

Supplementary Notes

Supplementary Note 1: PCS documentation for low-signal signed iterative random forest analysis

An HTML (webpage style) document that can be downloaded and viewed in the browser or via the website <https://yu-group.github.io/epistasis-cardiac-hypertrophy/>. Guided by the PCS framework for veridical data science, we transparently document and justify the many human judgment calls and modeling decisions (e.g., choice of dimension reduction and binarization thresholds) that were made throughout the lo-siRF pipeline. Furthermore, we provide additional exploration and stability analyses to ensure that our findings are stable and robust across these choices.

Supplementary Note 2: *CCDC141-IGF1R* and *CCDC141-TTN* interactions non-additively reshape cardiomyocyte size distribution

Given the diverse expressivity of cellular hypertrophy among diseased cells (indicated by the heavily positively-skewed size distribution in Fig. 6d), we further analyzed size changes of large and small cells focused into different microchannels separately. Our results show that silencing *TTN* ($p = 0.001$) or *CCDC141-TTN* ($p = 0.005$) predominantly targets cells that exhibit cellular hypertrophy (gray bars in Extended Data Fig. 8a, right) but has a trivial impact on small cells.

To explore this diverse expressivity further, we evaluated how gene silencing reshapes the cell size distribution (Extended Data Fig. 7a and p -values in 7b). QQ-plots showcasing the quantiles of gene-silenced cells against their scrambled controls suggest that digenic gene silencing leads to a stronger effect on larger cells (indicated by the larger deviation of points from the solid black line in Extended Data Fig. 7c). For example, the upper quantiles of diseased cells silencing *CCDC141-IGF1R* are much lower (the 0.9 quantile decreased by 14.4%, $p < 1E-4$) than the corresponding scrambled quantiles, indicating that this variant favors rescuing hypertrophic cells over small cells. However, this

size-dependent effect is mitigated in cells with monogenic silencing of *CCDC141* or *IGF1R*. This discrepancy in how these two genes affect cell size reinforces the hypothesized non-additive interaction between *CCDC141* and *IGF1R* (e.g., $p < 1E-4$ for non-additivity, 0.9 quantile). Importantly, stable non-additive interaction effects for hypertrophic cells (quantile levels higher than 0.6) are also observed for the *CCDC141-TTN* pair in both cell lines (Extended Data 8). Overall, these experimental results suggest both an epistatic interaction between *CCDC141* and the other two genes (*IGF1R* and *TTN*) as well as independent effects on modifying cardiomyocyte hypertrophy.

Supplementary Note 3: Legends for Extended Data Spreadsheets

Extended Data 1. GWAS results for left ventricular mass.

- A summary of the left ventricular mass (LVM) and left ventricular mass index (LVMI) genome-wide association study results using PLINK and BOLT-LMM. This includes a list of the LVM/LVMI risk loci, lead SNVs, independent significant SNVs, and positionally-mapped genes from FUMA GWAS.

Extended Data 2. List of GWAS-filtered SNVs used in the lo-siRF pipeline.

- A list of the 1405 SNVs that passed the genome-wide association study dimension reduction step in the lo-siRF pipeline. This table includes the rsID, chromosome, hg19 position, allelic information, its corresponding genomic loci in lo-siRF, and inferred (exonic) function.

Extended Data 3. The lo-siRF-prioritized SNVs and SNV-SNV epistatic interactions.

- A summary of the 283 lo-siRF-prioritized SNVs that show stable association with left ventricular mass index (LVMI), which are aggregated into six left ventricular hypertrophy risk loci (as shown in Fig. 2f) according to ANNOVAR gene annotation in lo-siRF. The spreadsheet also includes separate tabs showing the prioritized SNV-SNV pairs for each of the three locus-to-locus interactions (*CCDC141-IGF1R*, *CCDC141-TTN*, and *CCDC141-LOC157273;TNKS*) stably identified by all the three LVMI binarization thresholds in lo-siRF.

Extended Data 4. Functional annotations for lo-siRF-prioritized SNVs and their LD-linked SNVs extracted by FUMA, along with the functionally mapped genes.

- The spreadsheet includes a list of the 283 lo-siRF-prioritized SNVs and the FUMA-extracted candidate SNVs in LD with any of the 283 lo-siRF-prioritized SNVs, as well as their corresponding functional annotations from ANNOVAR, ChromHMM core-15 chromatin state, CADD score, RegulomeD score, eQTL and sQTL information from GTEx, and GWAS statistics. The spreadsheet also includes a list of protein-coding genes functionally linked from the 283 lo-siRF-prioritized SNVs and their LD-linked SNVs by positional, eQTL, and chromatin interaction mapping in FUMA.

Extended Data 5. ANNOVAR Enrichment analysis results.

- ANNOVAR enrichment of functional consequences of all SNVs (both the lo-siRF-prioritized SNVs and their LD-linked SNVs extracted by FUMA, details in Extended Data 4) associated to each of the six lo-siRF-prioritized risk loci for left ventricular hypertrophy.

Extended Data 6. Gene ontology and pathway co-association network.

- Co-association network between genes that are functionally linked to the lo-siRF-prioritized epistatic and hypostatic loci (Extended Data 4) and their shared GO or pathway terms. GOs and pathways are enriched from multiple annotated gene set libraries from Enrichr (Fig. 4c).

Extended Data 7. Transcription factor co-association network.

- The spreadsheet includes two lists of transcription factors and RNA-binding regulators enriched from lo-siRF-prioritized and lo-siRF-deprioritized genes (Fig. 4d), respectively. It also includes a summary of co-associations between lo-siRF-prioritized genes and their shared regulatory factors (Fig. 4e). Transcription factors and RNA-binding regulators in the co-association network are enriched from nine distinct annotated gene set libraries from Enrichr and ChEA3.

Extended Data 8. Single cardiomyocyte morphology assessment and statistical analysis.

- A summary of the epistatic relationships in cardiomyocyte morphology confirmed by gene perturbation experiment (Fig. 6). The spreadsheet includes evaluation of epistatic effects on cell size (at various quantile levels), cell textural irregularity, and cell boundary irregularity, supported by multiple statistical analysis and assessment of non-additivity for the investigated gene-gene interactions.