Genes & Diseases (2023) 10, 55–57

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.keaipublishing.com/en/journals/genes-diseases

RAPID COMMUNICATION

Identification and validation of cardiac nonconserved human-specific enhancers

Congenital heart disease (CHD) is one of the most common causes of neonatal mortality. The worldwide morbidity of CHD is 9.410‰, while the proportion of CHD patients who are reported to carry mutations in coding regions is <50%.[1,2] Enhancers play an important role in the spatio-temporal expression of target genes, and a lot of related variations are associated with CHD. There are conserved and nonconserved enhancers. It is common to study conserved enhancers in model organisms, while nonconserved enhancers lack systematic investigation and functional validation. Unfortunately, lots of CHD patients have no diagnosis even after whole exome sequencing (WES) analysis, while the significance of nonconserved regulatory regions, one of the potential pathogenic factors, is usually overlooked. Therefore, in this study, to facilitate studies on the etiology of CHD, we present a systematic genome-wide analysis and functional validations of nonconserved human-specific enhancers (NoHEs).

Here, we investigated cardiac human-specific enhancers based on big data analyses. ChIP-seq peaks identified in mouse samples are not suitable for the study of human-specific enhancers since these nonconserved regions do not exist in the mouse genome. Therefore, we included and analyzed 20 human heart-related and enhancer-associated ChIP-seq datasets downloaded from Cistrome DB (http://cistrome.org/db/#/) (Fig. 1A; Fig. S1A). Enhancer-associated histones were H3K4me1, H3K4me3 and H3K27ac. The number of enhancer peaks from these datasets was >850,000, and the total length was approximately 112.2 Mb. Detailed information on ChIP-seq datasets is shown in Table S1, and detailed methods are described in the supplementary materials.

Given that different thresholds, peak widths or data normalization methods can call variable peaks, data integration is commonly performed by intersection analysis on peaks.[3] Therefore, we intersected the 112.2 Mb enhancer-associated ChIP-seq peaks with a 92.4 Mb human genome region that covered 328 heart-related genes (Table S2) with upstream and downstream 100 kb. The coverage of enhancer-associated peaks around these genes is shown in Figure S1B. Then, to strictly screen out the cardiac human-specific peaks, we mapped these cardiac enhancer peaks to the mouse genome (mm10) using LiftOver with the minimum ratio of bases that must remap set to 0.01.[4] As shown in Figure 1B, there were ~1.3 Mb H3K4me1-related, ~0.8 Mb H3K4me3-related and ~1.6 Mb H3K27ac-related nonconserved human-specific peaks around cardiac genes.

To further generate a cardiac NoHE compendium and clarify the proportion of NoHEs in the total cardiac enhancer sequences, we mapped the total cardiac enhancer peaks and nonconserved peaks to the human genome (hg38). After that, those hg38 peaks were merged to form unique new sequences (termed cardiac enhancer sequences and cardiac NoHEs, respectively). Detailed methods are described in the supplementary materials. Consequently, we identified >4500 cardiac NoHEs, covering ~1.6 Mb of the human genome. Remarkably, in the ~5.5 Mb cardiac enhancer sequences (File S1; Fig. S1C), the proportion of cardiac NoHEs was >29% (1.6/5.5 Mb) (Fig. 1C). Moreover, as shown in Figure 1D, this 1.6 Mb compendium (File S2) was composed of ~1.0 Mb H3K4me1-related, ~0.3 Mb H3K4me3-related and ~0.7 Mb H3K27ac-related cardiac NoHEs. These data accentuate the conclusion that a large number of human-specific candidate cardiac enhancers will be of vital in studies of human heart regulatory regions.

Based on the cardiac NoHE compendium, we chose 30 candidate fragments (Table S3) as described in the supplementary materials and validated their enhancer activities in the human cardiomyocyte cell line AC16. Luciferase reporter assay showed that 50% (15/30) were active enhancers (Fig. 1E). Notably, most of the related possible target genes, such as *NKX2-5*, *NOTCH1*, *GATA4*, *MEF2A* and *TBX20*, were key genes related to CHD (Table S4). Detailed information on related histones and distances to possible target genes is shown in Figure. S1D–F. These results confirmed the efficiency and practicability of this cardiac NoHE compendium in identifying active enhancers that had putative effects on heart development.

To further investigate the function of these active cardiac NoHEs, we validated the function of 6 related SNPs (Table S5).

Peer review under responsibility of Chongqing Medical University.

https://doi.org/10.1016/j.gendis.2022.03.005

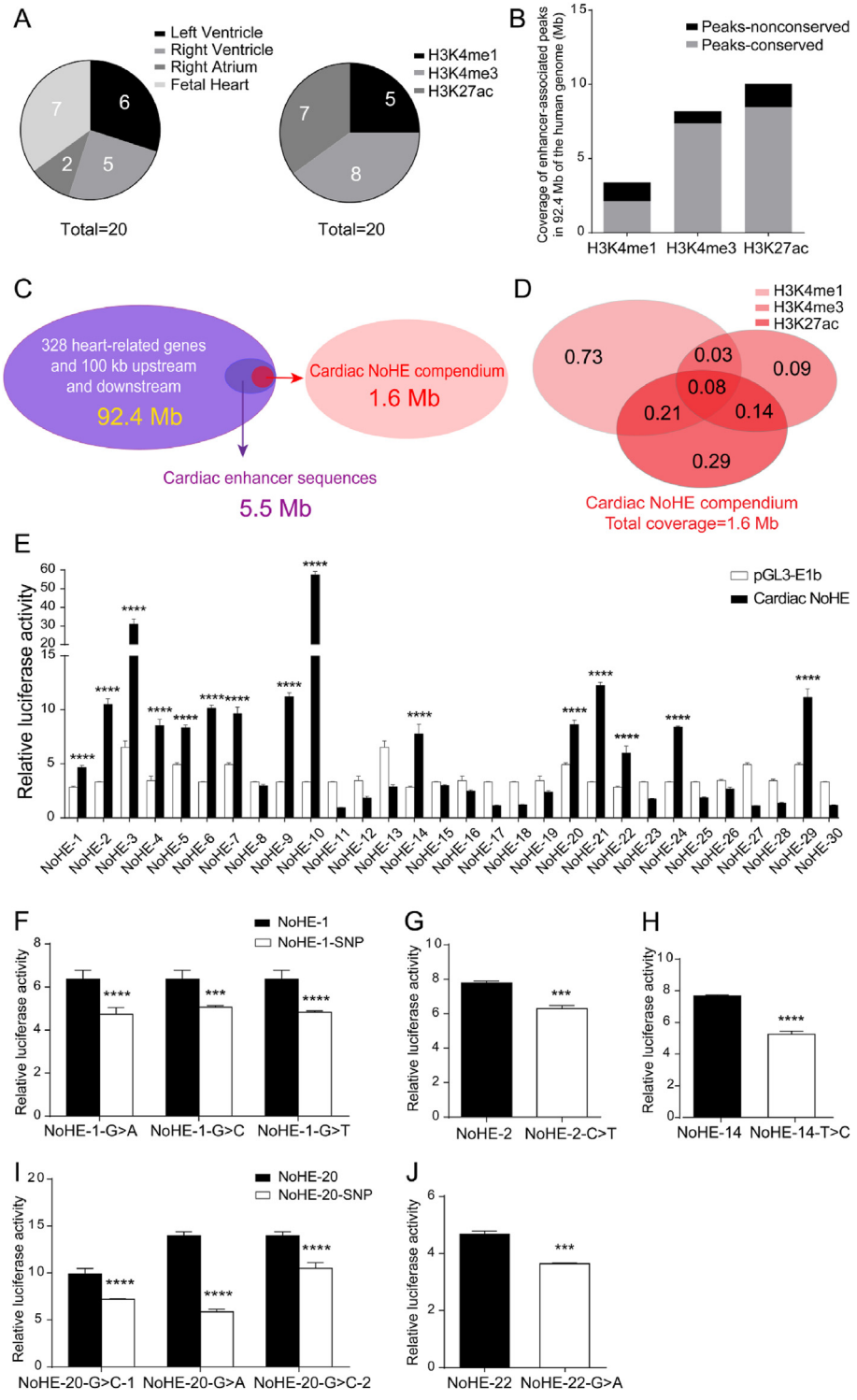

Figure 1 Identification and validation of cardiac nonconserved human-specific enhancers. **(A)** Information on included human heart-related and enhancer-associated ChIP-seq datasets. **(B)** Conserved and nonconserved enhancer-associated peaks in 92.4 Mb of the human genome. **(C, D)** Schematic diagram of the generation and constituent parts of the 1.6 Mb cardiac NoHE compendium. NoHE, nonconserved human-specific enhancer. **(E)** Validation of enhancer activity of the candidate cardiac NoHEs. Multiple *t* tests were performed for the statistical analysis between the NoHE groups and pGL3-E1b control groups. The assumption that all rows were sampled from populations with the same scatter (SD) was used to compute the individual *P* values. Correction for multiple comparisons was performed using the Sidak-Bonferroni method. ****$P < 0.0001$, $n = 3$. **(F–J)** Validation of the enhancer activity of cardiac NoHEs with related SNP-containing alleles. Multiple *t* tests were performed for the statistical analysis between the NoHE groups and NoHE-SNP groups. The assumption that all rows were sampled from populations with the same scatter (SD) was used to compute the individual *P* values. Correction for multiple comparisons was performed using the Sidak-Bonferroni method. Student's *t* test was used for the statistical analysis of NoHE with only one SNP-containing allele. ***$P < 0.001$; ****$P < 0.0001$, $n = 3$.

Compared with the WT groups (NoHE groups), the enhancer activities of 5 cardiac NoHEs with SNP-containing alleles were all significantly decreased in this study, which supported that rs78350132, rs113114656, rs12056541, rs139498, rs9611527 and rs3097165 represented functional variations (Fig. 1F—J). Remarkably, the mapped traits of rs78350132 and rs113114656 are systolic blood pressure and coronary artery disease. Considering that the correlation between the SNP and the corresponding trait will be more convincing if the functional verification is similar to the GWAS data,[5] these two heart-related intronic variations, which had an important effect on the human cardiomyocytes in this study, are very likely related to heart disease.

To date, many CHD patients have no diagnosis without the identification of the responsible exon locus.[2] This indicates that many pathogenic variations in regulatory regions, especially those in nonconserved regions, have not been well studied. Therefore, it is necessary to establish more integrative prediction methodologies and perform more experimental investigations to identify functional human-specific regulatory elements.[5] Our work focused on nonconserved regulatory regions and first identified 1.6 Mb human-specific cardiac enhancers by using rigorous criteria for the genome-wide analysis of ChIP-seq data. Notably, lots of enriched regions are usually not included in calling peaks due to the thresholds and/or peak widths set in ChIP-seq experiments,[3] which suggests that the cardiac NoHE compendium in this study is likely larger than 1.6 Mb. In addition, 50% of candidate cardiac NoHE and all related SNPs were functional. These results support the assertion that a large number of cardiac enhancers are nonconserved and functional. Therefore, in addition to analyzing conserved regions, capturing and studying human-specific functional elements should also be taken seriously.

Here, we presented a genome-wide analysis and functional validations of cardiac NoHEs that do not exist in the genome of a common model organism: the mouse. We first generated a 1.6 Mb cardiac NoHE compendium, which contained a large number of functional enhancers that exerted putative effects on heart development. This compendium will be valuable for cardiovascular research, especially studies on the etiology of CHD with variations in nonconserved regions.

Author contributions

Yonghao Gui and Qiang Li designed the study and reviewed the manuscript; Yawen Zhang performed the experiments and wrote the paper; Yawen Zhang, Yiting Gui, Xudong Chen, Feng Wang and Xu Wang collected and analyzed the data; Yiting Gui and Xudong Chen helped with the experiments; Yawen Zhang, Fang Wu, Youhua Wang and Xu Wang performed the statistical analyses. All authors have read and approved the final version of the manuscript.

Conflict of interests

The authors declare no conflict of interests.

Funding

This work was supported by The Natural Science Foundation of Shanghai (China) (No. 21ZR1410100); The National Natural Science Foundation of China (No. 81873481, 81771632, 81470442, 81741081 and 81170147); The National Key Research and Development Program of China (No. 2016YFC1000500); National Basic Research Program of China (973 Program) (No. 2013CB945401) and The Shanghai Key Laboratory of Birth Defects (China) (No. 13DZ2260600).

Acknowledgements

We are grateful to Jia Lin and Yinglan Zhang for their technical support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gendis.2022.03.005.

References

1. Liu Y, Chen S, Zühlke L, et al. Global birth prevalence of congenital heart defects 1970-2017: updated systematic review and meta-analysis of 260 studies. *Int J Epidemiol*. 2019;48(2):455—463.
2. Blue GM, Kirk EP, Giannoulatou E, et al. Advances in the genetics of congenital heart disease: a clinician's guide. *J Am Coll Cardiol*. 2017;69(7):859—870.
3. Hawkins RD, Hon GC, Ren B. Next-generation genomics: an integrative approach. *Nat Rev Genet*. 2010;11(7):476—486.
4. Song B, Sang Q, Wang H, et al. Complement genome annotation lift over using a weighted sequence alignment strategy. *Front Genet*. 2019;10:1046.
5. Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009;106(23):9362—9367.

Yawen Zhang[a,b], Yiting Gui[a,b], Xudong Chen[a], Feng Wang[a,b], Fang Wu[a,b], Youhua Wang[c], Xu Wang[d], Yonghao Gui[b,**], Qiang Li[a,*]

[a] *Translational Medical Center for Development and Disease, Shanghai Key Laboratory of Birth Defect, Institute of Pediatrics, Children's Hospital of Fudan University, National Children's Medical Center, Shanghai 201102, China*
[b] *Cardiovascular Center, NHC Key Laboratory of Neonatal Diseases (Fudan University), Children's Hospital of Fudan University, National Children's Medical Center, Shanghai 201102, China*
[c] *Department of Cardiology, Longhua Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai 200032, China*
[d] *Cancer Institute, Fudan University Shanghai Cancer Center, Shanghai 200032, China*

*Corresponding author.

**Corresponding author.
E-mail addresses: yhgui@fudan.edu.cn (Y. Gui), liq@fudan.edu.cn (Q. Li)

7 September 2021
Available online 30 March 2022