nature portfolio

Peer Review File

A self-supervised electrocardiogram foundation model for empowering cardiovascular disease prediction and genetic factor discovery

Corresponding Author: Professor Huiying Zhao

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This manuscript presents ECG-LFM, a self-supervised foundation model for electrocardiogram (ECG) analysis, pre-trained on over 10 million 12-lead ECGs. The model integrates contrastive learning and masked language modeling to capture both global and fine-grained ECG patterns, achieving state-of-the-art performance in predicting 8 cardiovascular diseases. ECG-LFM-derived features (EDFs) demonstrate interpretability by aligning with known ECG biomarkers and enable novel genetic discoveries through GWAS and Mendelian randomization. This is an interesting study indicating potential applications of the foundation model in diagnosis of cardiovascular disease. I have a few comments for the authors to address.

1.The ECG-LFM is a novel large-scale foundation model. It is necessary to compare the model to traditional machine-learning approaches in predicting risks of heart diseases through ECG, and additionally clarify how multi-task framework of ECG-LFM provides unique advantages.

2.The study evaluates the prediction of the model on six key measurements of heart functions (LV/LA/RV ejection fraction and stroke volume), demonstrating relevant clinical applicability. Applications of the model in predicting other clinically used measurements could strengthen the model's utility in cardiovascular function assessment.

3.The manuscript introduces a well-designed suite of three self-supervised tasks, contextual contrastive loss, MLM loss, and multi-segment contrastive loss for capturing diverse ECG characteristics. What are their complementary roles in ECG representation learning? Specifically, how does each objective contribute uniquely to capturing clinically relevant patterns?

4.The perturbation-based EDF interpretation (PEI) method is innovative. What are the relationships of ECG traits detected by this method with the established electrophysiological principles?

5.While the manuscript identifies individual genes (e.g., NFIA, TTN) associated with ECG-LFM-derived features (EDFs), a gene set enrichment analysis has been missed. This would help uncover broader biological pathways and mechanisms underlying the genetic associations.

6.In Methods, limited information on hyperparameter optimization for fine-tuned models are presented.

7.The manuscript reports SNPs significantly associated with EDFs. How many of these SNPs are known as associated with ECG traits? A pleiotropy analysis is needed to determine whether these SNPs influence multiple ECG traits beyond the EDFs.

8.ECG data preprocessing process needs to be visualized. Can you provide a figure to show the complete conversion process from the original ECG signal to the preprocessed ECG signal?

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
This is a paper by Siying Lin et al, titled “A self-supervised electrocardiogram foundation model for empowering cardiovascular disease prediction and genetic factor discovery.” The authors build a foundation model on 10+ million 12-lead ECGs, and fine-tune the model to CVD risk prediction. They extract EDF embeddings and perform GWAS and found a few SNPs, most of which were previously known. The paper is well written and flows well. There is enough description of the methods. The superior performance of ECG-LFM in predicting CVD-related diseases and phenotypes is clear. The GWAS results are good but not outstanding. I have the following comments (not sorted by minor or major comments):
1. Line 151: elaborate and explain more about individual recognition. What are the input data in the Chapman and CPSc-2018 datasets that were passed to the ECG-LFM? And what is the meaning of individual recognition? What is ACC measuring exactly?
2. Line 166: The authors used the embeddings of ECG-LFM to predict cardiac arrhythmia in the Chapman and CPSC-2018 datasets. How was the cardiac arrhythmia defined? Based on ICD codes? What ICD codes. How specific was it: AF, cardiomyopathy, etc? And how/what cardiac arrhythmia information was used in the fine-tuning process? Authors mentioned that the model was fine-tuned for CVD risk, but this needs more explanation to show what data went into the fine-tuning step and what CVD risk or information was accounted for. Also, what type of CVD modalities were included: Arrhythmias, CHD, long QT, AF, etc.? You mention 8 types of CVs on page 180. Although it might be detailed in the methods/supp methods, this should be explicitly mentioned in the results for better reading.
3. Line 278: What is the rational behind the selection of 16 important EDFs and 215 related phenotypes and what they are? This needs to be specified in supp material.
4. Line 297: It is important to add the effect size of associations.
5. Line 310: I wonder why GWAS did not discover snps/loci that were previously reported for ECG intervals? Phenome-wide association analysis showed very significant correlations between EDF and ECG intervals.
6. In Figure 5A, were the 8 CVD types included? I want to see AF, MI, etc. on that plot and see how they associate with the 16 EDF features. They should be annotated.
7. Authors should check how heritable the EDFs are. EDFs don’t seem to be well powered for the GWAS analysis. Not many loci were observed despite the relatively good sample size.
8. Italicize gene names throughout the manuscript.
9. In the MR analysis, Line 327, authors mention EDF or CVDs as exposure. I understood that IVs EDF CVDs is the MR framework, which makes CVDs as outcomes. Is that correct? Or authors tried IDs CVD EDFs? This should be clearly mentioned in both methods and results.
10. Line 364: I would not call TTN a novel finding. TTN is well know for ECG-related phenotypes, AF, cardiomyopathy, arrhythmias, etc.
11. What was the rational behind performing PCA on top of the 1024 EDFs? The embeddings are representations of the ECGs, then PCA is a linear combination of these embeddings. GWAS was performed on 16 EDFs, then now on 10 PCs. What do we expect to find and what is the biological difference between the use of both of these traits. How do authors interpret these results since the tested phenotypes are not explainable anymore?
12. Line 366 – 371: consistency lacks in reporting results. Sometimes Z-scores are reported and sometimes are not.
13. Line 372: in this section, GWAS was performed on PCs. But here on this line, we go back to EDFs and test gene-level associations. This should be moved to the previous section. Moreover, MAGMA was performed on the meta-GWAS results: was it for the multi-trait analysis performed by METAL on all PCs? It is confusing since authors mention EDFs in this section and not PCs, given that no multi-trait analysis on EDF was performed.
14. Typo gene name on Line 380. PLM should PLN?.
15. Line 385: tone down that sentence. Based on the results, it is not convincing that “genes identified as associated with EDFs potentially play roles in regulating ECG and CVDs.” Most of what was shown is only “association” from GWAS studies.
16. In the discussion: the authors claim the generalizability of ECG-LFM to diverse populations and clinical scenarios. They need to mention how this was checked. They need to show statistics about the datasets they used and that they contain individuals from diverse ancestral groups and are exposed to various environmental exposure, lifestyle, etc.
17. In the discussion, line 427 paragraph, I disagree with the claim that a groundbreaking aspect of the study was used to uncover genetic association underlying ECG morphology. GWAS on the 16 EDFs did not reveal many loci despite the relatively large sample size, and the multi-trait analysis on PCs revealed some loci that are already known for CVD (AF, ECG intervals, etc.).
18. Contributions: authors should add who performed the various types of analysis, building the foundation model, data cleaning, GWAS, etc.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
This article aims to extend ECG analysis toward the use of large-scale foundation models for cardiac signal interpretation. Traditional ECG algorithms—whether rule-based or task-specific deep learning models—are limited by their reliance on manually labeled data and their narrow focus on predefined diagnostic tasks. In contrast, an Electrocardiogram Large-Scale

Foundation Model (ECG-FM) leverages massive unlabeled datasets and self-supervised learning to capture the full spectrum of temporal and morphological patterns embedded in ECG signals.

In addition, the authors attempt to relate the outcomes of their proposed model to genotypes. Overall, this is a very interesting paper, which could benefit from a deeper GWAS and post-GWAS analysis.

1/ Some caution is needed regarding claims of novelty. First, HEARTBEiT was trained on more than 8 million observations, which is of the same order of magnitude (although somewhat smaller) than the present study. Second, as I understand it, the authors aim to associate latent factors generated by the model with genetic variants. While this is a commendable objective, it is not entirely novel from a methodological standpoint—other studies have already linked ECG-derived features to genotypes (e.g., work from the van der Harst group in the UK Biobank). This should be acknowledged and perhaps rephrased. That said, this does not diminish the relevance or interest of the analysis itself.

2/ Although ECG-LFM benefits from large-scale self-supervised pretraining, the article clearly shows that its downstream performance still depends on supervised fine-tuning using labeled ECG datasets. High-quality annotations therefore remain necessary to adapt the foundation model to specific clinical prediction tasks.

3/ GWAS analysis. This genetic component is important because it can help reveal the new information extracted by the proposed algorithms. However, it seems somewhat limited by sample size, and it is a bit disappointing to identify only 19 independent association signals, given what is known about the genetic architecture of standard ECG traits and CVDs. Since this section is central to the paper, I recommend a more comprehensive comparison between the genetic architecture of the EDFs and those of ECG or CVD traits as reported in the literature. A sensible first step would be to report genetic correlations rather than focusing exclusively on genome-wide significant SNPs (for example, via LD-score regression). The current PheWAS and GWAS analyses alone do not provide a sufficiently clear picture of genetic correlations (line 798).

4/ On a minor note, I did not see any QQ plots or reports of inflation for the GWAS. These could also be reported using LD-score regression (in this case to estimate SNP heritability and potential inflation).

5/ Using only MAGMA to relate the EDFs to genes (rather than to genotypes) seems somewhat limited. If the authors intend to pursue this direction, they may consider methods that integrate multiple sources of evidence to link SNP associations to genes—MAGMA being one component but not the only relevant tool.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have adequately addressed the previous comments. I have no further comments on this work.

(Remarks on code availability)
The authors have made the code available and provided adequate documentation, including instructions for installation and usage. Based on the provided materials, the code is a usable resource for the community and should allow reproduction of the main results.

Reviewer #2

(Remarks to the Author)
I have no further comments.

(Remarks on code availability)

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

This manuscript presents ECG-LFM, a self-supervised foundation model for electrocardiogram (ECG) analysis, pre-trained on over 10 million 12-lead ECGs. The model integrates contrastive learning and masked language modeling to capture both global and fine-grained ECG patterns, achieving state-of-the-art performance in predicting 8 cardiovascular diseases. ECG-LFM-derived features (EDFs) demonstrate interpretability by aligning with known ECG biomarkers and enable novel genetic discoveries through GWAS and Mendelian randomization. This is an interesting study indicating potential applications of the foundation model in diagnosis of cardiovascular disease. I have a few comments for the authors to address.

RE: We are grateful to the Reviewer for taking the time to review our manuscript, and for the positive feedback on our paper.

1.The ECG-LFM is a novel large-scale foundation model. It is necessary to compare the model to traditional machine-learning approaches in predicting risks of heart diseases through ECG, and additionally clarify how multi-task framework of ECG-LFM provides unique advantages.
RE: We thank the reviewer for this constructive suggestion. We have comprehensively compared ECG-LFM with traditional machine learning approaches, ECG-ResNet and DCNN-LSTM, across multiple datasets and tasks. The result indicated ECG-LFM has achieved significantly (two-sided t-test *P-value* < 0.05) better performance on AUC than other methods. Additionally, ECG-LFM has shown advantages on representation learning, generalization, and interpretability. The advantages of ECG-LFM are stated as follows.

(1) Better Performance in predicting disease risks
ECG-LFM consistently outperformed traditional methods across all evaluation scenarios. On the PTB-XL dataset, ECG-LFM achieved an average AUROC of 0.951, significantly surpassing ECG-ResNet (average AUROC = 0.893) and DCNN-LSTM (average AUROC = 0.890) by 6.5% and 6.8%, respectively. This advantage was maintained across all five CVD categories, with particularly strong performance in predicting conduction disturbance (AUROC=0.986) and myocardial infarction (AUROC=0.958). The superiority of ECG-LFM was further validated on the Chapman dataset, where it achieved an average AUROC of 0.993, significantly better than traditional methods ECG-ResNet (average AUROC = 0.976) and DCNN-LSTM (average AUROC = 0.972) (two-sided t-test *P-value* $<8.88\times10^{-6}$). Importantly, this performance advantage generalized to external test sets (CODE-15, CPSC-2018, and UKB), demonstrating the strong generalization capability of the model.

(2) **Other Advantages of ECG-LFM**
ECG-LFM is pre-trained on multiple self-supervised tasks and large-scale datasets, which allows it to learn a rich feature space and can be effectively fine-tuned for diverse downstream tasks, such as CVD prediction and cardiac function assessment. In comparison, the traditional methods rely solely on convolutional operations or other approach for feature extraction, and are trained end-to-end for a single prediction task, which has limited ability to conduct multi-tasks. The multi-task learning framework of ECG-LFM provides several key advantages.

- Comprehensive Representation Learning. ECG-LFM combined contextual contrastive loss, masked language modelling, and multi-segment contrastive loss, which enable the model to capture both local morphological features and global temporal dependencies, simultaneously. This is particularly important for ECG analysis, where both fine-grained waveform details and long-range temporal patterns are clinically relevant.
- Enhanced Generalization. By learning from multiple self-supervised tasks during pre-training, ECG-LFM develops more robust and generalizable representations. This explains its superior performance on external test sets, where traditional models often suffer from performance degradation due to domain shift.
- Patient-Level Consistency. The multi-segment contrastive loss enforces consistency between different ECG segments from the same patient, helping the model learn patient-specific characteristics that are stable across time. This reduces the impact of inter-individual variability and improves model robustness.
- Interpretability through PEI and GWAS Analysis. ECG-LFM enables superior interpretability through our perturbation-based EDF interpretation (PEI) method and Genome-Wide Association Study (GWAS) analysis. PEI successfully mapped EDFs to clinically meaningful ECG regions, such as EDF#102 to ST-segment (associated with MI and ST/T changes), EDF#19 to P-wave (associated with AFIB), and EDF#945 to QRS morphology (associated with hypertrophy). Furthermore, applications of the EDFs in GWAS identified 24 significant single nucleotide polymorphisms (SNPs) (P-value$<5\times10^{-8}$, LD $r^2<0.01$) associated with ECG, including 8 novel findings. This dual interpretability approach, morphological mapping through PEI and genetic validation through GWAS, provides unprecedented insights into how deep learning features relate to both clinical phenotypes and underlying genetic architecture.

We have added the above content to the Discussion section of the manuscript (page 20):

"ECG-LFM leverages a multi-task learning strategy that combines contextual contrastive loss, masked language modeling (MLM) loss, and multi-segment contrastive loss, enabling the model to align local ECG features with their contextual representations, reconstruct masked segments to learn long-range relationships, and enforce patient-level consistency across ECG segments, respectively."

"A key contribution of this study is on using EDFs to uncover genetic associations underlying ECG morphology. Traditional GWAS rely on handcrafted ECG features (e.g., QT interval), which are limited in their ability to capture complex waveform interactions. In contrast to traditional models that necessitate high-quality labelled datasets for downstream analysis, ECG-LFM leverages self-supervised pre-training to extract comprehensive, label-independent representations, referred to as ECG-LFM-derived features (EDFs). While previous studies have linked deep learning-derived features to genetic variants (PMID: 40822352), our approach uniquely combines deep representations from an ECG foundation model with comprehensive genetic and phenotypic analyses to provide novel insights into the genetic architecture of cardiac electrophysiology."

2.The study evaluates the prediction of the model on six key measurements of heart functions

(LV/LA/RV ejection fraction and stroke volume), demonstrating relevant clinical applicability. Applications of the model in predicting other clinically used measurements could strengthen the model's utility in cardiovascular function assessment.

RE: Thank you for this valuable suggestion to expand the clinical applicability of our model. We have now evaluated the predictive performance of ECG-LFM on two additional clinically relevant cardiac measurements, LV Circumferential Strain Global and LV Mean Myocardial Wall Thickness Global. LV circumferential strain is able to provide important information on myocardial contractility and is increasingly used in clinical practice for early detection of LV dysfunction (PMID: 22245798). LV wall thickness assessment is crucial for diagnosing hypertrophic cardiomyopathy (PMID: 34147459). Accurately predicting these additional evaluations significantly strengthen the clinical utility of our model.

In predicting LV Circumferential Strain Global, ECG-LFM achieved a Pearson correlation coefficient (PCC) of 0.64 on the test set (n = 8,247), demonstrating strong performance in assessing myocardial deformation. In predicting LV Mean Myocardial Wall Thickness Global, ECG-LFM achieved a PCC of 0.72 on the test set (n = 8,247), indicating accurate prediction of myocardial hypertrophy.

We have added the above content to the Results section (page 12):

"Finally, ECG-LFM was further fine-tuned to predict eight heart functional phenotypes, including left ventricular (LV) ejection fraction, LV stroke volume, left atrial (LA) ejection fraction, LA stroke volume, right ventricular (RV) ejection fraction, RV stroke volume, LV circumferential strain global and LV mean myocardial wall thickness global by using 32,988 individuals as training set (80%) from UKB. It was independently tested by using 8,247 individuals (20%) from the UKB dataset. ECG-LFM achieved an average Pearson correlation coefficient (PCC) of 0.649 that is significantly (P-value $<7.74\times10^{-4}$) better than HeartBEiT (average PCC = 0.607), ECG-ResNet (average PCC = 0.594), and DCNN-LSTM (PCC = 0.592) (Figure 3C, Supplementary Figure S4)."

3.The manuscript introduces a well-designed suite of three self-supervised tasks, contextual contrastive loss, MLM loss, and multi-segment contrastive loss for capturing diverse ECG characteristics. What are their complementary roles in ECG representation learning? Specifically, how does each objective contribute uniquely to capturing clinically relevant patterns?

RE: Thank you for this insightful question about the complementary roles of our three self-supervised tasks. Each objective contributes uniquely to capturing clinically relevant ECG patterns by addressing different aspects of representation learning.

(1). **Contextual Contrastive Loss: Local Feature Alignment**

This loss ensures that the model learns to align local ECG features with their contextual representations. By maximizing the similarity between the target context vectors (from quantization) and the predicted contextual representations (from the embedding module), contextual contrastive loss forces the model to capture fine-grained morphological patterns that are crucial for distinguishing subtle ECG abnormalities. This is particularly important for detecting localized changes such as ST-segment elevation in myocardial infarction or T-wave inversions in ischemia.

(2). **Masked Language Modeling Loss: Token-Level Reconstruction**

The MLM task requires the model to predict masked ECG tokens based on surrounding context, which promotes learning of long-range dependencies and temporal relationships in the ECG signal. This is essential for capturing clinically relevant patterns that span multiple heartbeats, such as arrhythmia sequences, conduction abnormalities, and repolarization dynamics. The reconstruction objective ensures that the model learns a comprehensive representation of the entire ECG waveform.

(3). **Multi-Segment Contrastive Loss: Patient-Level Consistency**

This loss enforces consistency between different segments from the same patient while distinguishing segments from different patients. By maximizing similarity between consecutive ECG segments from the same individual, Multi-segment contrastive loss helps the model learn patient-specific characteristics that are stable across time, such as baseline morphology, heart axis, and individual electrophysiological signatures. This is critical for personalized ECG analysis and reduces the impact of inter-individual variability.

We further performed ablation experiments to assess the contribution of three main components of ECG-LFM's loss function. We removed each of these components individually and obtained three models, "ECG-LFM without the contextual contrastive loss", "ECG-LFM without the masked language modeling loss" and "ECG-LFM without the multi-segment contrastive loss". These models were pre-trained and fine-tuned using the same large-scale pretraining data and training data as the full version of ECG-LFM. The results show that ECG-LFM without the contextual contrastive loss exhibited an average 5.11% decrease in AUROC compared to the full version of ECG-LFM (Supplementary Table S2 and Supplementary Figure S3). Specifically, the performance of the model drops ranging from 3.36% to 9.72% in the CVD classification task for prediction in PTB-XL (3.36%, average AUROC=0.919), Chapman (3.65%, average AUROC=0.957), CODE-15 (5.04%, average AUROC=0.869), CPSC-2018 (3.80%, average AUROC=0.934), and UKB (9.72%, average AUROC=0.731) datasets. ECG-LFM without the masked language modeling loss showed an average 8.04% reduction in AUROC relative to the full version of ECG-LFM. In the CVD classification task, the performance decreases 7.87% for PTB-XL (average AUROC=0.876), 6.02% for Chapman (average AUROC=0.933), 8.46% for CODE-15 (average AUROC=0.838), 6.35% for CPSC-2018 (average AUROC=0.909), and 11.5% for UKB datasets (average AUROC=0.717). ECG-LFM without the multi-segment contrastive loss showed an average 1.75% reduction in AUROC relative to the full version of ECG-LFM. In the CVD classification task, the performance decreases 0.88% for PTB-XL (average AUROC=0.942), 0.59% for Chapman (average AUROC=0.987), 3.20% for CODE-15 (average AUROC=0.886), 1.22% for CPSC-2018 (average AUROC=0.959), and 2.84% for UKB datasets (average AUROC=0.787).

We have added the above content to the Discussion section of the manuscript (page 20):

"ECG-LFM leverages a multi-task learning strategy that combines contextual contrastive loss, masked language modeling (MLM) loss, and multi-segment contrastive loss, enabling the model to align local ECG features with their contextual representations, reconstruct masked segments to learn long-range relationships, and enforce patient-level consistency across ECG segments, respectively."

and Supplementary Material (page 6):

"We further performed ablation experiments to assess the contribution of three main components of ECG-LFM's loss function: the contextual contrastive loss, the masked language

modeling loss and the multi-segment contrastive loss. These components were removed individually, which leads to three models, "ECG-LFM without the contextual contrastive loss", "ECG-LFM without the masked language modeling loss" and "ECG-LFM without the multi-segment contrastive loss". These models were pre-trained and fine-tuned using the same large-scale pretraining data and training data as the full version of ECG-LFM. The results show that ECG-LFM without the contextual contrastive loss exhibited an average 5.11% decrease in AUROC compared to the full version of ECG-LFM (Supplementary Table S2 and Supplementary Figure S3). Specifically, the performance of the model drops ranging from 3.36% to 9.72% in the CVD classification task for prediction in PTB-XL (3.36%, average AUROC=0.919), Chapman (3.65%, average AUROC=0.957), CODE-15 (5.04%, average AUROC=0.869), CPSC-2018 (3.80%, average AUROC=0.934), and UKB (9.72%, average AUROC=0.731) datasets. ECG-LFM without the masked language modeling loss showed an average 8.04% reduction in AUROC relative to the full version of ECG-LFM. In the CVD classification task, the performance decreases 7.87% for PTB-XL (average AUROC=0.876), 6.02% for Chapman (average AUROC=0.933), 8.46% for CODE-15 (average AUROC=0.838), 6.35% for CPSC-2018 (average AUROC=0.909), and 11.5% for UKB datasets (average AUROC=0.717). ECG-LFM without the multi-segment contrastive loss showed an average 1.75% reduction in AUROC relative to the full version of ECG-LFM. In the CVD classification task, the performance decreases 0.88% for PTB-XL (average AUROC=0.942), 0.59% for Chapman (average AUROC=0.987), 3.20% for CODE-15 (average AUROC=0.886), 1.22% for CPSC-2018 (average AUROC=0.959), and 2.84% for UKB datasets (average AUROC=0.787)."

4.The perturbation-based EDF interpretation (PEI) method is innovative. What are the relationships of ECG traits detected by this method with the established electrophysiological principles?

RE: Thank you for this insightful question about the electrophysiological basis of our PEI method. The PEI method works by systematically perturbing each EDF and observes the resulting changes in reconstructed ECG morphology, which allows us to map deep learning features to specific ECG regions. The ECG traits detected by this method demonstrate strong alignment with established electrophysiological principles.

For example, the ECG trait, EDF#102 was identified as representing the ST-segment on V4 lead, with high importance for predicting myocardial infarction (MI) and ST/T changes. This aligns with the well-established principle that ST-segment elevation or depression reflects underlying myocardial ischemia or infarction, due to altered repolarization patterns in ischemic myocardium. Another the ECG trait, EDF#19 was mapped to P-wave duration and amplitude on V2 lead, serving as a key predictor for atrial fibrillation (AFIB). This is consistent with the known association between prolonged P-wave duration and atrial electrical remodeling, which creates the substrate for AF by prolonging atrial conduction time and increasing heterogeneity. Multiple ECG traits, EDF#945, EDF#238, and EDF#381 were identified as representing QRS morphology features, particularly increased R wave amplitude on V5 lead and deep S waves on V3 lead. These findings align with the classic ECG criteria for left ventricular hypertrophy, where increased ventricular mass leads to greater electrical forces and altered conduction patterns, resulting in characteristic voltage changes.

5.While the manuscript identifies individual genes (e.g., NFIA, TTN) associated with ECG-LFM-derived features (EDFs), a gene set enrichment analysis has been missed. This would help uncover broader biological pathways and mechanisms underlying the genetic associations.

RE: Thank you for this valuable suggestion regarding gene-based association analysis. To strengthen our gene-level findings, we performed genome-wide gene association analysis using FUMA (PMID: 29184056) on the meta-GWAS result. Loci obtained by meta-GWAS were mapped to genes in FUMA using three strategies.
(1). SNPs were mapped to genes according to physical distance (within a 10-kb window) from known protein-coding genes in the human reference assembly (GRCh37/hg19).
(2). SNPs were also mapped to genes if they have shown significant eQTL association with the genes. The eQTL mapping was performed using two data repositories (GTEx v8 and eQTL catalogue), and was based on cis-eQTLs that can map SNPs to genes up to 1 megabase apart. We used a false discovery rate of 0.05 to define significant eQTL associations.
(3). SNPs were mapped to genes using the three-dimensional spatial interactions between SNPs and genes, which potentially involve long-range interactions. This procedure was performed using Hi-C data from the study by Schmitt et al (PMID: 27851967) in the FUMA platform. The candidate genes were selected if the SNP-gene interaction region overlaps with an enhancer region according to the data from Heart tissue in the Roadmap Epigenomics Project (PMID: 25693563), and is located in the promoter region defined by Roadmap Epigenomics Project and 250 bp upstream to 500 bp downstream of the transcription start site. The interactions are considered as significant if the binomial test indicated a false discovery rate $< 1 \times 10^{-5}$ (PMID: 27851967) and modified to account for the differences in the cell lines used in this study.

FUMA analysis identified 62 EDF-associated genes. Specifically, positional mapping identified 23 genes within 10 kb of the lead SNPs, eQTL mapping revealed 21 genes with significant expression associations in heart-relevant tissues, and chromatin interaction mapping identified 42 genes through three-dimensional spatial interactions (Supplementary Table S15).
Additionally, we performed MAGMA analysis and revealed 9 genes (NFIA, TTN, SAP30L, HAND1, CDKN1A, CEP85L, PLN, KLHL38 and SIPA1L1) significantly (*P-value*<0.05/18762 $\approx 2.7\times10^{-6}$) associated with the EDFs (Figure 6C, Supplementary Table S14).

In total, 63 unique genes were identified by FUMA analysis or MAGMA. For these 63 identified genes, the gene ontology (GO) enrichment analysis by clusterProfiler (PMID: 39019974) indicated these genes are enriched in the biological process (BP) of cardiac muscle contraction ($FDR = 9.81\times10^{-3}$), cardiac muscle tissue development ($FDR = 9.81\times10^{-3}$), and cardiac muscle cell contraction ($FDR = 9.81\times10^{-3}$) (Supplementary Table S18). The KEGG enrichment analysis shows that these genes are enriched in adrenergic signalling in cardiomyocytes ($FDR = 5.31\times10^{-3}$) and dilated cardiomyopathy ($FDR = 1.06\times10^{-2}$). Thus, these genes identified by GWAS using the EDFs are potentially associated with heart functions.

We have added the above content to the Results section of Supplementary Material (page 11):
"In total, 63 unique genes were identified by FUMA analysis or MAGMA. For these 63 identified genes, the gene ontology (GO) enrichment analysis by clusterProfiler (PMID:

39019974) indicated these genes are enriched in the biological process (BP) of cardiac muscle contraction (FDR = 9.81×10^{-3}), cardiac muscle tissue development (FDR = 9.81×10^{-3}), and cardiac muscle cell contraction (FDR = 9.81×10^{-3}) (Supplementary Table S18). The kyoto encyclopedia of genes and genomes (KEGG) enrichment analysis shows that these genes are enriched in adrenergic signalling in cardiomyocytes (FDR = 5.31×10^{-3}) and dilated cardiomyopathy (FDR = 1.06×10^{-2}). Thus, these genes identified by GWAS using the EDFs are potentially associated with heart functions."

6.In Methods, limited information on hyperparameter optimization for fine-tuned models are presented.

RE: We thank the reviewer for this constructive suggestion. ECG-LFM was pre-trained on 11,571,587 ECGs using two A800 120GB GPUs with distributed data parallelism and gradient accumulation. In the pre-training process, we employed the Adam optimizer with batch size of 128, adam eps of 1×*10^{-6}*, learning rate of 5×*10^{-5}*, β_1 of 0.9 and β_2 of 0.98. The training was halted after 100 epochs. For model fine-tuning, we randomly partitioned the training data into a training set (80%) and a validation (20%) set. ECG-LFM was optimized by running on one A800 120GB GPU. Checkpoints were selected on the respective validation set using early stopping conditioned on validation loss. Early stopping is configured to halt training if the validation loss does not improve within 10 epochs. Models were trained with batch size 256 for a maximum of 100 epochs using the Adam optimizer with adam eps of 1×*10^{-8}*, learning rate of 1×*10^{-6}*, β_1 of 0.9 and β_2 of 0.98 in each epoch using an exponential scheduler. For the individual recognition task, we extracted the embedding from the pre-trained model, and fed it into a single-layer perceptron for prediction. The single-layer perceptron was optimized using the Adam algorithm with backpropagation, trained for 500 epochs with batch size of 512, adam eps of 1×*10^{-8}*, learning rate of 1×*10^{-3}*, β_1 of 0.9 and β_2 of 0.99.

We have added the above content to the Methods section (page 31):
"ECG-LFM was pre-trained on 11,571,587 ECGs using two A800 120GB GPUs with distributed data parallelism and gradient accumulation. In the pre-training process, we employed the Adam optimizer with batch size of 64, adam eps of 1×10^{-6}*, learning rate of 5*×10^{-5}*, β_1 of 0.9 and β_2 of 0.98. The training was halted after 100 epochs. For model fine-tuning, we randomly partitioned the training data into a training set (80%) and a validation (20%) set. ECG-LFM was optimized by running on one A800 120GB GPU. Checkpoints were selected on the respective validation set using early stopping conditioned on validation loss. Early stopping is configured to halt training if the validation loss does not improve within 10 epochs. Models were trained with batch size 256 for a maximum of 100 epochs using the Adam optimizer with adam eps of 1*×10^{-8}*, learning rate of 1*×10^{-6}*, β_1 of 0.9 and β_2 of 0.98 in each epoch using an exponential scheduler. For the individual recognition task, we extracted the embedding from the pre-trained model, and fed it into a single-layer perceptron for prediction. The single-layer perceptron was optimized using the Adam algorithm with backpropagation, trained for 500 epochs with batch size of 512, adam eps of 1*×10^{-8}*, learning rate of 1*×10^{-3}*, β_1 of 0.9 and β_2 of 0.99."*

7.The manuscript reports SNPs significantly associated with EDFs. How many of these SNPs are

known as associated with ECG traits? A pleiotropy analysis is needed to determine whether these SNPs influence multiple ECG traits beyond the EDFs.

RE: We thank the reviewer for this constructive suggestion. Totally, 24 independent SNPs at 19 loci were identified as significantly (*P-value* < 5×10^{-8}, r^2<0.01) associated with the EDFs (Figure 6A, Supplementary Table S10). Of the 24 significant SNPs, 16 were located <100 kb away from the lead SNPs reported by the most recently published GWAS based on more than 77,898 samples for the ECG traits including spatial QRS-T angle, PR interval, QT interval, P wave duration and QRS duration (Supplementary Figure S10 and Supplementary Table S11).
We further performed colocalization analysis to assess whether EDF-associated SNPs influence multiple ECG traits by using COLOC (PMID: 24830394). In total, we detected 11 ECG trait pairs that had colocalized signals within 7 genomic regions (PP_4 > 90%, Supplementary Table S17). The analysis indicated that the region of rs4151702 (chr6:36645988) showed colocalization (PP_4 > 0.99) among spatial QRS-T angle, PR interval and QRS duration. Besides, QT interval and PR interval are colocalized (PP_4 > 0.98) in three regions of rs7373065 (chr3:38710315), rs72967533 (chr6:118655020) and rs35165228 (chr4:114408680).
These results reveal pleiotropic effects of specific genetic loci on multiple ECG traits.

We have added the above content to Methods section of Supplementary Material (page 5):
"*Colocalization analysis was performed to assess whether the ECG trait pairs share common genetic causal variants in a genomic region. We implemented colocalization analysis by using COLOC (PMID: 24830394) to detect shared causal SNPs between the two traits. COLOC assumes at most one association per trait in a test region and uses Approximate Bayes Factor computation to generate posterior probabilities (PP) of all possible configurations between two traits: 1) H0: no association with either trait; 2) H1: association with trait 1, not with trait 2; 3) H2: association with trait 2, not with trait 1; 4) H3: association with trait 1 and trait 2, two distinct SNPs; 5) H4: association with trait 1 and trait 2, one shared SNP. The PP of each configuration is respectively denoted by PP_0, PP_1, PP_2, PP_3, and PP_4. A large PP_4 (e.g., PP_4> 90%) was considered to be strong support for colocalization in the original method publication (PMID: 24830394), which indicated a shared variant between ECG traits. Genomic regions were defined within 100kb of the SNP*."

Results section of Supplementary Material (page 11)
*"**Colocalization analysis***
Totally, 24 independent SNPs at 19 loci were identified as significantly (P-value < 5×10^{-8}, r^2<0.01) associated with the EDFs (Figure 6A, Supplementary Table S10). Of the 24 significant SNPs, 16 were located <100 kb away from the lead SNPs reported by the most recently published GWAS based on more than 77,898 samples for the ECG traits including spatial QRS-T angle, PR interval, QT interval, P wave duration and QRS duration (Supplementary Figure S10 and Supplementary Table S11). We further performed colocalization analysis to assess whether EDF-associated SNPs influence multiple ECG traits by using COLOC (PMID: 24830394). In total, we detected 11 ECG trait pairs that had colocalized signals within 7 genomic regions (PP_4 > 90%, Supplementary Table S17). The analysis indicated that the region of rs4151702 (chr6:36645988) showed colocalization (PP_4 > 0.99) among spatial QRS-T angle, PR interval and QRS duration. Besides, QT interval and PR interval are colocalized (PP_4 > 0.98) in three regions of rs7373065 (chr3:38710315), rs72967533 (chr6:118655020) and rs35165228

(chr4:114408680). These results reveal pleiotropic effects of specific genetic loci on multiple ECG traits."

8.ECG data preprocessing process needs to be visualized. Can you provide a figure to show the complete conversion process from the original ECG signal to the preprocessed ECG signal?

RE: We thank the reviewer for this constructive suggestion. We have now added a figure (Supplementary Figure S12) to visualize the complete ECG data pre-processing pipeline.

Reviewer #2 (Remarks to the Author):

This is a paper by Siying Lin et al, titled "A self-supervised electrocardiogram foundation model for empowering cardiovascular disease prediction and genetic factor discovery." The authors build a foundation model on 10+ million 12-lead ECGs, and fine-tune the model to CVD risk prediction. They extract EDF embeddings and perform GWAS and found a few SNPs, most of which were previously known. The paper is well written and flows well. There is enough description of the methods. The superior performance of ECG-LFM in predicting CVD-related diseases and phenotypes is clear. The GWAS results are good but not outstanding. I have the following comments (not sorted by minor or major comments):

RE: We are grateful to the Reviewer for taking the time to review our manuscript, and for the positive feedback on our paper.

1. Line 151: elaborate and explain more about individual recognition. What are the input data in the Chapman and CPSc-2018 datasets that were passed to the ECG-LFM? And what is the meaning of individual recognition? What is ACC measuring exactly?

RE: We thank the reviewer for this constructive suggestion.

In this work, "Individual Recognition" refers to the task of identifying the individual ID for an input ECG record. Briefly, given an ECG record, the model aims to identify which individual the ECG belongs to. It is fundamentally a multi-class classification problem at the individual level. This was completed by a procedure including two stages:

1) Feature extraction: The ECG-LFM model converts a pre-processed ECG record into a 1,024-dimensional embedding vector that encodes individual-specific cardiac patterns.
2) Classification: The embedding was mapped to the individual by training a single-layer perceptron via Multi-Class Loss (Cross-Entropy Loss):

$$\mathcal{L}_{CE} = -log(\frac{exp(s_y)}{\sum_{j=1}^{N} exp(s_j)}),$$

where s_j is the j-th logit (unnormalized score) output by the perceptron, and s_y is the logit corresponding to the correct individual ID y, and N is the number of individuals.

Regarding the data from Chapman and CPSC-2018, we have performed pre-processing using a two-stage strategy before inputting into ECG-LFM. First, a high-pass filter (1 Hz cutoff) was applied to remove baseline drift, followed by a second-order Butterworth lowpass filter (30 Hz

cutoff) to attenuate high-frequency noise, and finally a notch filter (50/60 Hz) was employed to suppress powerline interference. If the duration of ECG was less than 10 seconds (5,000 data points), zero padding was used to extend the ECG signal into 10 seconds for each lead. All signals were normalized using z-score standardization and segmented into non-overlapping 5s segments. For each individual with multiple ECG segments, we randomly selected one segment per individual to construct the test set, and all remaining segments construct the training set. Finally, from Chapman dataset, 10,646 segments were used to construct the training set and 10,646 segments were used to construct the test set. From CPSC-2018 dataset, 13,286 segments were used to construct the training set and 6,876 segments were used to construct the test set.

The performance of individual recognition was evaluated by Top-3 Classification Accuracy (Top3-ACC). The formula is:
Top3-ACC = (Number of test samples where the true individual ID is among the top-3 predicted labels) / (Total number of test samples)
During testing, the single-layer perceptron computes the scores (after softmax) to obtain a probability distribution over all candidate individuals. The predicted individual IDs are then ranked in descending order of their corresponding probabilities. A prediction is considered correct if the true individual ID appears within the top 3 of this ranked list. The Top-3 ACC aggregates this result over all test segments from all test individuals, providing the overall fraction of test samples for which the true identity is recovered within the model's three highest-confidence predictions.

We have added this description to the Methods section (page 32): "
Individual recognition using ECG-LFM derived features
To evaluate the capability of ECG-LFM for individual recognition, we pre-processed the ECG records as described in the 'ECG Data Pre-=processing' section. For each individual, we randomly selected one 5-second segment to construct the test set, and all remaining segments construct the training set. By using this strategy, from Chapman dataset, 10,646 segments were used to construct the training set and 10,646 segments were used to construct the test set. From CPSC-2018 dataset, 13,286 segments were used to construct the training set and 6,876 segments were used to construct the test set.

Each ECG segment was passed through the ECG-LFM to obtain the 1024-dimensional embedding, then we trained a single-layer perceptron to map the embedding to one of the N individual IDs via cross-entropy (CE) loss:

$$\mathcal{L}_{CE} = -log(\frac{exp(s_y)}{\sum_{j=1}^{N} exp(s_j)}),$$

where s_j is the j-th logit (unnormalized score) output by the perceptron, and s_y is the logit corresponding to the individual ID y_i. During testing, we computed the scores (softmax normalized probabilities) and the predicted individual ID is considered correct if the true individual ID appears within the top 3 of the ranked scores. We evaluated the individual recognition performance using Top-3 Classification Accuracy (Top-3 ACC). The overall accuracy is calculated as:

$$Top-3\ ACC = \frac{\sum_{i=1}^{M} \mathbb{I}(y_i \in Top_3(p_i))}{M}$$

where M is the total number of test samples, y_i is the true individual ID of the i-th test sample, p_i is the predicted normalized probability vector output by the model for the i-th sample"

2. Line 166: The authors used the embeddings of ECG-LFM to predict cardiac arrhythmia in the Chapman and CPSC-2018 datasets. How was the cardiac arrhythmia defined? Based on ICD codes? What ICD codes. How specific was it: AF, cardiomyopathy, etc?

RE: We thank the reviewer for this constructive suggestion.
In the Chapman and CPSC-2018 datasets, the cardiac arrhythmia is annotated by experts according to electrocardiographic diagnostic criteria rather than ICD codes. The electrocardiographic diagnostic criteria were defined according to standard clinical guidelines (e.g., AHA/ACC/HRS recommendations) for the standardization and interpretation of the electrocardiogram. Specifically, each ECG recording was independently labelled as representing certain diseases or not by licensed cardiologists.

Here, cardiac arrhythmia is considered as a disease cluster including atrial fibrillation (AFIB), conduction disturbance (CD), sinus bradycardia (SB), general supraventricular tachycardia (GSVT), and ventricular ectopics (VE). The individuals with ECG recorded diagnosed as one of the above diseases are considered as cardiac arrhythmia cases. Otherwise, they are considered as controls.

And how/what cardiac arrhythmia information was used in the fine-tuning process?

RE: For the CVD prediction, we performed full-model fine-tuning. Here, the cardiac arrhythmia information was used as a ground-truth label for signals.
First, 12-lead ECG signals were input into the model, and the model was supervised to predict expert-annotated labels from datasets. In the model training process, we added a classification head to the ECG-LFM. During fine-tuning, the weights of both the feature encoder and the transformer layers were updated using a binary cross-entropy (BCE) loss function to minimize the difference between the predicted probability and the ground-truth labels for each CVD. With these strategies, the model was allowed to transition from a general-purpose representation (learned during pre-training on 10 million ECGs) to a highly specialized diagnostic tool for specific clinical categories.

Authors mentioned that the model was fine-tuned for CVD risk, but this needs more explanation to show what data went into the fine-tuning step and what CVD risk or information was accounted for.

RE: The input data for fine-tuning were the pre-processed 12-lead ECG signals as described in Methods section (e.g., 10-second windows, resampled to 500 Hz, filtered, and normalized). Briefly, to standardize ECG data from multiple sources, we applied linear interpolation to sample the data to force the sampling frequency to be 500Hz. To address signal artifacts, we implemented a series of filtering steps. First, a high-pass filter (1 Hz cutoff) was applied to remove baseline drift, followed by a second-order Butterworth lowpass filter (30 Hz cutoff) to attenuate high-frequency noise, and finally a notch filter (50/60 Hz) was employed to suppress powerline interference. If the duration of ECG was less than 10 seconds (5,000 data points), zero

padding was used to extend the ECG signal into 10 seconds for each lead. During fine-tuning, the model was trained using a Binary Cross-Entropy (BCE) loss function to minimize the difference between the predicted probability and the ground-truth labels for each CVD.

Also, what type of CVD modalities were included: Arrhythmias, CHD, long QT, AF, etc.? You mention 8 types of CVs on page 180. Although it might be detailed in the methods/supp methods, this should be explicitly mentioned in the results for better reading.

RE: We thank the reviewer for this suggestion. The ECG-LFM was fine-tuned for predicting 8 types of CVD. These CVDs are conduction disturbance (CD), Hypertrophy (HYP), Myocardial infarction (MI), ST/T change (STTC), Ventricular ectopics (VE), Sinus bradycardia (SB), Atrial fibrillation (AFIB) and General supraventricular tachycardia (GSVT). As suggested by the reviewer, these 8 types of CVD now have been described in the Results section.

We have revised the Results section to include more detailed description of the CVDs (page 10): *"To evaluate the performance of ECG-LFM in predicting risks of CVDs, we fine-tuned the model for predicting eight types of CVDs that are conduction disturbance (CD), hypertrophy (HYP), myocardial infarction (MI), ST/T change (STTC), ventricular ectopics (VE), sinus bradycardia (SB), atrial fibrillation (AFIB) and general supraventricular tachycardia (GSVT) for individuals from the PTB-XL and Chapman datasets."*

3. Line 278: What is the rational behind the selection of 16 important EDFs and 215 related phenotypes and what they are? This needs to be specified in supp material.

RE: Thank you for this critical question, which allows us to clarify the rationale behind selecting the 16 important ECG-LFM Derived Features (EDFs) and the 215 cardiovascular disease (CVD)-related phenotypes, and to specify their biological meaning.

(1). Rationale for selecting the 16 important EDFs
The selection of the 16 EDFs was based on their quantified importance in predicting the eight CVDs in the fine-tuning task, as determined by a rigorous feature attribution method. The importance of EDFs in predicting CVDs is quantified by Integrated Gradients (IG) that computes an importance score for each of the 1,024 EDFs. The feature importance evaluation was performed on 2,000 samples from the PTB-XL and Chapman test sets.
From 1,024 EDFs, the most important top 5-EDFs for predicting each CVD have been selected for further analysis, resulting in a total of 16 unique EDFs. Analysing the biological meaning of these important EDFs, rather than all 1,024, allows us to focus interpretation and downstream genetic/phenotypic association analyses on the most clinically relevant signal, reducing multiple testing burden and enhancing biological discovery.

(2). The selection of phenotypes for PheWAS analysis
To obtain the represent landscape associated with cardiovascular risk factors, we employed quality control (QC) to select phenotypes from the UK Biobank (UKB). Totally, 264 phenotypes are considered, which are physical measures, blood assays, cardiac magnetic resonance imaging (CMRI) traits, ECG traits, and disorder outcomes. The criteria of QC used for phenotype

selection are followings. First, if less than 100 cases have the measurement of phenotypes, these phenotypes are excluded from the analysis. Moreover, if the individuals missed more than 20% phenotypes, the individuals are excluded from the analysis. For continuous phenotypes, the normality is assessed via Shapiro-Wilk tests, with non-normal distributions transformed (e.g., log-based inverse normalization) and extreme outliers (beyond ±4 SD) are excluded. Binary phenotypes with low prevalence (case counts <100) are filtered out to avoid unstable associations. Phenotypic correlations are computed to remove highly redundant traits (pairwise $r > 0.9$). Following these QC procedures, a total of 215 phenotypes remain for PheWAS analysis, including 34 physical measures, 16 blood assays, 147 cardia magnetic resonance imaging (CMRI) traits, 11 ECG traits, and 7 disorder outcomes. Detailed information about these 215 phenotypes is shown in Supplementary Table S5.

We have added the above content to Supplementary Material (page 2):
"We collected a total of 264 phenotypes to represent a comprehensive landscape of known and putative cardiovascular risk factors from the UK Biobank (UKB), including 56 physical measures, 34 blood assays, 155 cardiac magnetic resonance imaging (CMRI) traits, 12 ECG traits, and 7 disorder outcomes. Then we performed phenotype quality control (QC) for these 264 phenotypes before phenotype-wide association study (PheWAS). Phenotypes with excessive missingness (case counts <100) are excluded, and individuals with high rates of missing data (>20% of phenotypes) are removed. For continuous phenotypes, normality is assessed via Shapiro-Wilk tests, with non-normal distributions transformed (e.g., log-based inverse normalization) and extreme outliers (beyond ±4 SD) are excluded. Binary phenotypes with low prevalence (case counts <100) are filtered out to avoid unstable associations. Phenotypic correlations are computed to remove highly redundant traits (pairwise r > 0.9). Following these QC procedures, a total of 215 phenotypes remain for PheWAS analysis, including 34 physical measures, 16 blood assays, 147 cardia magnetic resonance imaging (CMRI) traits, 11 ECG traits, and 7 disorder outcomes. Detailed information about these 215 phenotypes is shown in Supplementary Table S5."

4. Line 297: It is important to add the effect size of associations.

RE: We thank the reviewer for this constructive suggestion. We have revised the Results section (page 15) to include effect size alongside the *P-value* for all reported associations:
*"To identify the associations between 16 important EDFs and 215 CVD-related phenotypes, a phenome-wide association (PheWAS) study was performed by using information on 41,235 individuals from the UKB dataset. A total of 1,387 significant associations were identified (Supplementary Table S6, P-value<0.05/(16*215)* $\approx 1.5\times10^{-5}$*) between these EDFs and 167 CVD-related phenotypes that are 32 physical measures, 5 blood assays, 112 cardia magnetic resonance imaging (CMRI) traits, 11 ECG traits, and 7 disorder outcomes (Supplementary Table S6). The associations of EDFs with the CMRI traits and ECG traits are more significant than with other phenotypes (Figure 5A). Specifically, the EDFs are more likely to be associated with the ECG traits, including ventricular rate (beta=-0.321, P-value <* 1.0×10^{-250}*), QT interval (beta=0.241, P-value <* 1.0×10^{-250}*), RR interval (beta=0.321, P-value <* 1.0×10^{-250}*), PP interval (beta=0.337, P-value <* 1.0×10^{-250}*) and QRS num (beta=-0.488, P-value <* 1.0×10^{-250}*), indicating the electrophysiological linkages between EDFs and fundamental cardiac conduction properties. Moreover, the EDFs are significantly associated with CMRI traits, including left atrium max*

volume - 2 channel (beta=-0.131, P-value < 1.0×10^{-250}), heart rate during PWA (beta=-0.205, P-value < 1.0×10^{-250}), number of beats in waveform average for PWA (beta=0.152, P-value = 1.0×10^{-176}) and LV circumferential strain AHA 9 (beta=-0.139, P-value = 6.7×10^{-153}). This result indicates that EDFs can reflect underlying cardiac structure and function as measured by imaging techniques.
Interestingly, the EDFs are found significantly associated with the phenotypes of blood assays that are Cholesterol (beta=0.049, P-value = 2.7×10^{-20}), Sodium in urine (beta= -0.035, P-value = 1.4×10^{-11}), Alanine aminotransferase (beta= -0.031, P-value = 5.3×10^{-10}), and Alkaline phosphatase (beta=-0.030, P-value = 9.6×10^{-9}). Many physical measures are also identified as significantly associated with the EDFs, which include leg fat mass (right) (beta= -0.095, P-value = 4.3×10^{-114}), leg fat mass (left) (beta= -0.094, P-value = 1.2×10^{-113}), whole body fat mass (beta= -0.117, P-value = 2.0×10^{-112}), and trunk fat mass (beta=-0.117, P-value = 1.3×10^{-107}). These findings have demonstrated that the EDFs have the potential to monitor the risk factors associated with cardiovascular diseases."

5. Line 310: I wonder why GWAS did not discover snps/loci that were previously reported for ECG intervals? Phenome-wide association analysis showed very significant correlations between EDF and ECG intervals.

RE: We thank the reviewer for this constructive suggestion, and apologize for the unclear description of the result.

Totally, 12 SNPs were identified as associated with ECG by our GWAS. Among them, 4 were reported as associated with ECG intervals by previous studies. For example, SNP rs6587924 has been identified as associated with EDF#900 by the GWAS in this study and known to be associated with PR interval (PMID: 32439900) and QT interval (PMID: 36050321). Additionally, the SNP rs7373065 was identified as associated with EDF#945 by the GWAS in this study, and known to be associated with QT interval (PMID: 3627187). Furthermore, these 12 SNPs are adjacent to 7 genes, PLEKHA3, HAND1, FAM53A, SCN5A, NFIA, MIR297, and TBX3. Out of them, HAND1, the regulator of conduction system development is known to be associated with QT interval in previous GWAS studies (PMID: 36050321), SCN5A is known to be associated with PR interval (PMID: 21347284) and QT interval (PMID: 19305408), NFIA has been reported in GWAS to be associated with PR interval (PMID: 32439900) and QT interval (PMID: 36050321), and TBX3 is involved in cardiac conduction system development and has been associated with PR interval (PMID: 30679814) and RR interval (PMID: 38507016). This finding is consistent with PheWAS analysis showing significant associations (*P-value* < 1.0×10^{-250}) between EDF and ECG intervals.

We have revised the Results section to make more clearly description (page 16).
"Notably, several of the identified genes (HAND1, SCN5A, NFIA, and TBX3) have been reported to be associated with PR interval (PMID: 21347284), RR interval (PMID: 38507016), and QT interval (PMID: 19305408), consistent with the PheWAS analysis showing significant associations (P-value < 1.0×10^{-250}) between EDF and ECG intervals."

6. In Figure 5A, were the 8 CVD types included? I want to see AF, MI, etc. on that plot and see how they associate with the 16 EDF features. They should be annotated.

RE: We thank the reviewer for this constructive suggestion. We have revised Figure 5A to annotate the associations between EDF and CVDs.

7. Authors should check how heritable the EDFs are. EDFs don't seem to be well powered for the GWAS analysis. Not many loci were observed despite the relatively good sample size.

RE: We thank the reviewer for this constructive suggestion.
We estimated the inflation factor (λ) of GWAS for the 16 important EDFs, and achieved a mean λ of 1.027 (ranging from 1.000 to 1.047), indicating minimal genomic inflation. The QQ plots for all 16 EDFs were shown in Supplementary Figure S9, which confirms the absence of systematic inflation.

Additionally, we estimated the liability-scale heritability (h^2) of the 16 EDFs using single-trait linkage disequilibrium score regression (LDSC). We conducted quality control for SNPs inputting into LDSC analysis by filtering SNPs using the HapMap3, removing the SNPs if they were strand-ambiguous, removing SNPs whose minor allele frequency (MAF)<0.01 and removing SNPs located in the major histocompatibility complex (MHC) region (chromosome 6: 28,477,797–33,448,354). As shown in Supplementary Table S8, among 16 EDFs, 14 EDFs have heritability (h^2) with false discovery rate (FDR) < 0.05 and a median h^2 of 0.103 (ranging from 0.030 to 0.152), suggesting that EDFs are highly polygenic.

We have added the above content to Results section of Supplementary Material (page 9):
"We estimated the inflation factor (λ) for the 16 important EDFs. GWAS of the EDF achieved a mean λ of 1.027 (ranging from 1.000 to 1.047), indicating minimal genomic inflation. The QQ plots for all 16 EDFs were shown in Supplementary Figure S9, which visually confirm the absence of systematic inflation. Additionally, we estimated the liability-scale heritability (h^2) of the 16 EDFs using single-trait linkage disequilibrium score regression (LDSC). SNPs inputting into LDSC performed quality control by filtering SNPs using the HapMap3, removing the SNPs if they were strand-ambiguous, removing SNPs whose minor allele frequency (MAF)<0.01 and removing SNPs located in the major histocompatibility complex (MHC) region (chromosome 6: 28,477,797–33,448,354). As shown in Supplementary Table S8, among 16 EDFs, 14 EDFs have heritability (h^2) with false discovery rate (FDR) < 0.05 and a median h^2 of 0.103 (ranging from 0.030 to 0.152), suggesting that EDFs are highly polygenic."

8. Italicize gene names throughout the manuscript.
RE: We thank the reviewer for this constructive suggestion. We have revised the manuscript to italicize all gene names.

9. In the MR analysis, Line 327, authors mention EDF or CVDs as exposure. I understood that IVs ◊ EDF ◊ CVDs is the MR framework, which makes CVDs as outcomes. Is that correct? Or authors tried IDs ◊ CVD ◊ EDFs? This should be clearly mentioned in both methods and results.

RE: We thank the reviewer for this constructive suggestion.
The reviewer is correct. The MR analysis followed the standard framework: IVs → Exposure → Outcome. Specifically, we performed bidirectional MR analyses to test causal relationships in both directions. When EDF is taken as exposure and CVD as outcome, the MR is to test whether

genetic variants associated with EDFs causally influence CVD risk. When CVD is considered as exposure and EDF is considered as outcome, the MR is to test whether genetic variants associated with CVDs causally influence EDF expression.

We have revised both the Methods section (page 35) and Results section (page 17) to clearly specify the bidirectional nature of our MR analysis.

Methods

"We performed bidirectional Mendelian Randomization (MR) analyses to investigate causal relationships between EDFs and CVDs. The MR analysis is to test two causal directions: (1) EDF as exposure and CVD as outcome, and (2) CVD as exposure and EDF as outcome. When EDF is taken as exposure and CVD as outcome, the MR is to test whether genetic variants associated with EDFs causally influence CVD risk. When CVD is considered as exposure and EDF is considered as outcome, the MR is to test whether genetic variants associated with CVDs causally influence EDF expression."

Results

"In MR analysis, a significant causal relationship was determined if the P-value surpassed the Bonferroni-corrected significance level (i.e., P-value $< 0.05/(5\times16\times2) \approx 3.1 \times 10^{-4}$, considering two causal directions). The bidirectional MR analysis revealed significant causal relationships between EDFs and the CVDs (Figure 5B). When using CVD as exposure and EDF as outcome, the analysis indicated AF as a putative causal effect on the decreased EDF#700 ($P_{IVW} = 6.9\times10^{-16}$, OR = 0.968, 95% CI of 0.961 to 0.976), EDF#19 ($P_{IVW} = 3.7\times10^{-9}$, OR = 0.983, 95% CI of 0.978 to 0.989) and EDF#238 ($P_{IVW} = 1.4\times10^{-4}$, OR = 0.982, 95% CI of 0.973 to 0.992) (Figure 5B, Supplementary Table S9)."

10. Line 364: I would not call TTN a novel finding. TTN is well know for ECG-related phenotypes, AF, cardiomyopathy, arrhythmias, etc.

RE: Thank you for this important clarification. We agree that TTN is indeed a well-established gene for ECG-related phenotypes, AF, cardiomyopathy, and arrhythmias. We apologize for the inaccurate description. We have revised the manuscript to claim that TTN is an established gene and emphasize the novelty of the specific variant at the Discussion section (page 23):

"Among these novel SNPs, the lead SNP rs2562829 is near to TTN that is a gene known to be associated with QT interval (PMID: 29874175) and PR interval (PMID: 30046033). While TTN is an established gene for arrhythmia (PMID: 29290336), the specific variant rs2562829 represents a novel association with cardiac conduction."

11. What was the rational behind performing PCA on top of the 1024 EDFs? The embeddings are representations of the ECGs, then PCA is a linear combination of these embeddings. GWAS was performed on 16 EDFs, then now on 10 PCs. What do we expect to find and what is the biological difference between the use of both of these traits. How do authors interpret these results since the tested phenotypes are not explainable anymore?

RE: Thank you for these insightful questions regarding our analytical approach. We appreciate the opportunity to clarify the rationale behind performing PCA on the 1024 EDFs and the biological interpretation of our findings.

The 1,024 EDFs collectively capture the complete ECG signal representation learned by our deep learning model, but their high dimensionality poses challenges for genetic analysis. Thus, the primary motivation for applying PCA to the 1,024 EDFs is to capture the major axes of electrocardiographic variability in a more interpretable and statistically efficient manner.

Besides, from these 1,024 EDFs, 16 EDFs were selected for GWAS, which have shown high importance in predicting risks of CVDs. The biological meaning of these EDFs was interpreted by a perturbation-based EDF interpretation (PEI) method. The EDF was perturbed and used to reconstruct ECG signals. The differences between the reconstructed ECG signals and the original ECGs were evaluated by paired t-tests. Regions with higher absolute t-values were identified as most likely to be represented by the EDFs, and the smoothed t-value maps highlighted these regions. The method was used on 16 EDFs, and successfully mapped EDFs to clinically meaningful ECG features. For example, EDF#102 is indicated as ST-segment on V4 lead (key for MI and STTC prediction), EDF#19 is indicated as P-wave duration and amplitude on V2 lead (key for AF prediction), and EDF#945, #238, #381 are indicated as QRS morphology features (key for HYP prediction). These findings are biologically consistent with the established ECG-pathology relationships that ST-segment changes reflect ischemic processes in MI, and P-wave abnormalities are known markers of atrial conduction issues in AF and QRS morphological changes indicate ventricular hypertrophy.

To make a clear description, we have revised the Results sections (page 18).
"Most of the GWAS for ECG are performed only for the known risk factors associated with CVDs because of the difficulty to reduce the data dimension and keep the biological meaning of ECG signal. To expand genetic insights into ECG, we extracted the first 10 principal components (PCs) from 1,024 EDFs, and performed GWAS for each PC on 41,235 individuals of UKB dataset. These 10 PCs accounted for 85.2% of the total variance in 1,024 EDFs, suggesting they capture the majority of electrocardiographic variability of the individuals. For 10 PCs of the EDFs, multi-trait genome-wide meta-analysis was performed using METAL."

12. Line 366 – 371: consistency lacks in reporting results. Sometimes Z-scores are reported and sometimes are not.

RE: Thank you for pointing out this inconsistency in our results reporting. We have revised the manuscript to report all association results using beta coefficients consistently throughout the Results section (page 19):
"Especially, the variant rs6843082, located near MIR297 is reported to be significantly associated with atrial fibrillation (beta=-0.371, P-value=3.4×10^{-155}), generalized supraventricular tachycardia (beta=-0.100, P-value=1.3×10^{-8}), and heart failure (beta=-0.075, P-value=2.6×10^{-15}), indicating its pleiotropic effects on cardiac conduction and ventricular function (Figure 6B)."

13. Line 372: in this section, GWAS was performed on PCs. But here on this line, we go back to

EDFs and test gene-level associations. This should be moved to the previous section. Moreover, MAGMA was performed on the meta-GWAS results: was it for the multi-trait analysis performed by METAL on all PCs? It is confusing since authors mention EDFs in this section and not PCs, given that no multi-trait analysis on EDF was performed.

RE: We thank the reviewer for this constructive suggestion. We appreciate the opportunity to clarify our analytical approach and will restructure the manuscript accordingly. Our analysis includes three steps:
Step 1: Perform GWAS on 10 PCs of 1,024 EDFs;
Step 2: Perform multi-trait meta-analysis (METAL) on the 10 PCs GWAS for EDFs;
Step 3: Use the meta-analysis summary statistics from Step 2 for MAGMA gene-based association analysis.

We have revised the Methods and Results sections (page 19) to clarify the description.

Methods
"*Meta-analysis for EDF*
To identify SNPs and genes associated with EDFs, the meta-analysis was performed by three stages. First, we performed GWAS on 10 PCs of 1,024 EDFs. Then, multi-trait meta-analysis was performed on the 10 PCs GWAS for EDFs using METAL. Finally, the meta-analysis summary statistics from Step 2 were used for gene-based association analysis to identify genes associated with EDFs.
"

Results
"To identify genes associated with the EDFs, Multi-marker Analysis of GenoMic Annotation (MAGMA) (PMID: 25885710) and FUMA (PMID: 29184056) on the meta-GWAS were performed. In total, 63 unique genes were identified by at least one of the three FUMA gene mapping strategies or considered as significant by MAGMA. Briefly, MAGMA analysis revealed 9 genes (NFIA, TTN, SAP30L, HAND1, CDKN1A, CEP85L, PLN, KLHL38 and SIPA1L1) as significantly (P-value<0.05/18762 $\approx 2.7 \times 10^{-6}$) associated with the EDFs (Figure 6C, Supplementary Table S14), FUMA identified a total of 62 candidate genes (Supplementary Table S15). Out of these 63 genes, 39 are reported as associated with ECG traits, such as QRS duration (PMID: 31217584), QT interval (PMID: 24952745), and PR interval (PMID: 32439900). Moreover, 28 of them were found by previous studies to be associated with CVDs in the GWAS Catalog (PMID: 30445434), such as myocardial infarction (PMID: 38072966), heart failure and (PMID: 31919418) atrial fibrillation (PMID: 29892015). The MAGMA analysis indicated SAP30L as the most significantly associated with the EDFs (P-value=2.8×10^{-13}). SAP30L is known to play roles in the regulation of cardiac development and heart function (PMID: 22821512). These findings suggest that the genes identified in our analysis may be involved in the genetic architecture of ECG features and cardiovascular diseases."

14. Typo gene name on Line 380. PLM should PLN?.

RE: We sincerely thank the reviewer for their meticulous reading and for pointing out this typo. The gene name on Line 380 should indeed be PLN, not PLM. We have corrected "PLM" to "PLN" in the revised manuscript.

15. Line 385: tone down that sentence. Based on the results, it is not convincing that "genes identified as associated with EDFs potentially play roles in regulating ECG and CVDs." Most of what was shown is only "association" from GWAS studies.
RE: Thank you for this important feedback regarding the strength of our conclusions. We have revised the sentence to more accurately reflect the evidence (page 20):
"These findings suggest that the genes identified in our analysis may be involved in the genetic architecture of ECG features and cardiovascular diseases."

16. In the discussion: the authors claim the generalizability of ECG-LFM to diverse populations and clinical scenarios. They need to mention how this was checked. They need to show statistics about the datasets they used and that they contain individuals from diverse ancestral groups and are exposed to various environmental exposure, lifestyle, etc.

RE: We thank the reviewer for this constructive suggestion.
The generalizability of ECG-LFM was evaluated across multiple dimensions. First, we tested the model on independent datasets with diverse demographic characteristics: PTB-XL (*n*=18,118, European ancestry), Chapman (*n*=10,646, Chinese ancestry), CPSC-2018 (n=6,877, Chinese ancestry), and UK Biobank (*n*=41,235, mainly European ancestry). These datasets represent populations from different geographic regions and healthcare systems. Second, the datasets include individuals with varying clinical presentations, ranging from healthy controls to patients with acute myocardial infarction, arrhythmias, and conduction disorders. Third, the model was evaluated on ECGs acquired using different recording devices and protocols, demonstrating robustness to technical variations. Detailed information for these datasets was shown in Supplementary Table S1 and Supplementary Figure S1. The consistent performance of ECG-LFM across these diverse settings (average AUROC of 0.930, ranging from 0.810 to 0.993) supports its generalizability to different populations and clinical scenarios.

We have added above content in the Discussion section to explicitly address generalizability (page 21). *"The generalizability of ECG-LFM was evaluated across multiple dimensions. First, we tested the model on independent datasets with diverse demographic characteristics: PTB-XL (N=18,118, European ancestry), Chapman (N=10,646, Chinese ancestry), CPSC-2018 (N=6,877, Chinese ancestry), and UK Biobank (N=41,235, mainly European ancestry). These datasets represent populations from different geographic regions and healthcare systems. Second, the datasets include individuals with varying clinical presentations, ranging from healthy controls to patients with acute myocardial infarction, arrhythmias, and conduction disorders. Third, the model was evaluated on ECGs acquired using different recording devices and protocols, demonstrating robustness to technical variations. Detailed information for these datasets was shown in Supplementary Table S1 and Supplementary Figure S1. The consistent performance of ECG-LFM across these diverse settings (average AUROC of 0.930, ranging from 0.810 to 0.993) supports its generalizability to different populations and clinical scenarios."*

We have also added a supplementary table (Supplementary Table S1) with detailed demographic and clinical characteristics of each dataset such as Sample size, age, sex, ancestry information and ECG information.

17. In the discussion, line 427 paragraph, I disagree with the claim that a groundbreaking aspect of the study was used to uncover genetic association underlying ECG morphology. GWAS on the 16 EDFs did not reveal many loci despite the relatively large sample size, and the multi-trait analysis on PCs revealed some loci that are already known for CVD (AF, ECG intervals, etc.).

RE: Thank you for this important feedback regarding the strength of our claims about the genetic discovery aspect of our study. We agree that the term "groundbreaking" may be overly strong and revised our language to more accurately reflect the contribution of our work (page 22):
"A key contribution of this study is on using EDFs to uncover genetic associations underlying ECG morphology."

18. Contributions: authors should add who performed the various types of analysis, building the foundation model, data cleaning, GWAS, etc.
RE: Thank you for your suggestion to clarify author contributions. We have revised the "Contributions" section to provide specific details about each author's role in the various analyses and manuscript preparation:
"H.Z., Y.Y. and S.L. conceived the basic idea. S.L. designed the algorithm and built the model. Q.W. and Y.C. collected the data, pre-processed the data and carried out benchmark experiments. S.L. and Z.L. participated in the case analysis. S.L., H.Z. and Y.Y. prepared the figures and wrote the manuscript. All authors performed literature searches, verified the underlying data, edited the text, critically read the manuscript and approved the final version for submission. H.Z. provided final edits to the manuscript."

Reviewer #3 (Remarks to the Author):

This article aims to extend ECG analysis toward the use of large-scale foundation models for cardiac signal interpretation. Traditional ECG algorithms—whether rule-based or task-specific deep learning models—are limited by their reliance on manually labeled data and their narrow focus on predefined diagnostic tasks. In contrast, an Electrocardiogram Large-Scale Foundation Model (ECG-FM) leverages massive unlabeled datasets and self-supervised learning to capture the full spectrum of temporal and morphological patterns embedded in ECG signals.

In addition, the authors attempt to relate the outcomes of their proposed model to genotypes. Overall, this is a very interesting paper, which could benefit from a deeper GWAS and post-GWAS analysis.
RE: We are grateful to the Reviewer for taking the time to review our manuscript, and for the positive feedback on our paper.
1/ Some caution is needed regarding claims of novelty. First, HEARTBEiT was trained on more than 8 million observations, which is of the same order of magnitude (although somewhat smaller) than the present study. Second, as I understand it, the authors aim to associate latent factors generated by the model with genetic variants. While this is a commendable objective, it is

not entirely novel from a methodological standpoint—other studies have already linked ECG-derived features to genotypes (e.g., work from the van der Harst group in the UK Biobank). This should be acknowledged and perhaps rephrased. That said, this does not diminish the relevance or interest of the analysis itself.

RE: We thank the reviewer for this constructive suggestion.

We have revised the manuscript accordingly.
The sentence that states "existing models are typically trained on relatively small datasets" has been revised as "*While some recent foundation models like HeartBEiT have been trained on substantial datasets (over 8 million ECGs), the potential of models pre-trained on more than 10 million ECGs remains to be fully explored*."

Moreover, we have cited relevant work from the van der Harst group who have linked analyses in deep learning-derived features the UK Biobank to genetic variants. "*While previous studies have linked deep learning-derived features to genetic variants (PMID: 40822352), our approach uniquely combines deep representations from an ECG foundation model with comprehensive genetic and phenotypic analyses to provide novel insights into the genetic architecture of cardiac electrophysiology*."

2/ Although ECG-LFM benefits from large-scale self-supervised pretraining, the article clearly shows that its downstream performance still depends on supervised fine-tuning using labeled ECG datasets. High-quality annotations therefore remain necessary to adapt the foundation model to specific clinical prediction tasks.

RE: We thank the reviewer for this constructive suggestion, and apologize for the unclear description of the method.
The application of ECG-LFM in downstream analysis requires input original ECG data but not labelled ECG data. This is different from the traditional model that requires high-quality annotated ECG as input to train a model for disease risk prediction. We have added this content in the Discussion section (page 23).

Discussion
"Traditional GWAS rely on handcrafted ECG features (e.g., QT interval), which are limited in their ability to capture complex waveform interactions. In contrast to traditional models that necessitate high-quality labelled datasets for downstream analysis, ECG-LFM leverages self-supervised pre-training to extract comprehensive, label-independent representations, referred to as ECG-LFM derived features (EDFs). While previous studies have linked deep learning-derived features to genetic variants (PMID: 40822352), our approach uniquely combines deep representations from an ECG foundation model with comprehensive genetic and phenotypic analyses to provide novel insights into the genetic architecture of cardiac electrophysiology."

3/ GWAS analysis. This genetic component is important because it can help reveal the new information extracted by the proposed algorithms. However, it seems somewhat limited by sample size, and it is a bit disappointing to identify only 19 independent association signals, given what is known about the genetic architecture of standard ECG traits and CVDs. Since this

section is central to the paper, I recommend a more comprehensive comparison between the genetic architecture of the EDFs and those of ECG or CVD traits as reported in the literature. A sensible first step would be to report genetic correlations rather than focusing exclusively on genome-wide significant SNPs (for example, via LD-score regression). The current PheWAS and GWAS analyses alone do not provide a sufficiently clear picture of genetic correlations (line 798).

RE: We thank the reviewer for this constructive suggestion.
To investigate the genetic architecture of EDF, we extracted the first 10 principal components (PCs) from 1,024 EDFs (EDF-PC), and performed GWAS for each PC on 41,235 individuals of UKB dataset. Then we used cross-trait (LDSC) to estimate genetic correlations between the GWAS of 10 EDF-PCs and 5 ECG traits (Spatial QRS-T angle, PR interval, QT interval, P wave duration, QRS duration) and 7 CVDs (AF, AP, CD, GSVT, HYP, HF, MI). For ECG traits, a total of 6 pairs of the 12 EDF-PCs and ECG traits were found with a nominally significant r_g (*P-value*<0.05, $|r_g|$ ranging from 0.134 to 0.445, Supplementary Table S16). Moreover, we identified 7 nominally significant genetic correlations (*P-value*<0.05, $|r_g|$ ranging from 0.150 to 0.400) between EDF-PCs and the CVDs, atrial fibrillation, coronary artery disease, and heart failure (Supplementary Table S16). These findings suggest that the genetic architecture underlying our deep learning-derived EDFs is correlated with standard ECG traits and CVDs.
We have added above content to Methods and Results section of Supplementary Material.

Methods (page 4 of Supplementary Material)
*"**Genetic correlation estimation***
Linkage disequilibrium score regression (LDSC) is used to estimate genetic correlations between traits based on GWAS summary statistics. Firstly, we conducted quality control for SNPs input into LDSC analysis by: (1). filtering SNPs using the HapMap370; (2). removing the SNPs if they were strand-ambiguous; (3). removing SNPs whose minor allele frequency (MAF)<0.01; (4). removing SNPs located in the major histocompatibility complex (MHC) region (chromosome 6: 28,477,797–33,448,354). The European-ancestry population of 1000 Genomes Project Phase 3 was used as the reference panel for the analysis. Then, LDSC was performed, under a weighted linear model, by regressing the product of Z-statistics of two traits with the LD scores of the genome-wide SNPs."

Results (page 10 of Supplementary Material)
*"**Genetic correlations of ECG-LFM derived features with ECG traits and CVDs***
To investigate the genetic architecture of EDF, we extracted the first 10 principal components (PCs) from 1,024 EDFs (EDF-PCs), and performed GWAS for each PC on 41,235 individuals of UKB dataset. Then we used cross-trait (LDSC) to estimate genetic correlations (r_g) between the GWAS of 10 EDF-PCs and 5 ECG traits (Spatial QRS-T angle, PR interval, QT interval, P wave duration, QRS duration) and 7 CVDs (AF, AP, CD, GSVT, HYP, HF, MI). For ECG traits, a total of 12 pairs of the EDF-PCs and ECG traits were estimated with a nominally significant r_g (P-value<0.05, $|r_g|$ ranging from 0.134 to 0.445, Supplementary Table S16). For CVDs, we identified 7 nominally significant genetic correlations (P-value<0.05, $|r_g|$ ranging from 0.150 to 0.400) between EDF-PCs and various CVDs, including atrial fibrillation, coronary artery disease, and heart failure (Supplementary Table S16). These findings suggest that the genetic

architecture underlying our deep learning-derived EDFs is correlated with standard ECG traits and CVDs.
”

4/ On a minor note, I did not see any QQ plots or reports of inflation for the GWAS. These could also be reported using LD-score regression (in this case to estimate SNP heritability and potential inflation).

RE: We thank the reviewer for this constructive suggestion.

We estimated the inflation factor (λ) for the 16 important EDFs. GWAS of the EDF achieved a mean λ of 1.027 (ranging from 1.000 to 1.047), indicating minimal genomic inflation. The QQ plots for all 16 EDFs were shown in Supplementary Figure S9, which visually confirm the absence of systematic inflation. Additionally, we estimated the liability-scale heritability (h^2) of the 16 EDFs using single-trait linkage disequilibrium score regression (LDSC). We first conducted quality control for SNPs inputting into LD-score regression (LDSC) analysis by: (1). filtering SNPs using the HapMap3; (2). removing the SNPs if they were strand-ambiguous; (3). removing SNPs whose minor allele frequency (MAF)<0.01; (4). removing SNPs located in the major histocompatibility complex (MHC) region (chromosome 6: 28,477,797–33,448,354). As shown in Supplementary Table S8, among 16 EDFs, 14 EDFs have heritability (h^2) with false discovery rate (*FDR*) < 0.05 and a median h^2 of 0.103 (ranging from 0.030 to 0.152), suggesting that EDFs are highly polygenic.

We have added above content to Methods and Results section of Supplementary Material:

Methods (page 4 of Supplementary Material)

*“**Genetic correlation estimation***

Linkage disequilibrium score regression (LDSC) is used to estimate genetic correlations between traits based on GWAS summary statistics. Firstly, we conducted quality control for SNPs input into LDSC analysis by: (1). filtering SNPs using the HapMap370; (2). removing the SNPs if they were strand-ambiguous; (3). removing SNPs whose minor allele frequency (MAF)<0.01; (4). removing SNPs located in the major histocompatibility complex (MHC) region (chromosome 6: 28,477,797–33,448,354). The European-ancestry population of 1000 Genomes Project Phase 3 was used as the reference panel for the analysis. Then, LDSC was performed, under a weighted linear model, by regressing the product of Z-statistics of two traits with the LD scores of the genome-wide SNPs.”

Results (page 9 of Supplementary Material)

*“**Estimating the inflation factor and liability-scale heritability for GWAS of ECG-LFM derived features***

We estimated the inflation factor (λ) for the 16 important EDFs. GWAS of the EDF achieved a mean λ of 1.027 (ranging from 1.000 to 1.047), indicating minimal genomic inflation. The QQ plots for all 16 EDFs were shown in Supplementary Figure S9, which visually confirm the absence of systematic inflation. Additionally, we estimated the liability-scale heritability (h^2) of the 16 EDFs using single-trait linkage disequilibrium score regression (LDSC). SNPs inputting into LDSC performed quality control by filtering SNPs using the HapMap3, removing the SNPs if they were strand-ambiguous, removing SNPs whose minor allele frequency (MAF)<0.01 and removing SNPs located in the major histocompatibility complex (MHC) region (chromosome 6: 28,477,797–33,448,354). As shown in Supplementary Table S8, among 16 EDFs, 14 EDFs have

heritability (h^2) with false discovery rate (FDR) < 0.05 and a median h^2 of 0.103 (ranging from 0.030 to 0.152), suggesting that EDFs are highly polygenic."

5/ Using only MAGMA to relate the EDFs to genes (rather than to genotypes) seems somewhat limited. If the authors intend to pursue this direction, they may consider methods that integrate multiple sources of evidence to link SNP associations to genes—MAGMA being one component but not the only relevant tool.

RE: Thank you for this valuable suggestion regarding gene-based association analysis. To strengthen our gene-level findings, we have further performed genome-wide gene association analysis using FUMA (PMID: 29184056) on the meta-GWAS result. Loci obtained by meta-GWAS were mapped to genes in FUMA using three strategies.
First, SNPs were mapped to genes based on physical distance (within a 10-kb window) from known protein-coding genes in the human reference assembly (GRCh37/hg19). Secondly, SNPs were also mapped to genes if they have shown significant eQTL association with the genes. The eQTL mapping was performed using two data repositories (GTEx v8 and eQTL catalogue), and was based on cis-eQTLs that can map SNPs to genes up to 1 megabase apart. We used a false discovery rate of 0.05 to define significant eQTL associations. Thirdly, SNPs were mapped to genes using the three-dimensional spatial interactions between SNPs and genes, which potentially involve long-range interactions. This procedure was performed using Hi-C data from the study by Schmitt et al (PMID: 27851967) in the FUMA platform. The candidate genes were selected if the SNP-gene interaction region overlaps with an enhancer region according to the data from Heart tissue in the Roadmap Epigenomics Project (PMID: 25693563), and is located in the promoter region (250 bp upstream to 500 bp downstream of the transcription start site) defined by Roadmap Epigenomics Project. The interactions are considered significant if the binomial test indicated a false discovery rate $< 1 \times 10^{-5}$ (PMID: 27851967) and modified to account for the differences in the cell lines used in this study.

FUMA analysis identified 62 EDF-associated genes. Specifically, positional mapping identified 23 genes within 10 kb of the lead SNPs, eQTL mapping revealed 21 genes with significant expression associations in heart-relevant tissues, and chromatin interaction mapping identified 42 genes through three-dimensional spatial interactions (Supplementary Table S15).
Additionally, we performed MAGMA analysis and revealed 9 genes (NFIA, TTN, SAP30L, HAND1, CDKN1A, CEP85L, PLN, KLHL38 and SIPA1L1) significantly (*P-value*<0.05/18762 $\approx 2.7 \times 10^{-6}$) associated with the EDFs (Figure 6C, Supplementary Table S12).
We have added the above content in Results and Methods sections.

Results (page 19 of manuscript)
"To identify genes associated with the EDFs, multi-marker analysis of GenoMic annotation (MAGMA) (PMID: 25885710) and FUMA (PMID: 29184056) on the meta-GWAS result were performed. In total, 63 unique genes were identified by at least one of the three FUMA gene mapping strategies or considered as significant by MAGMA. Briefly, MAGMA analysis revealed 9 genes (NFIA, TTN, SAP30L, HAND1, CDKN1A, CEP85L, PLN, KLHL38 and SIPA1L1) as significantly (P-value<0.05/18762 $\approx 2.7 \times 10^{-6}$) associated with the EDFs (Figure 6C, Supplementary Table S14), FUMA identified a total of 62 candidate genes (Supplementary Table S15). Out of these 63 genes, 39 are reported as associated with ECG traits, such as QRS

duration (PMID: 31217584), QT interval (PMID: 24952745), and PR interval (PMID: 32439900). Moreover, 28 of them were found by previous studies to be associated with CVDs in the GWAS Catalog (PMID: 30445434), such as myocardial infarction (PMID: 38072966), heart failure and (PMID: 31919418) atrial fibrillation (PMID: 29892015). The MAGMA analysis indicated SAP30L as the most significantly associated with the EDFs (P-value=2.8 $\times 10^{-13}$). SAP30L is known to play roles in the regulation of cardiac development and heart function (PMID: 22821512). These findings suggest that the genes identified in our analysis may be involved in the genetic architecture of ECG features and cardiovascular diseases."

Methods (page 37 of manuscript):

*"**Gene mapping using FUMA***

To strengthen our gene-level findings, we have further performed genome-wide gene association analysis using FUMA (PMID: 29184056) on the meta-GWAS result. Loci obtained by meta-GWAS were mapped to genes in FUMA using three strategies.

First, SNPs were mapped to genes based on physical distance (within a 10-kb window) from known protein-coding genes in the human reference assembly (GRCh37/hg19).

Secondly, SNPs were also mapped to genes if they have shown significant eQTL association with the genes. The eQTL mapping was performed using two data repositories (GTEx v8 and eQTL catalogue), and was based on cis-eQTLs that can map SNPs to genes up to 1 megabase apart. We used a false discovery rate of 0.05 to define significant eQTL associations.

Thirdly, SNPs were mapped to genes using the three-dimensional spatial interactions between SNPs and genes, which potentially involve long-range interactions. This procedure was performed using Hi-C data from the study by Schmitt et al (PMID: 27851967) in the FUMA platform. The candidate genes were selected if the SNP-gene interaction region overlaps with an enhancer region according to the data from Heart tissue in the Roadmap Epigenomics Project (PMID: 25693563), and is located in the promoter region (250 bp upstream to 500 bp downstream of the transcription start site) defined by Roadmap Epigenomics Project. The interactions are considered significant if the binomial test indicated a false discovery rate $< 1 \times 10^{-5}$ (PMID: 27851967) and modified to account for the differences in the cell lines used in this study."

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

The authors have adequately addressed the previous comments. I have no further comments on this work.

Reviewer #1 (Remarks on code availability):

The authors have made the code available and provided adequate documentation, including instructions for installation and usage. Based on the provided materials, the code is a usable resource for the community and should allow reproduction of the main results.

RE: We sincerely thank the reviewer for reviewing our revised manuscript and for the positive feedback. We appreciate your recognition that our revisions have adequately addressed the previous comments, as well as your positive assessment of our code availability and documentation.

Reviewer #2 (Remarks to the Author):

I have no further comments.

RE: We sincerely thank the reviewer for reviewing our revised manuscript and for the positive feedback.