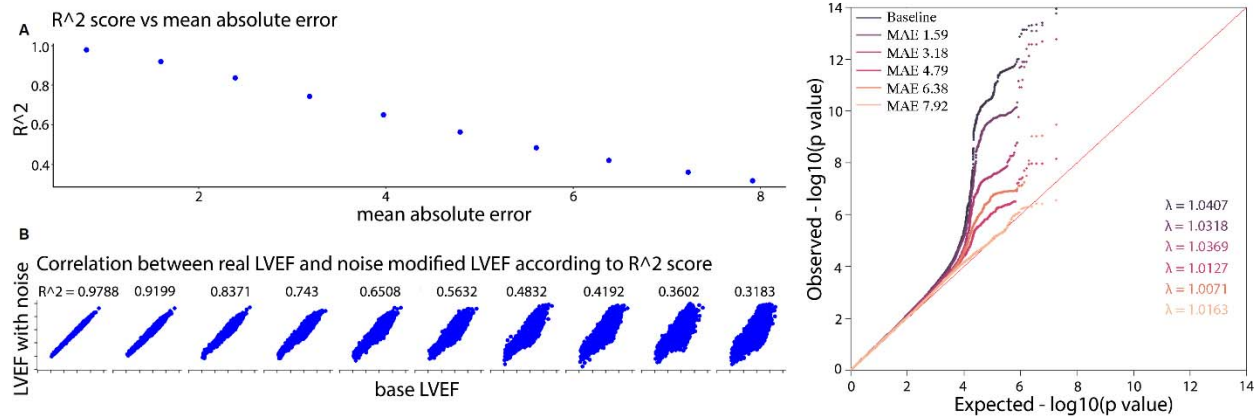
## Figures



**Figure 1** Manhattan plots for genome-wide association studies on UK Biobank reported left ventricular ejection fraction

**a**, GWAS on continuous LVEF measurements **b**, GWAS on Normal/Abnormal LVEF where the range for normal is 52-72 in male and 54-74 female population **c**, GWAS on LVEF bucketed to the nearest multiple of 5

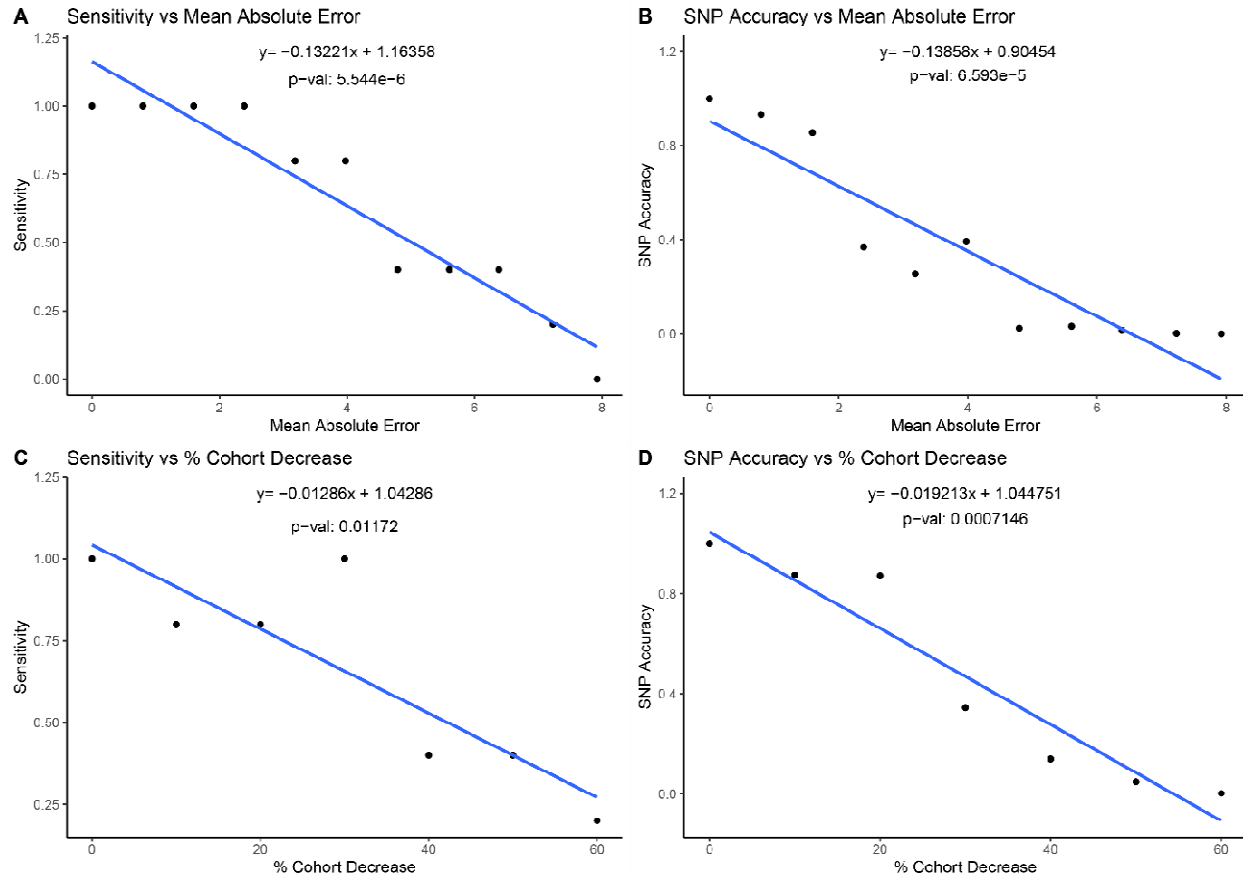




**Figure 2** Impact of noise in LVEF on GWAS

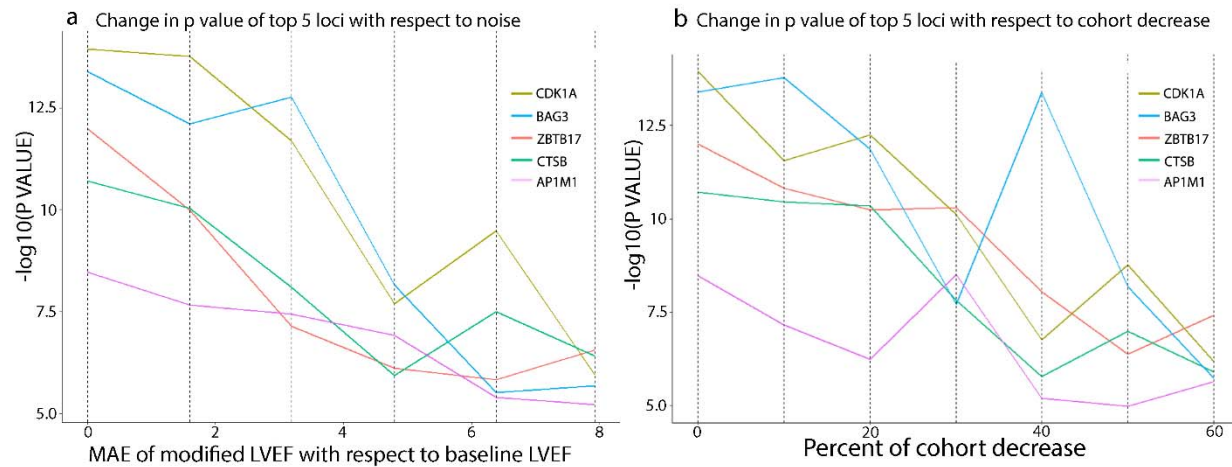
**a**, Visualizing r<sup>2</sup> score, mean absolute error, and the distribution of noise-modified-LVEF with respect to the baseline LVEF

**b**, Q-Q plots of P values from GWAS summary statistics for different levels of noise

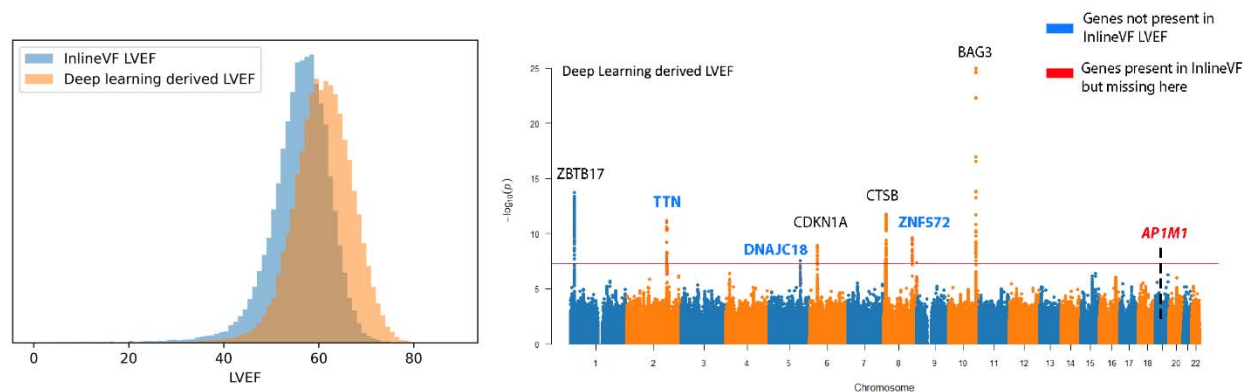


**Figure 3** Impact of cohort decrease and noise generation on GWAS power.

**a**, Regression analysis on the impact of measurement error quantified by a mean absolute error on sensitivity. **b**, Regression analysis on the impact of the mean absolute error on SNP accuracy. **c**, Regression analysis of the impact of cohort size decline on sensitivity. **d**, Regression analysis of the impact of cohort size decline on SNP accuracy.



**Figure 4 a**, Slope chart shows the change in the P value of the top 5 loci with respect to mean absolute error; **b**, Slope chart shows the change in a P value of top 5 loci with respect to the cohort decrease; each locus is named after the closest gene



**Figure 5** Differences in distribution and GWAS summary statistics between two methods of obtaining LVEF from MRI

**a**, Histograms of InlineVF derived LVEF and Deep Learning derived LVEF **b**, Manhattan plot from GWAS performed on Deep Learning derived LVEF; genes colored in blue don't appear in InlineVF LVEF GWAS (Figure 1a); genes colored in red appear in InlineVF LVEF GWAS but not in deep learning derived LVEF GWAS

## Tables

**Table 1. Cohort baseline characteristics**

<b>Characteristic</b>	<b>Mean or n</b>
N	39624
Age at MRI	54.9 ± 7.47
Male	18933 (47.8%)
Self-identified White British	33726 (85.1%)
Body mass index (kg/m <sup>2</sup> )	26.5 ± 4.19
Hypertension	2487 (6.3%)
Pulse rate	67.9 ± 10.9
LV ejection fraction (%)	55.4 (6.78)
LV end diastolic volume (mL)	141
LV end systolic volume (mL)	64.1

**Table 2. Metrics of genetic signal for each increase in SD**

<b>Noise SD</b>	<b>SNP Accuracy</b>	<b>Loci Sensitivity</b>
<b>0%</b>	1.0	1.0
<b>1%</b>	0.9377	1.0
<b>2%</b>	0.8547	1.0
<b>3%</b>	0.3675	1.0
<b>4%</b>	0.2537	0.8
<b>5%</b>	0.3921	0.8
<b>6%</b>	0.0228	0.4
<b>7%</b>	0.0307	0.4
<b>8%</b>	0.0145	0.4
<b>9%</b>	0.0020	0.2
<b>10%</b>	0	0

## Supplemental Material

### Supplementary Tables

#### Supplementary Table 1. Mapping between Gaussian Noise SD and MAE

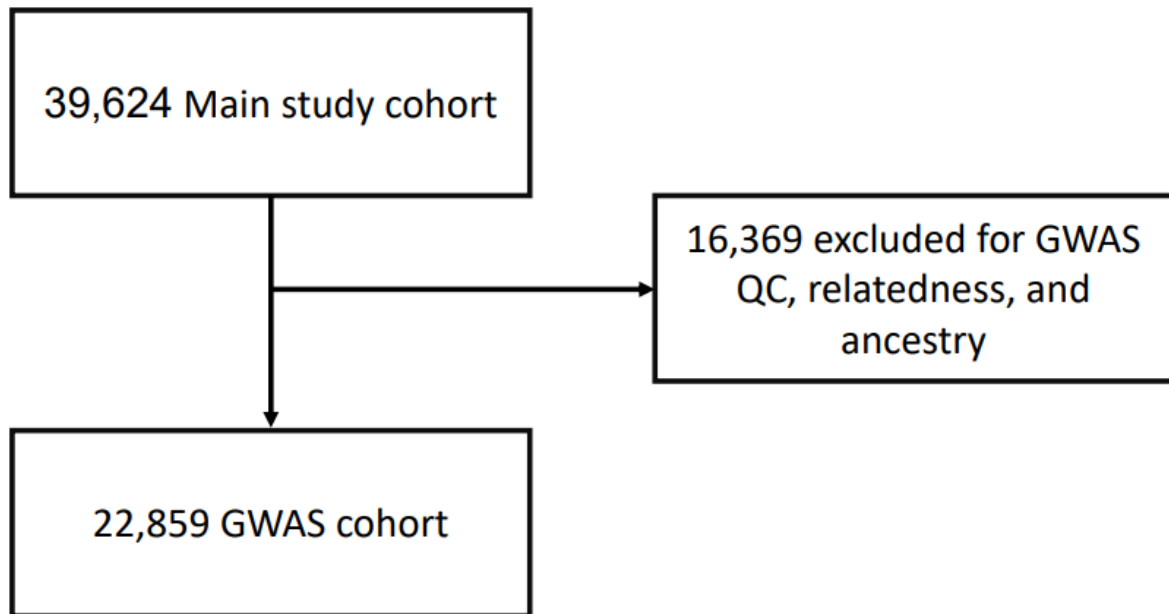
SD	MAE	R2
0	0	1
1	0.797489	0.9788
2	1.594416	0.9199
3	2.386753	0.8371
4	3.183924	0.743
5	3.974958	0.6508
6	4.793956	0.5632
7	5.604129	0.4832
8	6.380848	0.4192
9	7.228321	0.3602
10	7.920860	0.3183

### Supplementary Table 2. Metrics of genetic signal for each decrease in cohort size

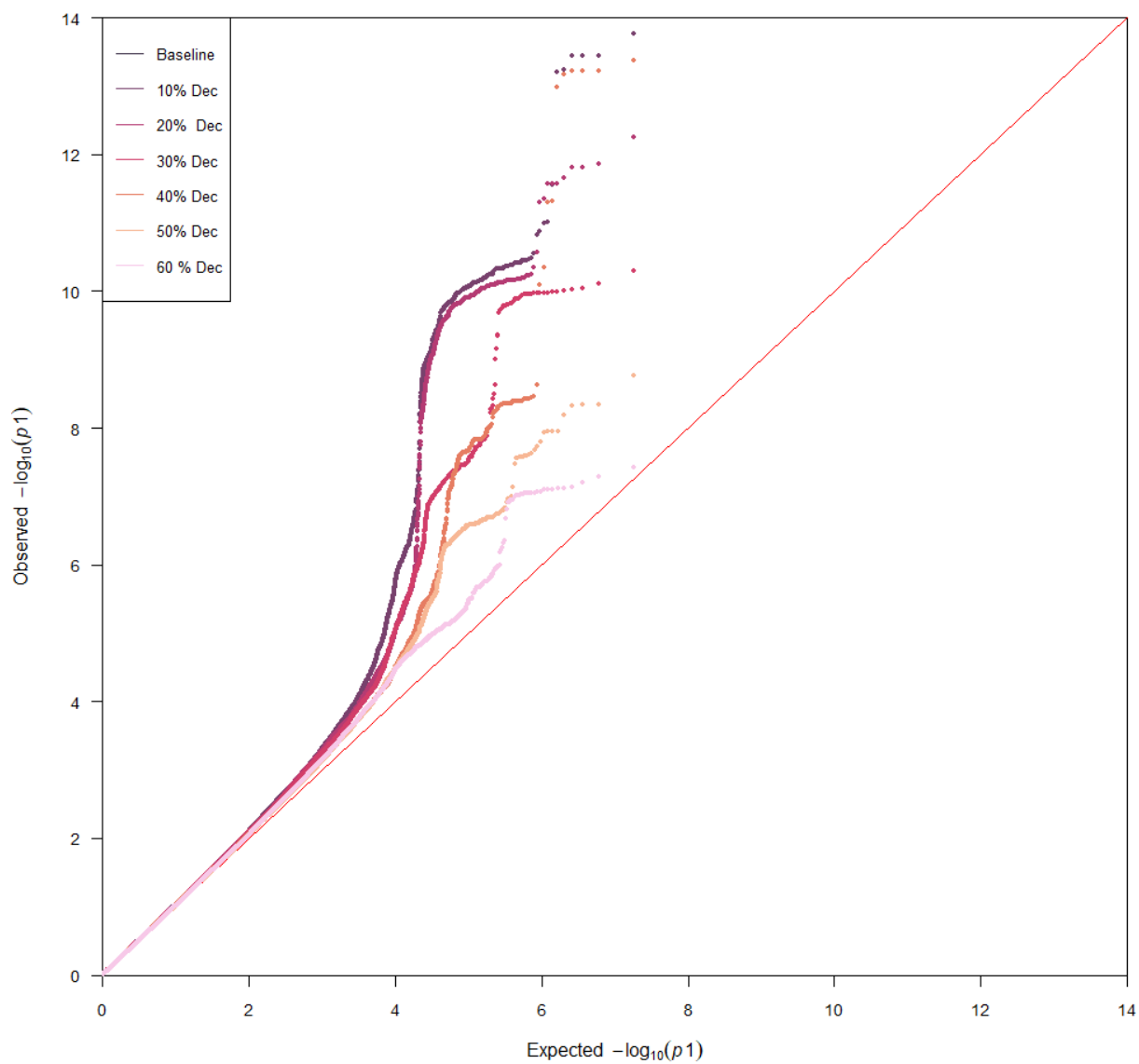
<b>Cohort decrease</b>	<b>SNP Accuracy</b>	<b>GWAS Sensitivity</b>
<b>0%</b>	1.0	1.0
<b>10%</b>	0.8744	0.8
<b>20%</b>	0.8713	0.8
<b>30%</b>	0.3436	1.0
<b>40%</b>	0.1392	0.4
<b>50%</b>	0.0477	0.4
<b>60%</b>	0.0019	0.2



## Supplementary Figures



Supplementary Figure 1. Cohort diagram



**Supplementary Figure 2.** *Q-Q plots of P values from GWAS summary statistics for different percentages of cohort decrease*