A community-oriented, data-driven resource to improve protocol design for cardiac modelling from human pluripotent stem cells

Chris Siu Yeung Chow[1], Sumedha Negi[1], Shaine Chenxin Bao[1], Chen Fang[1,2], James E. Hudson[3], Woo Jun Shim[1], Yuanzhao Cao[1,4], Nathan J. Palpant[1,4,*]

[1]*Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia*

[2]*Ochsner Clinical School-The University of Queensland School of Medicine, New Orleans, Louisiana*

[3]*Queensland Institute for Medical Research, Brisbane, QLD, Australia*

[4]*Co-senior authors*

[*]Corresponding author

Associate Professor Nathan J. Palpant
Institute for Molecular Bioscience, The University of Queensland
306 Carmody Road, St. Lucia, QLD 4072 Australia
Email: n.palpant@uq.edu.au

Abstract

Protocols for culturing cardiac cells derived from human-induced pluripotent stem cells (hiPSCs) are critical for studying mechanisms of development and disease. This study provides analysis of a community-oriented database to assess parameters used in published cardiac differentiation protocols, including a web-accessible portal for user-guided customisation of protocols. Validation studies support context-dependent roles for media composition in deriving functional maturation of hiPSC-derived cardiomyocytes.

After decades optimising methods for cardiac differentiation from hiPSCs, there is need for data-driven protocol design that draws on this collective knowledge. To address this, Ewoldt et al. recently provided a systematic analysis of 300 studies focusing on 63 protocol annotations used for differentiation and maturation of hiPSC-derived cardiomyocytes (hiPSC-CMs)[1]. The study found that protocols are often exclusive to a few research groups and lack appropriate comparisons to benchmark performance. Furthermore, there is insufficient data to systematically identify protocol parameters that yield the best functional outcomes. In this study, we develop a feature selection approach that provides unsupervised, data-driven recommendations for protocol optimisation to achieve physiological maturation of cardiac cells using diverse experimental variables.

We first supplemented the foundational database with 14 additional studies using metabolic techniques for cardiac maturation and disease modelling. Using hPSCReg[2], we also added categories for cell line sex and ancestry aiming to understand their importance on hiPSC modelling studies and to promote awareness of equity and representation in hiPSC cardiac modelling studies[3]. To enhance protocol interpretability and comparability, we considered the feature selection strategy as a categorisation problem which includes biological annotations, addresses feature cardinality, and improves the practicality of future data submission and dissemination. We then performed feature importance analyses to evaluate protocol variables associated with 15 hiPSC-CM physiological maturation parameters. Since study designs cannot be imputed, we assumed informative missingness and applied one-hot encoding for all categorised features. Collectively, these changes expanded the original database to 314 published protocols and 441 features (**Supplemental Table 1**).

We determined feature enrichment with both Shannon Entropy (information gain via the reduction of entropy by splitting the dataset with the feature) and Gini Impurity (the likelihood that the target functional outcome, is incorrectly miscategorised in a subset after splitting with the protocol category) to capture maximum protocol-outcome associations given the limited data. For each physiological parameter, we fit default random forest models in the Sklearn package[4] with all studies have a reported value. We constructed 2000 stochastic datasets to form null distributions for both metrics by 15 functional outcomes as a one-tailed bootstrapping approach. To evaluate any unintended bias in protocol sampling that would dilute biologically relevant feature enrichment, we performed a Jaccard Index analysis and show that the Boolean features in the new database are highly unique (**Extended Data Fig. 1, Supplemental Table 2**). A summary of the results is provided in **Fig. 1a** showing the enrichments for experimental variables as they inform physiological hiPSC-CM maturation parameters. Unexpectedly, the sample sizes of published protocols did not correlate linearly with their number of significant enrichments (**Fig. 1b**). However, the number of significant enrichments strongly correlates with the fold-change dissimilarity for significant versus all enriched protocol features (**Extended Data Fig. 2**), showing that the bootstrapping approach to apply a statistical threshold leads to independent enrichments, retaining unique signals across the limited pool of 314 protocols without introducing ascertainment bias. Subsequent hierarchical clustering by statistical significance revealed distinct protocol parameters associated with measures of physiological maturation (**Fig. 1c** and **Supplemental Table 2, 3)**.

Taken together, the outcomes of enriched parameters align with well-established mechanisms of hiPSC-CM maturation including the use of non-cardiomyocyte cell types in 3D organoid models, the type, induction, and duration of Wnt signalling activity, and the use of diverse stimulation methods and culturing substrates (**Fig. 1d** and **Extended Data Fig. 3**). We next evaluated RNA-seq data from 11 protocols in the database to assess maturation benchmarking to the adult human heart based on myofilament isoform switching of myosin light chain and troponin I[5] (**Extended Data Fig. 4**). As expected, these data reinforce a significant need for improved *in vitro* maturation methods of both 2D and 3D protocols to achieve molecular and functional maturation of the adult human heart.

Noting the limited sampling of the database and the ongoing development of new methods, we designed an online portal (available upon final publication) to assist with customisation of protocol design, efficient retrieval of protocols, and uploading of new protocol results for community-oriented knowledge gain. New published protocols can be uploaded and benchmarked to the database using this portal. Standards for human stem cell studies developed by the International Society for Stem Cell Research[6] will be required and monitored to ensure data quality in the database.

Next, we demonstrate application of the database in guiding hiPSC-CM maturation and disease modeling. Recent efforts have aimed at improving hiPSC protocols to model ischaemic sensitivity for cardiovascular drug testing[8] to

address the high disease burden of ischaemic heart disease (IHD)[7]. Using a small molecule Wnt modulation protocol for hiPSC differentiation into cardiomyocytes[9], we evaluated the physiological relationship between hiPSC-CM maturation and ischaemia **(Fig. 2a)**. scDRS analysis[10] of single cell RNA-seq data across cardiac differentiation[9] demonstrates that the base protocol has limited enrichment for gene associations in pathological disorders like myocardial infarction and IHD compared to general cardiovascular structural/functional traits (**Fig. 2b**). We therefore used the database to define the most common backbone media, growth matrix, metabolic substrates, and signaling factors for protocols used to model myocardial ischemia (**Fig. 2c**). We experimentally compared the top two backbone media compositions: DMEM+FA versus RPMI+B27 **(Fig. 2d)**, noting that DMEM and lipid mix supplementation were also enriched for contractility in protocol mining (**Fig. 1d)**. Impedance-based analysis of contractile force confirmed that DMEM+FA significantly increases contraction amplitude compared to RPMI+B27 **(Fig. 2e)**, which is likely due to the increased calcium present in DMEM media[11]. RNA-seq analysis provided further evidence that DMEM+FA upregulates well-established genes involved in ischaemic responses and fatty acid beta-oxidation, and downregulates genes involved in glycolysis **(Fig. 2f** and **Extended Data Fig. 5).** Using SBayesRC[12] we show that DMEM+FA induced gene programs are exclusively aligned with the heritability enrichment of both heart structural/functional traits as well as chronic IHD (**Fig. 2g**).

While media supplementation with metabolic maturation-promoting factors is common (**Fig. 2c**), we investigated the impact of backbone media on modulating the physiological effects of these factors (**Supplemental Tables 4 and 5**). We show that RPMI+B27 coupled with growth factors (T3) and metabolic pathway modulators (GW) markedly enhanced contractility (**Fig. 2h)**. Also, all maturation supplements increase post-IR cell death compared to RPMI+B27 alone (**Fig. 2i**). However, compared to RPMI+B27, standalone DMEM+FA media markedly increased hiPSC-CMs contractility and none of the maturation factors provided additional value (**Fig. 2h** and **Extended Data Fig. 6**) whereas their addition mildly improved ischaemia sensitivity compared to DMEM+FA media alone (**Fig. 2i** and **Extended Data Fig. 6**). Collectively, this confirms our feature selection results indicating that DMEM and Lipid Mix are the most enriched composition impacting hiPSC-CM force of contraction (**Fig 1d**), and likely replace the need for growth hormone and supplements required in RPMI+B27.

In conclusion, this study provides significant value for studies aiming for data-driven protocol design. It overcomes the near-incomparability of context-dependent hiPSC-CM protocols to reveal design-outcome associations. Users can estimate the direction of effect by assessing the proportion of quantised functional outcomes and retrieving similar protocol designs through the web portal. Our processed data can be used for benchmarking, support experimental testing, and streamline literature review to inform decisions made for protocol design. We acknowledge the following considerations in using this database. First, although we have shown the dissimilarity of enrichments, they are still based on most reproduced study designs. Emerging modifications that impact protocol improvements may not be identified as significant in this analysis simply because of limitations in representation. Second, it remains challenging to generalise the biological role of protocol variables which may be enriched to counteract other protocol variables that inhibit maturation. Third, tree models paired with one-hot encoding do not retain ordinal relationships between one-hot-encoded categories, therefore the original study characteristic can be used in the same path. Lastly, contradictions in enrichment for maturation and modeling may arise because the database overlooks critical differences between assay designs, such as the cellular complexity and nutrient requirements of 2D versus 3D culture models. Collectively, we anticipate that community-driven development of this database will accelerate the utility of hiPSC cardiac models for research and translational applications.

Acknowledgements

This work has been supported by grant funding from the NHMRC (MRFCDDM000033 to NP and 2007625 to NP), the Ian Potter Foundation (31111380 to NP), and the National Heart Foundation of Australia (106721 to NP). We thank Meredith Redd for her assistance with project design.

Disclosures

No conflicts of interest are reported for any of the authors.

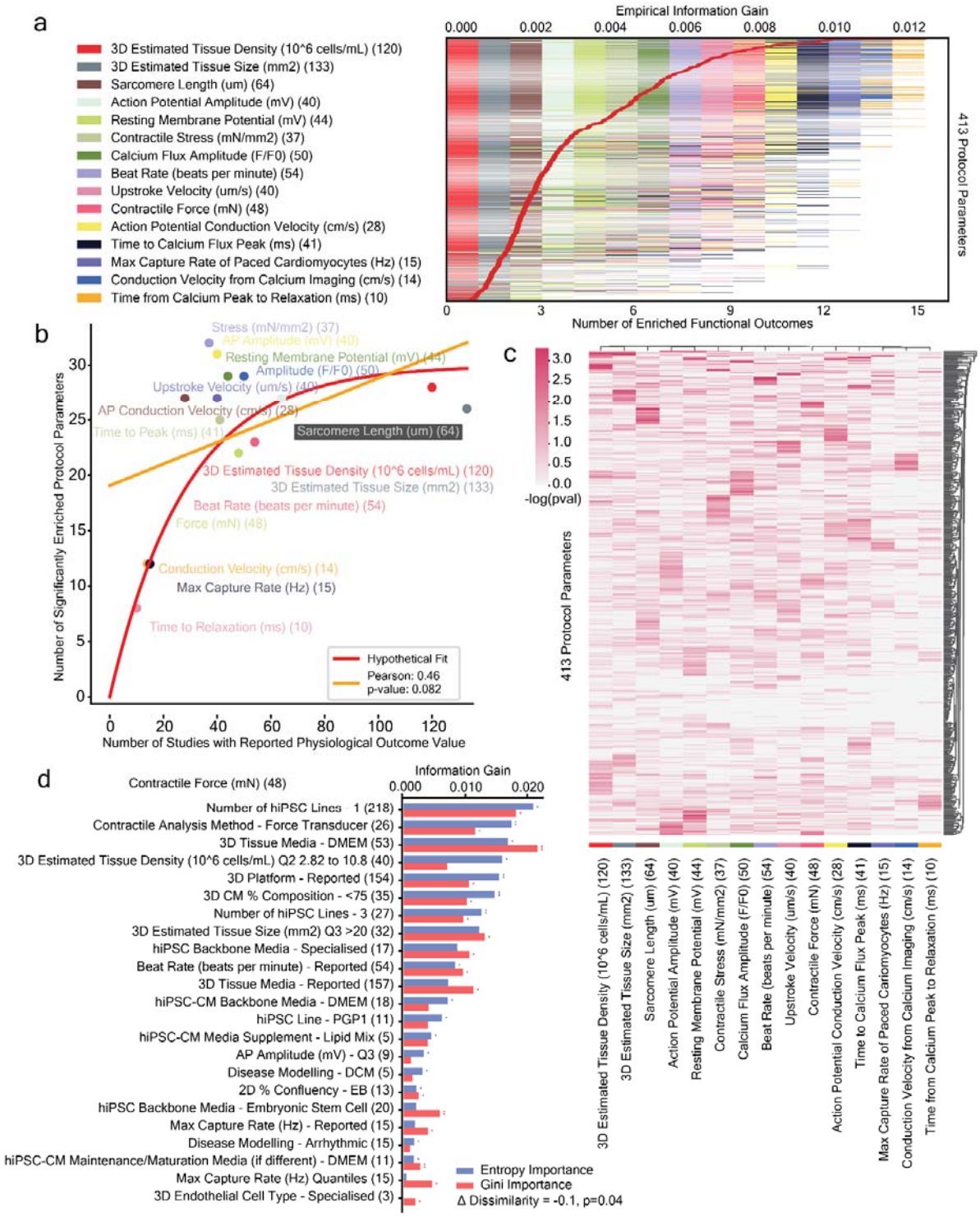

Figure 1. Data mining analysis of published hiPSC-CM protocols. (a) Overview illustrating the enrichments of 413 protocol features across 15 affected biological outcomes. Empirical information gain is measured as the average between the correlated Gini and Entropy importances (Pearson's R = 0.94, p=0.00) or the standalone enriched value if only one metric is enriched. **(b)** Relationship between the number of significantly enriched features and the size of training set. Hypothetical model (red) estimates a plateau of approximately 30 out of 413 hiPSC-CM protocol parameters that affect physiological metrics. **(c)** Clustered heatmap by statistical significance shows different physiological outcomes have distinct protocol parameter enrichments. **(d)** Bar chart of significantly enriched protocol annotations for contractile force. In brackets are the number of studies matching the feature description. We measured p-value as the probability that the 2000 stochastic datasets have higher enrichment value for the protocol feature. *P<0.05, **P<0.01, ***P<0.001, ***P<0.0001.

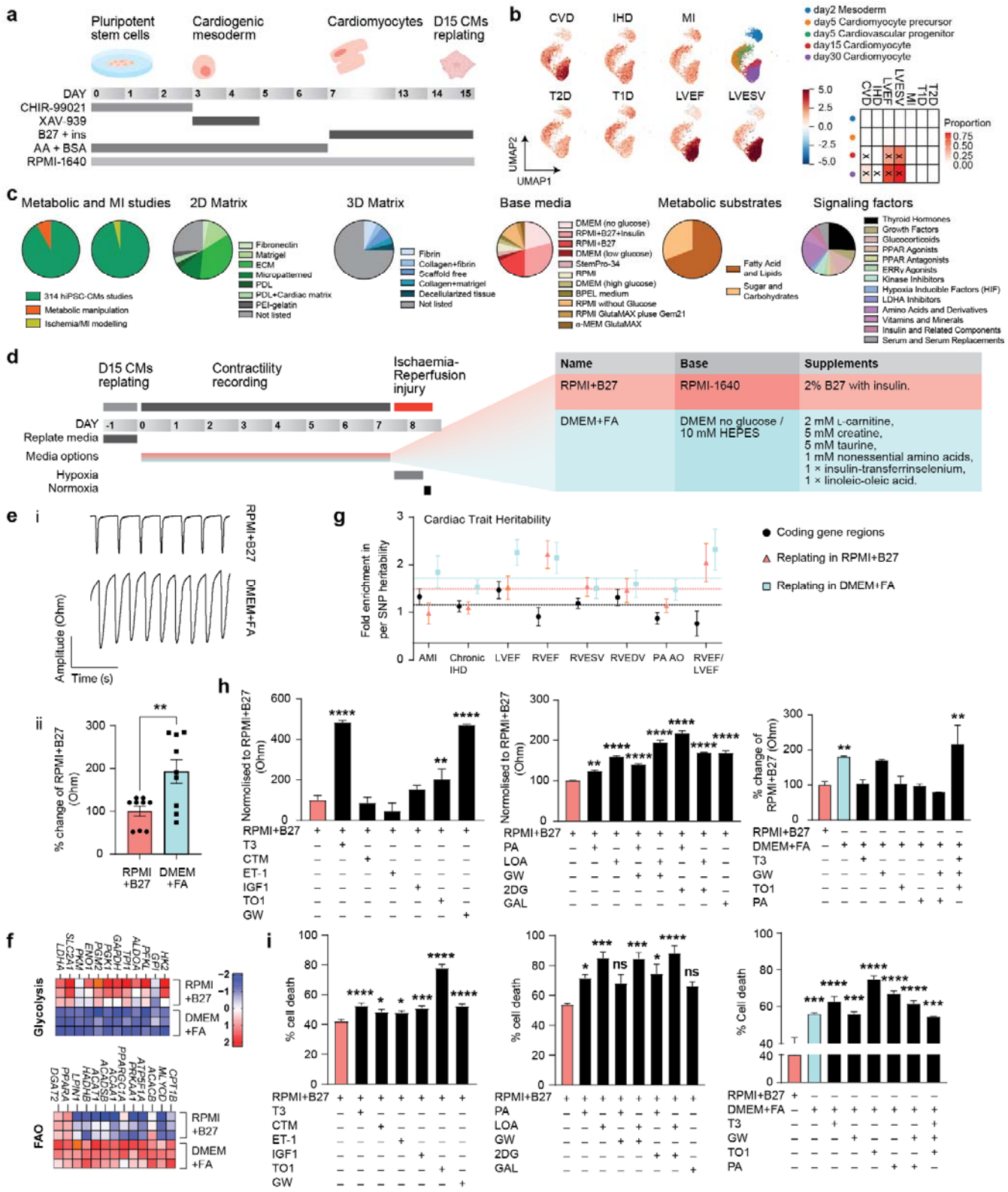

Figure 2. Comparative analysis of cardiometabolic factors influencing iPSC-CM maturation and ischaemia modelling. (a) Schematic of the applied hiPSC-CM protocol. **(b)** UMAPs shows the relevance of molecular signatures associating individual hiPSC-CMs to traits and diseases based on scDRS. Heatmap shows an increasing proportion of associated cells along differentiation timepoints, where x marks the group-level significance. **(c)** Pie charts show amongst 314 hiPSC-CM studies, the proportion that uses metabolic strategy and ischaemic modelling and those

studies' common 2D/3D matrix materials, backbone media, metabolic substrates, and signalling factors. **(d)** Protocol schematic with backbone media as RPMI+B27 or DMEM+FA to assess maturity and ischaemic sensitivity of hiPSC-CMs. **(e)** Comparisons of hiPSC-CM contractility showing representative traces (i), and normalised peak values (ii). **(f)** RNA-seq analysis of known gene panels shows metabolic reprogramming between DMEM+FA and RPMI+B27 cultured hiPSC-CMs. **(g)** Fold enrichment in per-SNP heritability for up-regulated genomic regions after DMEM+FA and RPMI+B27 conditions versus all protein-coding genomic regions with SBayesRC, with standard error bars. The dotted lines are mean enrichments across shown traits for each group. **(h)** Changes in peak contractile amplitude of hiPSC-CMs and **(i)** their proportion of cell death under ischaemia-reperfusion stress in backbone media RPMI+B27 or DMEM+FA with maturation-relevant signalling molecules and metabolic substrates. CVD, cardiovascular disease; T1D, T2D, type-1 or 2 diabetes; AMI, MI, acute or myocardial infarction; IHD, ischemic heart disease; LVEF, LVESV, left ventricular ejection fraction or end-systolic volume; RVEF, right ventricular ejection fraction; RVESV, right ventricular end-systolic volume; RVEDV, right ventricular end-diastolic volume; PA AO, pulmonary artery and aorta diameter ratio; RVEF/LVEF, inverse-normalised right ventricular ejection fraction to left ventricular ejection fraction ratio. Unless otherwise specified, 3 biological replicates and each with 1-3 technical replicates; Data are presented as mean ± standard error with student's *t*-test (e) and One-Way ANOVA (h, i). *P<0.05, **P<0.01, ***P<0.001, ***P<0.0001.

References

1 Ewoldt, J. D., Samuel; Jewett, Maggie; Karakan, M. Çağatay; Lin, Yih-Mei; Mir Hashemian, Paria; Gao, Xining; Lou, Linhua; McLellan, Micheal; Tabares, Jonathan; Ma, Marshall; Coariti, Adriana C. Salazar; He, Jin; Toussaint, Jr., Kimani C.; Bifano, Thomas; Ramaswany, Sharan; White, Alice; Agarwal, Arvind; Lejeune, Emma; Baker, Brendon; Chen, Christopher. Induced pluripotent stem cell-derived cardiomyocyte in vitro models: tissue fabrication protocols, assessment methods, and quantitative maturation metrics for benchmarking progress. *DRYAD* (2024). https://doi.org//10.5061/dryad.ksn02v7bh

2 Seltmann, S. *et al.* hPSCreg--the human pluripotent stem cell registry. *Nucleic Acids Res* **44**, D757-763 (2016). https://doi.org/10.1093/nar/gkv963

3 Ghosh, S., Nehme, R. & Barrett, L. E. Greater genetic diversity is needed in human pluripotent stem cell models. *Nat Commun* **13**, 7301 (2022). https://doi.org/10.1038/s41467-022-34940-z

4 Pedregosa, F. a. V., G. and Gramfort, A. and Michel, V., and Thirion, B. a. G., O. and Blondel, M. and Prettenhofer, P., and Weiss, R. a. D., V. and Vanderplas, J. and Passos, A. and & Cournapeau, D. a. B., M. and Perrot, M. and Duchesnay, E. Scikit-learn: Machine Learning in {P}ython. *Journal of Machine Learning Research* **12**, 2825--2830 (2011).

5 Yang, X., Pabon, L. & Murry, C. E. Engineering adolescence: maturation of human pluripotent stem cell-derived cardiomyocytes. *Circ Res* **114**, 511-523 (2014). https://doi.org/10.1161/CIRCRESAHA.114.300558

6 Ludwig, T. E. *et al.* ISSCR standards for the use of human stem cells in basic research. *Stem Cell Reports* **18**, 1744-1752 (2023). https://doi.org/10.1016/j.stemcr.2023.08.003

7 Roth, G. A. *et al.* Global, regional, and national burden of cardiovascular Ddseases for 10 causes, 1990 to 2015. *Journal of the American College of Cardiology* **70**, 1–25 (2017). https://doi.org/10.1016/j.jacc.2017.04.052

8 Redd, M. A. *et al.* Blockade of acid-sensing ion channel 1a is equipotent to sodium-hydrogen exchange inhibition in reducing cardiac injury in rodent models of myocardial infarction. *European Heart Journal*, in press (2023).

9 Friedman, C. E. *et al.* Single-Cell Transcriptomic Analysis of Cardiac Differentiation from Human PSCs Reveals HOPX-Dependent Cardiomyocyte Maturation. *Cell Stem Cell* **23**, 586-598.e588 (2018). https://doi.org/10.1016/j.stem.2018.09.009

10 Zhang, M. J. *et al.* Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. *Nat Genet* **54**, 1572-1580 (2022). https://doi.org/10.1038/s41588-022-01167-z

11 Tiburcy, M. *et al.* Defined Engineered Human Myocardium With Advanced Maturation for Applications in Heart Failure Modeling and Repair. *Circulation* **135**, 1832-1847 (2017). https://doi.org/10.1161/CIRCULATIONAHA.116.024145

12 Zheng, Z. *et al.* Leveraging functional genomic annotations and genome coverage to improve polygenic prediction of complex traits within and between ancestries. *Nature Genetics*, 1-11 (2024).