Rules of the Rod: *MYH7* Variant Effects Across the β-MHC Coiled-Coil Domain

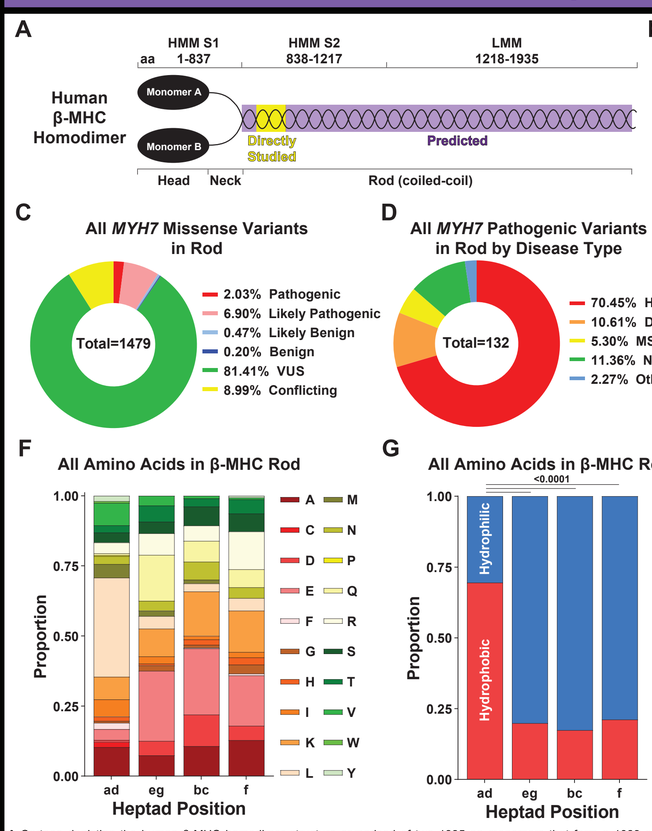
Clayton E. Friedman[1,2,3], Shawn Fayer[4], Sriram Pendyala[4], Wei-Ming Chien[1,2,3,5], Charles E. Murry[1,2,3,6,7,8], Lea M. Starita[4,9], Douglas M. Fowler[4,7,9], Kai-Chun Yang[1,2,3,5]

[1] Institute for Stem Cell and Regenerative Medicine, University of Washington, School of Medicine, Seattle, WA 98109, USA; [2] Center for Cardiovascular Biology, University of Washington, Seattle, WA 98109, USA; [3] Department of Medicine/Cardiology, University of Washington, Seattle, WA 98109, USA; [4] Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA; [5] Cardiology/Hospital Specialty Medicine, VA Puget Sound HCS, Seattle, WA 98108, USA; [6] Department of Pathology, University of Washington, Seattle, WA 98109, USA; [7] Department of Bioengineering, University of Washington, Seattle, WA 98195, USA; [8] Department of Stem Cell Biology and Regenerative Medicine, University of Southern California, Los Angeles, 90033, USA; [9] Brotman Baty Institute for Precision Medicine, Seattle, WA 98195, USA.

BACKGROUND

Hypertrophic cardiomyopathy (HCM) is characterized as an unexplained thickening of the cardiac left ventricle[1]. HCM patients with pathogenic variants in genes encoding sarcomeric proteins have worse prognoses and are at greater risk of sudden cardiac death and developing heart failure[2]. Pathogenic, autosomal-dominant variants in *myosin heavy chain 7* (*MYH7*, encoding β-MHC) account for ~33% of all HCM cases. Genetic testing of HCM patients has identified >1700 *MYH7* missense variants (ClinVar[3]), however, ~81% are classified as variants of uncertain significance (VUS) due to insufficient functional/clinical data required to interpret variant effect. In particular, while the *MYH7* S2 domain is sensitive to missense mutations, the leading *in silico* variant effect prediction algorithm (EVE[4]) fails to accurately predict known pathogenic variants in the proximal part of this domain, indicating the need for further functional analysis using *in vitro* models. While multiplexed assays of variant effect (MAVEs) can resolve variant effect at scale, this approach relies on examining salient phenotypes in easily mutagenized cells. Development of a method to gene edit hiPSCs at scale would enable MAVEs of genes in disease-relavant contexts by leveraging the differentiation capacity of hiPSCs.

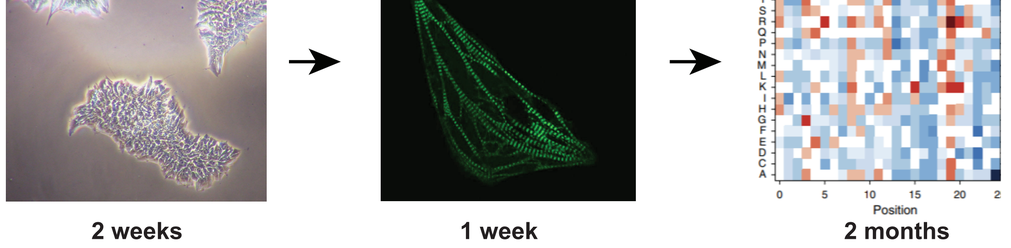
MYH7 missense variants and β-MHC coiled-coil rules



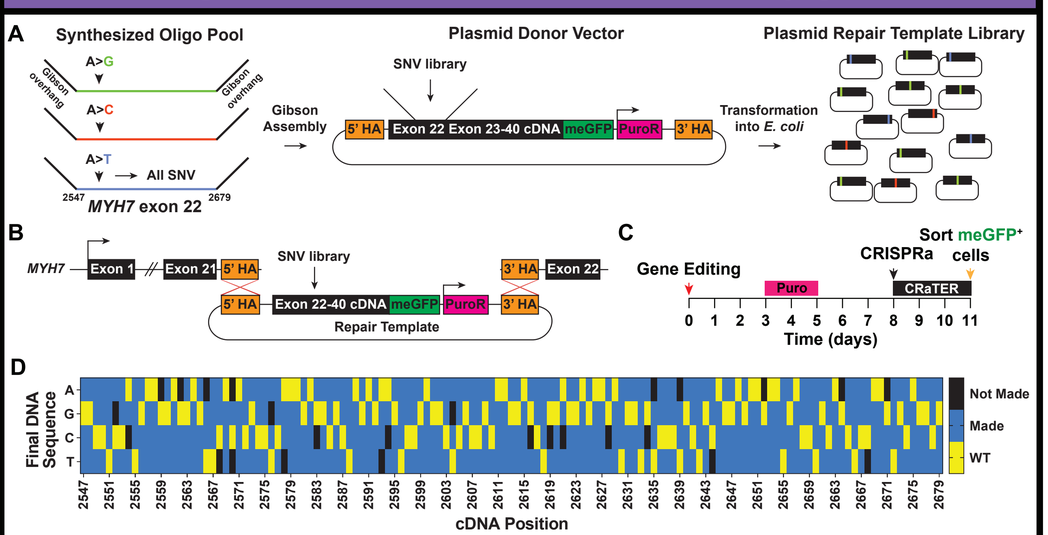
A Cartoon depicting the human β-MHC homodimer structure comprised of two 1935 aa monomers that form a 1089 aa α-helical coiled-coil domain spanning the heavy meromyosin (HMM) subdomain 2 (S2) through the light meromyosin (LMM) domain. Yellow=aa 849-893 region directly studied; purple=aa 838-848 and 894-1930 regions where variant effects are predicted.
B Cartoon cross section of a β-MHC α-helical coiled-coil domain depicting a repeating heptad motif 'abcdefg' with distances for register pairs from the interhelical interface (dotted vertical line) from nearest to furthest: 'a' and 'd', 'e' and 'g', 'b' and 'c', and 'f'.
C All *MYH7* missense variants in the β-MHC rod (aa 838-1935) from ClinVar (Accessed 11 March 2025). VUS=variant of uncertain significance.
D Disease types reported with pathogenic *MYH7* missense variants.
E Histogram of pathogenic *MYH7* missense variant position in 25 aa increments.
F The proportion of amino acids in the β-MHC rod for each heptad position paired by interhelical interface distance from smallest to largest (left to right).
G The proportion of hydrophobic (red; A, V, L, I, F, W, Y, and M) and hydrophilic (blue; C, D, E, G, H, K, N, P, Q, R, S, and T) amino acids in the β-MHC rod for each heptad position pair. P values determined by Fisher's exact test.
H Pathogenic *MYH7* variants at each heptad position pair. The blue section highlights substitutions at the 'a' and 'd' heptad positions that switch from hydrophobic to hydrophilic.

OBJECTIVE and STRATEGY

Objective: ***Develop a high-throughput pipeline to generate and functionally assess MYH7 SNVs in hiPSC-CMs to classify VUS and build a functionally-informed predictor of coiled-coil variant effect.***



MYH7 single nucleotide variant hiPSC library generation



A Cloning strategy used to generate a *MYH7* SNV library in plasmid DNA.
B Recombination of the *MYH7* SNV plasmid library into the endogenous *MYH7* locus.
C Schematic of the gene-editing strategy to knock-in a *MYH7* SNV library into hiPSCs and enrich for on-target edited hiPSCs using CRaTER[5].
D Heatmap of *MYH7* SNV generated in hiPSCs with the original cDNA sequence (horizontal) and final sequences (vertical). Yellow=WT; blue=made; black=not made.

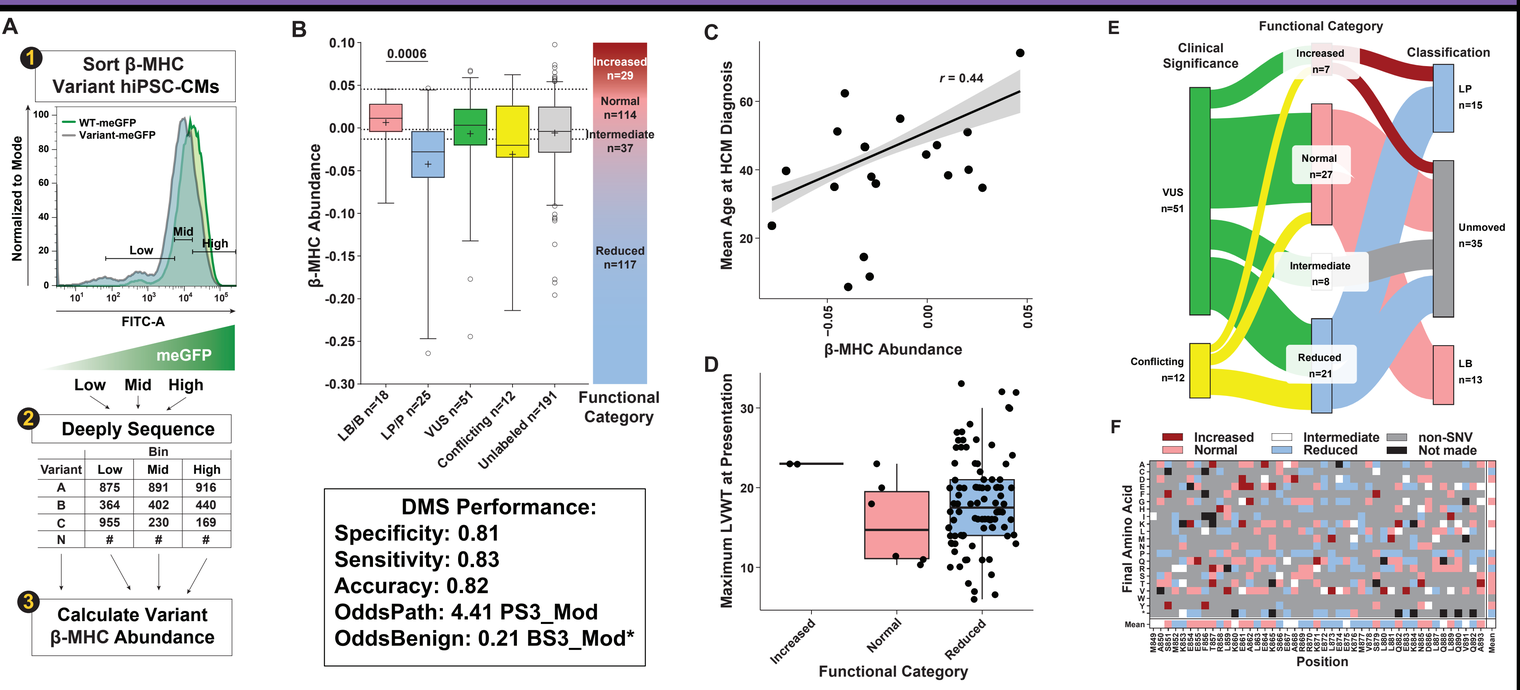
REFERENCES

1. Yogasundaram H, Alhumaid W, Dzwiniel T, Christian S, Oudit GY. Cardiomyopathies and Genetic Testing in Heart Failure: Role in Defining Phenotype-Targeted Approaches and Management. Can J Cardiol. 2021 Apr;37(4):547-559.
2. Marstrand P, Han L, Day SM, Olivotto I, Ashley EA, Michels M, Pereira AC, Wittekind SG, Helms A, Saberi S, Jacoby D, Ware JS, Colan SD, Semsarian C, Ingles J, Lakdawala NK, Ho CY; SHaRe Investigators. Hypertrophic Cardiomyopathy With Left Ventricular Systolic Dysfunction: Insights From the SHaRe Registry. Circulation. 2020 Apr 28;141(17):1371-1383.
3. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. 2016;44:D862-868.
4. Frazer J, Notin P, Dias M, Gomez A, Min JK, Brock K, Gal Y, Marks DS. Disease variant prediction with deep generative models of evolutionary data. Nature. 2021 Nov;599(7883):91-95.
5. Friedman CE, Fayer S, Pendyala S, Chien WM, Loiben A, Tran L, Chao LS, Mckinstry A, Ahmed D, Karbassi E, et al. CRaTER enrichment for on-target gene-editing enables generation of variant libraries in hiPSCs. J Mol Cell Cardiol. 2023.
6. Matreyek KA, Starita LM, Stephany JJ, Martin B, Chiasson MA, Gray VE, Kircher M, Khechaduri A, Dines JN, Hause RJ, et al. Multiplex assessment of protein variant abundance by massively parallel sequencing. Nat Genet. 2018.
7. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12:e1001779.
8. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020.
9. Harvey ML, Osborn E, Lancaster MC, Smith JE, Shen Q, Abe T, Cai R, Muhammad A, Kannankeril R, Yang T, et al. Mapping the Functional Landscape of KCNQ1 Variants to Define Ion Channel Mechanisms and Arrhythmia Risk. medRxiv. 2025.
10. Mirza H, Kumar A, Craiglow BG, Zhou J, Saraceni C, Torbeck R, Ragsdale B, Rehder P, Ranki A, Choate KA. Mutations Affecting Keratin 10 Surface-Exposed Residues Highlight the Structural Basis of Phenotypic Variation in Epidermolytic Ichthyosis. J Invest Dermatol. 2015;135:3041-3050.

SUPPORT and ACKNOWLEDGMENTS

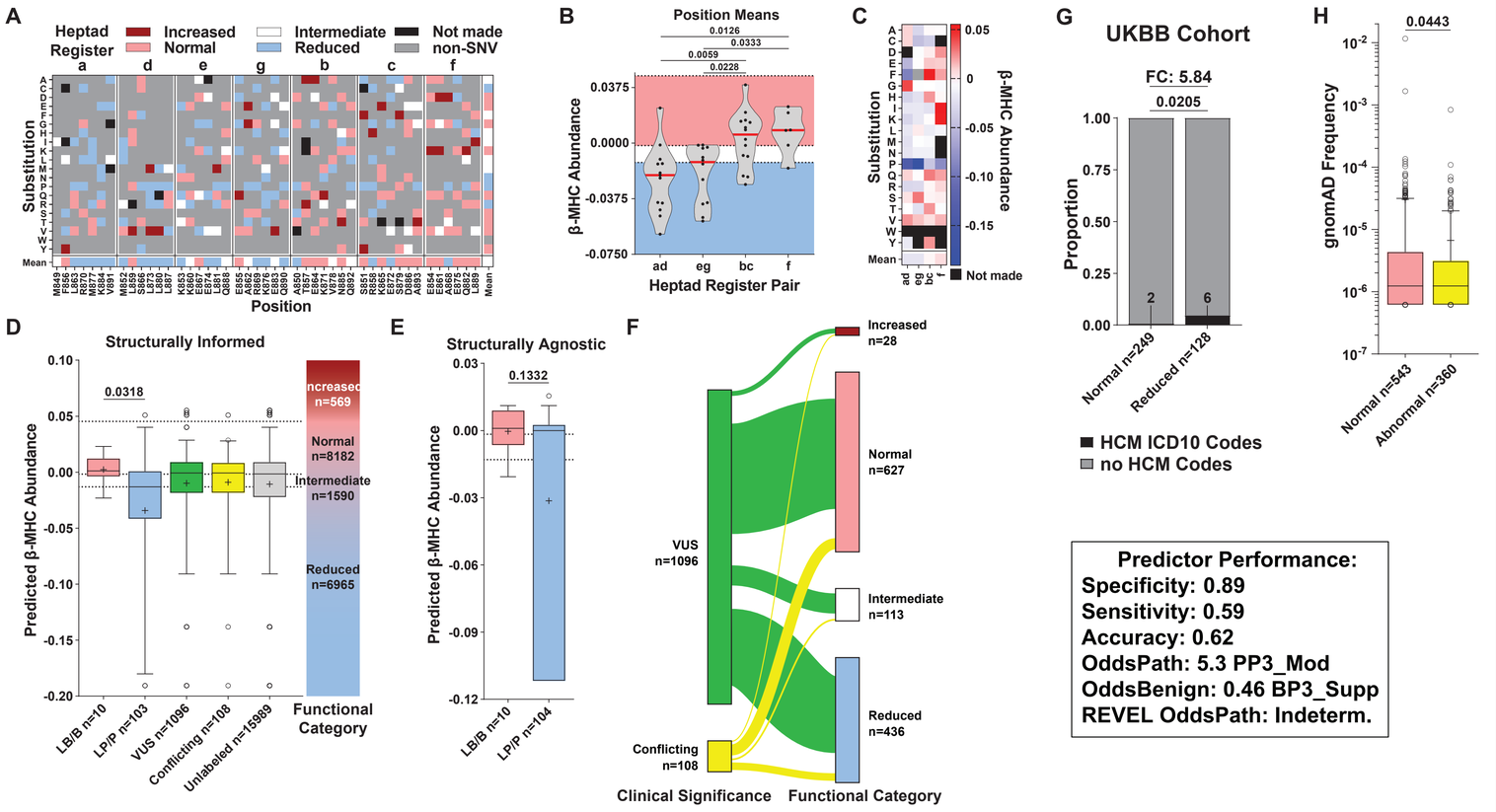
I thank the Yang and Murry laboratories for their support and thoughtful discussions. I also thank Lea Starita and Doug Fowler for their ongoing support and collaboration. This work was supported by the CMAP Postdoctoral Training Fund and external grants (NHLBI K99 Award K99HL177347, VA IK2BX004642, and 1R01HL171174-01A1).

Multiplexed assessment of β-MHC abundance in *MYH7* variant hiPSC-CMs



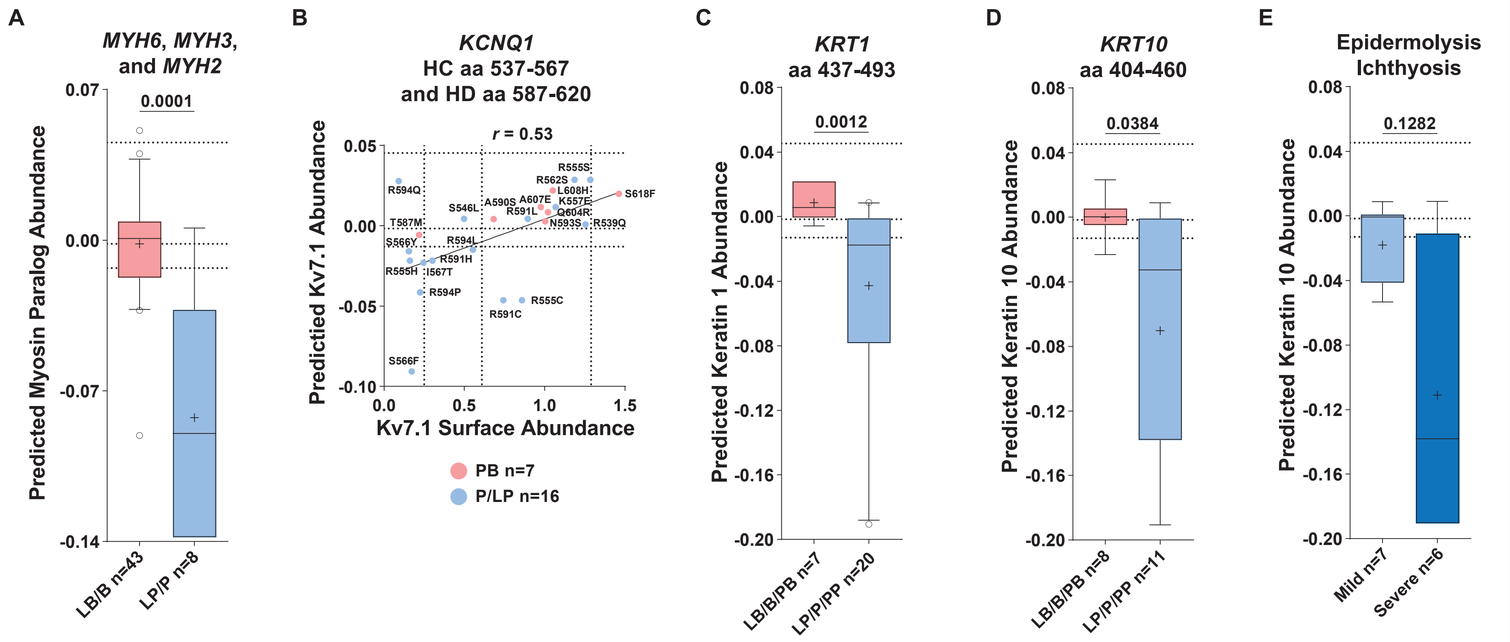
A Schematic of multiplexed β-MHC abundance assessment pipeline using VAMP-seq[6]: (1) Representative flow cytometry plot displaying FITC-A (normalized to mode) for T60 WT-MYH7-meGFP (grey) versus variant-MYH7-meGFP library (green) hiPSC-CMs with library cell sorting into low, mid, and high FITC-A bins, (2) deep sequencing of the mutagenized endogenous *MYH7* locus for each bin, and (3) calculation of β-MHC protein abundance for each variant. T=days since onset of cardiac-directed differentiation.
B Amino acid-level median β-MHC abundance (clear dots, n=297 unique variants) normalized to WT (n=2-5 biological replicates). ClinVar LB/B (pink), LP/P (blue), VUS (green), conflicting (yellow), and unlabeled variants (grey) are displayed in box-and-whisker plots. Lines in boxes=column medians; +=column mean; upper dotted line=increased β-MHC upper cutoff (0.045357); middle dotted line=normal/intermediate β-MHC cutoff (-0.001653605); lower dotted line=reduced β-MHC cutoff (-0.013).
C Linear regression for *MYH7* variants (black dots) displaying measured β-MHC abundance (x-axis) and mean age at HCM diagnosis (years) in SHaRe[2] patients (y-axis). Line=line of best fit; dotted lines=95% confidence intervals.
D Echocardiography measurements of cardiac left ventricular wall thickness (LVWT; mm) in HCM patients in the SHaRe registry for single *MYH7* variants (black dots) associated with increased, normal, or reduced β-MHC abundance. P values determined by Mann-Whitney U test.
E Sankey plot displaying the proportion of VUS and conflicting variants functionally categorized as increased, normal, intermediate, or reduced β-MHC abundance and classified as LP, LB, or unmoved.
F Heatmap of categorical β-MHC abundance for the original aa residue and position (horizontal) and the final amino acid (vertical). Grey=non-SNV; black=not made; final row=column means; final column=row means.

A functionally-informed predictor accurately segregates calibrating *MYH7* variants



A Heatmap of directly studied categorical β-MHC abundance for the original aa residue and position (horizontal) and the substituted amino acid (vertical), reorganized by register with interhelical interface distance increasing from left to right.
B Average β-MHC abundance for each aa position (black dots) at each register pair. Red lines=column medians. P values determined by one-way ANOVA with Tukey's multiple comparisons test.
C Heatmap of directly studied average β-MHC abundance for each register pair (horizontal) and the substituted amino acid (vertical).
D Predicted β-MHC abundance for all possible missense variants (clear dots) within the aa 838-848 and aa 894-1930 regions (n=17306 variants) extrapolating directly measured β-MHC abundance (aa 849-893) averaged according to register pairs (from **C**). ClinVar variants LB/B (pink), LP/P (blue), VUS (green), conflicting (yellow), and unlabeled variants (grey) are in box-and-whisker plots.
E Predicted β-MHC abundance for missense variants (clear dots) within the aa 838-848 and aa 894-1930 regions extrapolating directly measured β-MHC abundance (aa 849-893) averaged across all register pairs. ClinVar LB/B (pink) and LP/P (blue) variants are displayed in box-and-whisker plots.
F Sankey plot displaying the proportions of VUS and conflicting variants functionally categorized with increased, normal, intermediate, or reduced predicted β-MHC abundance.
G Proportion of UK Biobank[7] participants with HCM-related ICD10 codes (black) or without such codes (grey) for all *MYH7* variants predicted to have normal or reduced abundance β-MHC abundance. P values determined by Fisher's exact test.
H GnomAD[8] allelic frequency in the general population of *MYH7* variants (clear dots) predicted to have normal (pink) or abnormal β-MHC abundance (yellow) displayed in box-and-whisker plots.

Predictions are generalizable to myosin paralogs and unrelated coiled-coil domains



A Amino acid-level predicted myosin paralog abundances (clear dots) combining missense variants for *MYH6*, *MYH3*, and *MYH2*. ClinVar variants LB/B (pink) and LP/P (blue) are displayed in box-and-whisker plots.
B Linear regression for pseudobenign (PB; pink) and ClinVar P/LP (blue) *KCNQ1* variants (encoding Kv7.1) displaying measured Kv7.1 surface abundance[9] (x-axis) and predicted Kv7.1 abundance (y-axis). Solid line=line of best fit; left dotted line=loss of function/partial loss of function cutoff (0.25); middle dotted line=partial loss of function/normal function cutoff (0.61); right dotted line= normal function/gain of function cutoff (1.29).
Predicted Keratin 1 (**C**) and Keratin 10 (**D**) abundances (clear dots) for missense variants in *KRT1* and *KRT10*, respectively. ClinVar LB/B with PB variants (pink) and LP/P with PP variants (blue) are displayed in box-and-whisker plots.
E Amino acid-level predicted Keratin 10 abundance (clear dots) for missense variants in *KRT10*. Variants associated with mild epidermolysis ichthyosis[10] (EI) (light blue) and severe EI (dark blue) are displayed in box-and-whisker plots.

CONCLUSIONS and FUTURE DIRECTIONS

- Pooled β-MHC abundance assessment enables phenotyping of nearly 300 *MYH7* SNVs in hiPSC-CMs
- β-MHC abundance accurately segregates pathogenic and benign *MYH7* variants and correlates with patient heart wall thickness and age of HCM onset
- β-MHC abundance correlates with the interhelical interface distance and is substitution specific, informing accurate variant effect predicitons of >17000 *MYH7* variants across 1045 positions
- β-MHC-based, functionally informed predictions are accurate in unrelated coiled-coil domains, indicating variants in repetititive structural motifs can have repetitive and predictable functional effects
- What other multiplexable phenotypes can help to resolve unmoved *MYH7* VUS?

Predictor Performance:
Specificity: 0.89
Sensitivity: 0.59
Accuracy: 0.62
OddsPath: 5.3 PP3_Mod
OddsBenign: 0.46 BP3_Supp
REVEL OddsPath: Indeterm.

Seeking Tenure-Track Assistant Professor Position

Email: fried8@uw.edu