CMPortal: a community-oriented, data-driven resource to inform protocol design for cardiac modelling from human pluripotent stem cells

Chris Siu Yeung Chow[1], Sophie Shen[1], Amy Hanna[1], Sumedha Negi[1], Clarissa L. L. Tan[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

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

Protocol design and benchmarking is central to optimising model development using human pluripotent stem cell derived cardiomyocytes (hPSC-CMs). By applying data mining to decades of research and hundreds of peer reviewed studies, we evaluate how protocol variables associate with common properties of cardiac functional and physiological maturation. This resource is publicly accessible through CMPortal, a community-oriented website that provides data-driven tools for researchers to navigate leverage decades of knowledge for benchmarking protocol designs and outcomes for their dedicated applications in developmental biology, disease modelling, and drug screening.

Human pluripotent stem cell-derived cardiomyocytes (hPSC-CMs) are a widely used model for cardiac research, enabling studies of development, physiology, disease mechanisms, and drug responses.[1] However, maturing hPSC-CMs toward adult-like phenotypes remains an ongoing challenge for modeling adult cardiac biology.[2] Despite development of maturation strategies involving specialized media, co-culture systems, and various types of matrix and biomaterials, optimizing protocol conditions suitable for modelling adult heart phenotypes remain challenging due to high variability in experimental design and outcomes.

Ewoldt et al.[3,4] recently reviewed 300 study designs and demonstrate that while maturity outcomes have improved over time, no significant differences exist between maturation strategies. They highlight difficulties in parsing individual effects of protocol variables due to complex design variables and inconsistent reporting, drawing attention to the need for data-driven approaches and community-oriented benchmarking strategies to improve the accuracy and efficiency of hPSC-CM protocol design.

Here, we develop a data-driven platform that offers user-guided analyses of protocol features and their association with functional outcomes. Rather than assuming direct correlations exist between protocol factors and physiological maturation, we use an unsupervised approach that enables detection of bidirectionality and context-dependency allowing parameters to be independently assessed.

We first expanded the Ewoldt database[3] to 322 studies, including 14 metabolic maturation studies and 8 publications to capture recent developments in the field. We also added protocol design features including cell line sex, ancestry, and transcriptomic markers of myofilament isoform switching. The final database has standardized nomenclature across 400 protocol features spanning five categories: experimental design variables (e.g., media composition factors, plating densities, matrix), analysis methods (e.g., imaging techniques, electrophysiology approaches), cell profiles (details about cell line including sex, ancestry, and source), study characteristics (e.g., publication year, journal), and measured outcomes (e.g., measured sarcomere length, contractile force, calcium handling, myofilament isoforms) (**Figure 1A**).

We applied a hybrid strategy combining conventional statistics and data mining to assess protocol features against 117 target parameters, such as 18 maturity indicators, cell profiles, and applications in disease and pharmacological modeling (**Tables S1 and S2**). Numerical measurements were binned into quantiles (e.g. Q1, Q2, etc.) to enable analysis of protocols ranked by their maturity outcomes. One-hot encoding was used to derive a data format suitable for machine learning, which also accounts for the significant missing data and assumes protocol variables, when not used or reported, are still potentially causal.

For data mining, we used random forest classifiers with 10,000 permuted feature sets to associate protocol designs with target parameters through Shannon entropy and Gini impurity, allowing the detection of statistically robust associations even in small and sparsely reported datasets (**Figure 1A** and **Table S3**). Dimensionality reduction via UMAP demonstrated high protocol heterogeneity across studies, consistent with reported findings[3] (**Figure 1B**). Despite this heterogeneity, polynomial regression confirmed that the database sufficiently captures protocol features for target endpoints relating to best and worst performing quantiles in a data-limited setting (**Figure 1C**). Quality checks using Jaccard index analysis confirmed that data transformation preserved feature distinctiveness (**Figure S1**).

Data mining revealed alignment between protocol variables and expected biological impacts on cardiomyocyte maturation (**Figures 1D-E**). For example, Q1 contractility protocols were significantly associated with Wnt modulation, fibroblast inclusion, and matrix stiffness which are well established factors impacting sarcomere assembly and force generation.[5,6] Importantly, we identified that identical maturation strategies could be linked to both best and worst maturation outcomes, demonstrating that this approach identifies context-dependent effects of protocol variables and not simply assume their generalized effects (**Figure 1F**). Specifically, the findings suggest that maturation parameters can be independently modulated by distinct sets of protocol variables, challenging the assumption that improving one cardiac maturity metric necessarily enhances all others.

We next used a spearman correlation analysis of all continuous variables in the database (n=31) to assess whether maturity measurements recapitulate expected physiological relationships (**Table S4**). Indeed, the data demonstrates that across hundreds of peer-reviewed studies, there are significant positive correlations between expected parameters including resting membrane potential and conduction velocity ($\rho=1$, $p<0.01$), as well as contractile force and 3D tissue size (ρ=0.33, $p<0.05$) (**Figure 1G**). Interestingly, the database also identifies protocol correlations that provide important considerations in co-culture design. Different cell type compositions revealed positive, null, and negative effects on maturity indicators including calcium kinetics and sarcomere length (**Figure 1H**).

We integrated these findings into a community-oriented web resource, *CMPortal* (https://palpantlab.com/cmportal). The Database Viewer enables searching of the 322 curated protocols based on five feature categories using techniques from simple keyword matching to multi-criteria searches, with options to toggle column visibility, export customized datasets, and evaluate relevant protocols of interest (**Figures 2A-C**). The Variable Search enables researchers to identify studies and protocols based on topic areas and/or sets of protocol variables which enables unsupervised discovery of relevant protocols based on desired and/or associated protocol features, even when the published protocols do not report the selected metrics (**Figure 2D**). The Protocol Benchmarking tool enables researchers to evaluate differentiation and maturation protocols against database standards. By uploading or manually selecting protocol features and the user's own experimental measurements, the tool compares protocols through a radar chart that highlights maturity indicators across multiple categories and visually distinguishes between predicted, normal-range, and outlier experimental values (**Figures 2E-F**). Lastly, the Enrichment Browser allows researchers to identify protocol features significantly associated with specific outcomes which serves as the data foundation for the search and benchmarking features. (**Figure 2G**).

We aimed to validate an unexpected finding from the database showing that contractility and sarcomere length are not correlated maturation endpoints (**Figure 2H**), each having distinct protocol variable enrichments governing their maturation (**Table S2**). This finding suggests that the Frank-Starling effect, one of the most fundamental principles governing the relationship between heart pump function and load, develops through independent mechanisms. We investigated this by first identifying enriched protocol characteristics of Q1 and Q5 contractility, choosing backbone media compositions involving DMEM+fatty acids (FA) (Q1 contractility) and RPMI+B27 (Q5 contractility) (**Figure 2I**). DMEM+FA confers enhanced contractile maturity due to the high calcium concentration[7] and fatty acids[5] that mimic postnatal development. As shown in data mining, impedance-based measurements showed significantly higher peak contraction amplitude in DMEM+FA versus RPMI+B27 (**Figure 2J**), which was also supported by RNA-seq data showing higher metabolic activity from glycolysis to mature fatty acid oxidation in DMEM+FA cells (**Figure 2K**). Consistent with the database correlation, contractility changes occurred without significant variation in sarcomere length (**Figure 2L**). This supports our observation that fundamental features of cardiac maturation, including interdependent mechanisms controlling heart function like the Frank-Starling relationship, are not necessarily mechanistically linked in development. *CMPortal* allows for unsupervised discovery of these context-dependent protocol variables, enabling users to independently optimize protocols for specific functions.

CMPortal represents the first quantitative approach to overcome significant analytical difficulties of unreported, sparse protocol data, successfully associating context-dependent protocol variables with cell modelling outcomes. These findings challenge a fundamental but rarely challenged assumption that hPSC-CM maturity advances uniformly across all metrics. Instead, the results support a modularised paradigm to develop protocols through selected metrics based on experimental goals, aligning with recent literature demonstrating that attributes of hPSC-CM maturation can be compartmentalized.[8] We translate these findings into user-friendly tools provided on a public website for researchers to evaluate, design, and benchmark protocols. *CMPortal* provides a streamlined approach to literature review, inform experimental design, and facilitate protocol benchmarking for understanding cardiac biology, diseases, and drug responses as a community-oriented resource for hPSC-CM research.

Acknowledgements

This work has been supported by grant funding from the NHMRC (MRFCDDM000033 and 2007625 to NP), the Ian Potter Foundation (31111380 to NP), and the National Heart Foundation of Australia (106721 to NP). J.E.H. acknowledges the support of a Snow Medical Research Foundation Fellowship, Grant No. SMRF2019–060. Microscopy was performed at the Institute for Molecular Bioscience Microscopy Facility which was established with the support of the Australian Cancer Research Foundation (ACRF). We thank Meredith Redd for her assistance with project design.

Declaration of interests

N.J.P is cofounder and advisor for Infensa Bioscience, a biotechnology company focused on the development of peptide therapeutics for stroke and heart attack. J.E.H. is a cofounder, scientific advisor, and holds equity in Dynomics, a biotechnology company focused on the development of heart failure therapeutics. J.E.H. is a co-inventor on patents for bioengineered human cardiac tissue patches. J.E.H. is co-inventor on licensed patents for engineered heart muscle, and patents relating to cardiac organoid maturation and therapeutics.

Figure 1. *CMPortal* database construction and hybrid data mining approach.

(A) Schematic overview of the workflow for integrating and analysing 322 hPSC-CM studies to evaluate protocol designs and target parameters.

(B) UMAP visualizations project the database by five feature categories, enabling assessment of variability in protocol components and identification of shared trends, with annotations shown in color.

(C) Relationship between the number of significant enrichments and the number of training protocols. A polynomial model fitted to the best and worst quantiles of maturity indicators (pink and purple) suggests a plateau at 20–30 out of 413 protocol features significantly associated with maturity metrics. Grey indicates target parameters for intermediate maturity quantiles, cell profile, and study characteristics.

(D–E) Bar plots showing protocol feature metadata for the highest (D) and lowest (E) contraction force quantiles, based on enrichments with $p<0.025$. Yellow bars indicate the number of studies reporting the relevant variables; purple bars represent the confidence in true feature enrichment. Log-odds ratio (logOR) and Fisher's exact test (FET) estimate overall directionality of enrichment. Co-reporting bias is calculated as the percentage of enriched features co-published more often than expected by chance, using the mean Jaccard Index (JI) distribution for a matched number of features.

(F) Stacked bar plot compares maturation strategies revealing significant context-dependent effects on maturation properties.

(G) Table summarizing representative Spearman correlations among numerical protocol features, showing expected associations (full data in Table S4).

(H) Spearman correlation plots of cardiac cell ratios in 3D cultures show opposing relationships with sarcomere length and calcium relaxation time as maturity indicators.

Figure 2. *CMPortal* website interface and experimental use cases.

(A) Database Viewer allows users to explore 322 hPSC-CM protocols across 483 protocol features, and to install the filtered database.

(B) Pie charts of the database show media and substrate usage of metabolic maturation protocols.

(C) Example UMAPs show segregation of studies by the protocol feature categories. Cell line ancestry shows distinct trends of protocol variable and analysis method usage. Common media, like most features, show no clear trends.

(D) Variable Search enables protocols retrieval by any desired and/or significantly enriched features by target parameters.

(E) Protocol Benchmarking module enables users to upload and compare protocols against the database using experimental data and reference ranges estimated with the feature enrichments.

(F) Radar plot generated by *CMPortal* for comparing maturation outcomes between Protocol 80 and 250 which are amongst the most cited studies in the database. The plot shows protocol 80 being associated to properties of higher transcriptional and contractile maturation, but lower electrophysiological and calcium handling maturation. Hallow datapoints are estimated with enrichments when no experimental data is available.

(G) Enrichment Browser retrieves enrichments across the 117 target parameters.

(H) Spearman correlation of pairwise, co-reported values of contractility and sarcomere length in 322 protocols. No significance was found.

(I) *CMPortal*-informed experiment designed to compare the impact of media composition on contractility.

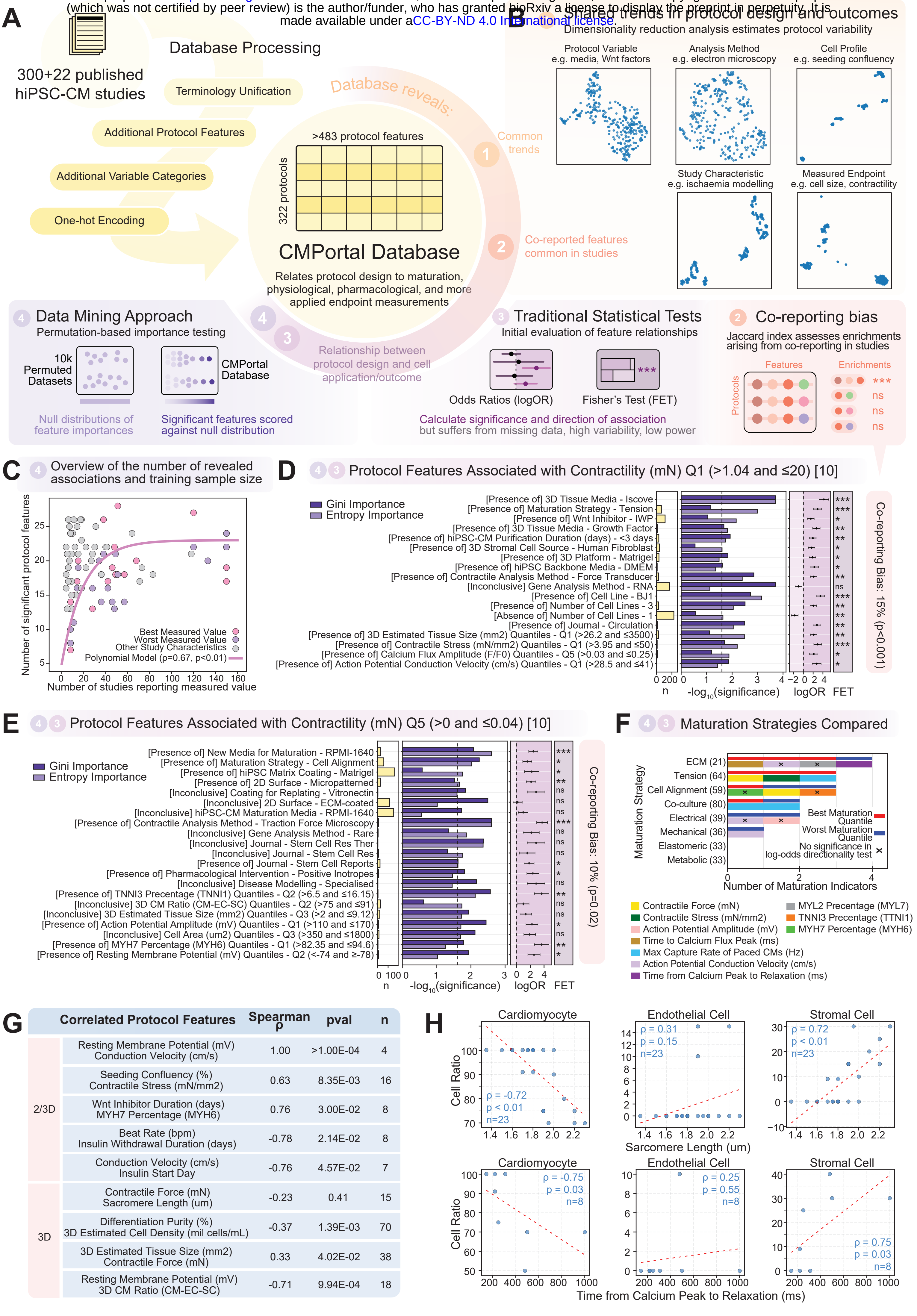
(J) Representative traces and bar plot of impedance-based contractile measurements in hPSC-CM across time show higher peak amplitude for DMEM+FA and RPMI+B27 cells.

(K) Bulk RNA-seq analysis of pathway-specific gene panels indicates metabolic reprogramming towards higher maturity in hPSC-CMs cultured in DMEM+FA compared to RPMI+B27.

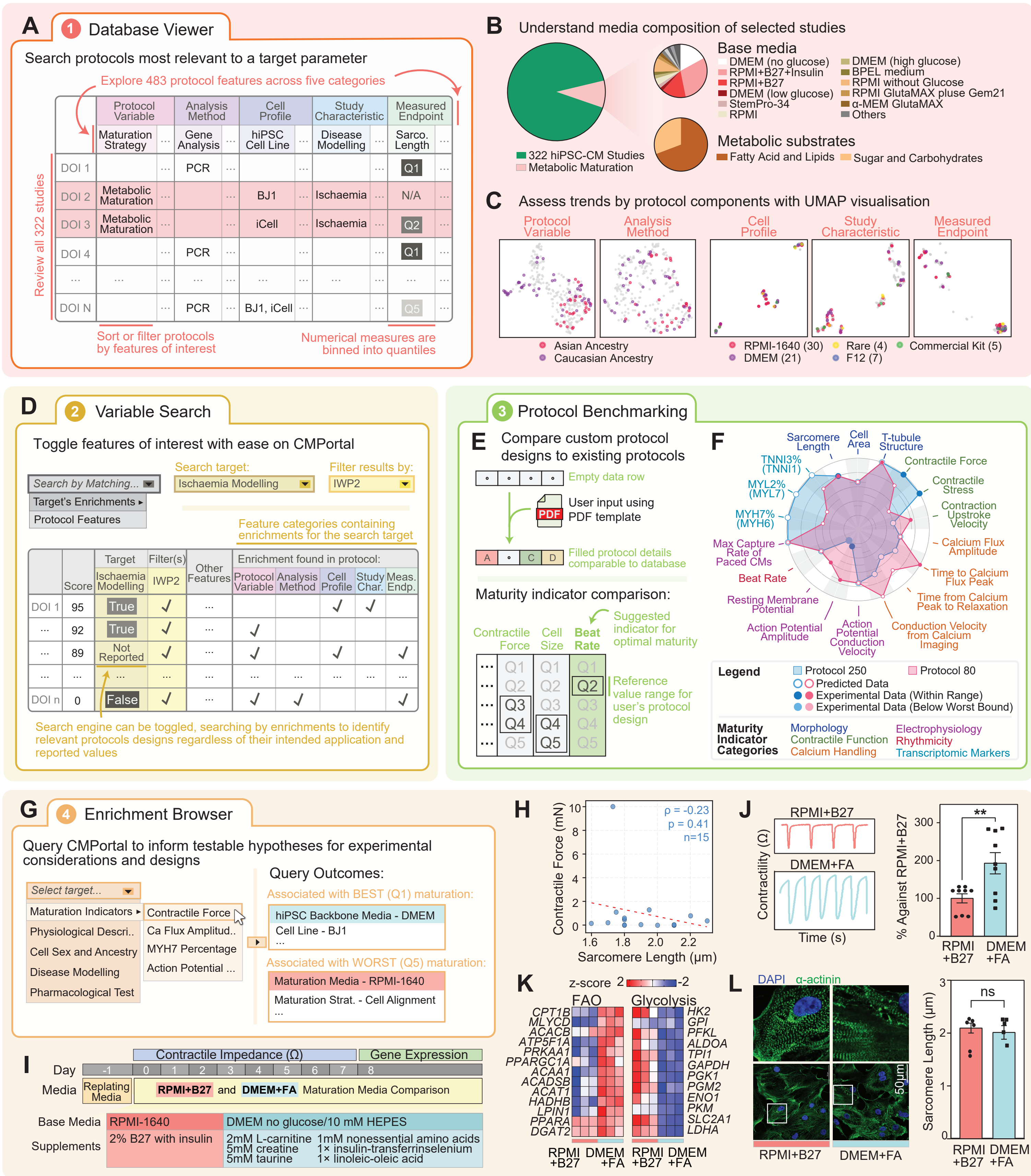
(L) Representative alpha-actinin staining (green), DAPI (blue) in fixed hPSC-CM confirms independence of contractility from sarcomere length. Bar charts values are mean ± standard error whiskers with student's *t*-test significance (G, I). *P<0.05, **P<0.01,***P<0.001, ****P<0.0001, ns: non-significant. Study Char. : Study Characteristic; Meas. Endp.: Measured Endpoint; Sarco. Length: Sarcomere Length.

References

1. Murry, C. E. & Keller, G. Differentiation of Embryonic Stem Cells to Clinically Relevant Populations: Lessons from Embryonic Development. *Cell* **132**, 661–680 (2008).
2. Yang, X., Pabon, L. & Murry, C. E. Engineering Adolescence: Maturation of Human Pluripotent Stem Cell–Derived Cardiomyocytes. *Circulation Research* **114**, 511–523 (2014).
3. Ewoldt, J. K. *et al.* Induced pluripotent stem cell-derived cardiomyocyte in vitro models: benchmarking progress and ongoing challenges. *Nat Methods* **22**, 24–40 (2025).
4. Ewoldt, J. *et al.* Induced pluripotent stem cell-derived cardiomyocyte in vitro models: tissue fabrication protocols, assessment methods, and quantitative maturation metrics for benchmarking progress. 192537 bytes Dryad https://doi.org/10.5061/DRYAD.KSN02V7BH (2024).
5. Lyra-Leite, D. M. *et al.* A review of protocols for human iPSC culture, cardiac differentiation, subtype-specification, maturation, and direct reprogramming. *STAR Protocols* **3**, 101560 (2022).
6. Lee, S. *et al.* Contractile force generation by 3D hiPSC-derived cardiac tissues is enhanced by rapid establishment of cellular interconnection in matrix with muscle-mimicking stiffness. *Biomaterials* **131**, 111–120 (2017).
7. Tiburcy, M. *et al.* Defined Engineered Human Myocardium With Advanced Maturation for Applications in Heart Failure Modeling and Repair. *Circulation* **135**, 1832–1847 (2017).
8. Fetterman, K. A. *et al.* Independent compartmentalization of functional, metabolic, and transcriptional maturation of hiPSC-derived cardiomyocytes. *Cell Reports* **43**, (2024).



	Correlated Protocol Features	Spearman ρ	pval	n
2/3D	Resting Membrane Potential (mV) Conduction Velocity (cm/s)	1.00	>1.00E-04	4
	Seeding Confluency (%) Contractile Stress (mN/mm2)	0.63	8.35E-03	16
	Wnt Inhibitor Duration (days) MYH7 Percentage (MYH6)	0.76	3.00E-02	8
	Beat Rate (bpm) Insulin Withdrawal Duration (days)	-0.78	2.14E-02	8
	Conduction Velocity (cm/s) Insulin Start Day	-0.76	4.57E-02	7
3D	Contractile Force (mN) Sacromere Length (um)	-0.23	0.41	15
	Differentiation Purity (%) 3D Estimated Cell Density (mil cells/mL)	-0.37	1.39E-03	70
	3D Estimated Tissue Size (mm2) Contractile Force (mN)	0.33	4.02E-02	38
	Resting Membrane Potential (mV) 3D CM Ratio (CM-EC-SC)	-0.71	9.94E-04	18


A
1 Database Viewer
Search protocols most relevant to a target parameter
Explore 483 protocol features across five categories
Protocol Variable
Analysis Method
Cell Profile
Study Characteristic
Measured Endpoint
Maturation Strategy
Gene Analysis
hiPSC Cell Line
Disease Modelling
Sarco. Length
DOI 1
DOI 2
DOI 3
DOI 4
DOI N
PCR
Metabolic Maturation
BJ1
iCell
BJ1, iCell
Ischaemia
N/A
Q1
Q2
Q5
Review all 322 studies
Sort or filter protocols by features of interest
Numerical measures are binned into quantiles
B
Understand media composition of selected studies
322 hiPSC-CM Studies
Metabolic Maturation
Base media
DMEM (no glucose)
RPMI+B27+Insulin
RPMI+B27
DMEM (low glucose)
StemPro-34
RPMI
DMEM (high glucose)
BPEL medium
RPMI without Glucose
RPMI GlutaMAX pluse Gem21
α-MEM GlutaMAX
Others
Metabolic substrates
Fatty Acid and Lipids
Sugar and Carbohydrates
C
Assess trends by protocol components with UMAP visualisation
Protocol Variable
Analysis Method
Cell Profile
Study Characteristic
Measured Endpoint
Asian Ancestry
Caucasian Ancestry
RPMI-1640 (30)
DMEM (21)
Rare (4)
F12 (7)
Commercial Kit (5)
D
2 Variable Search
Toggle features of interest with ease on CMPortal
Search by Matching...
Target's Enrichments
Protocol Features
Search target:
Ischaemia Modelling
Filter results by:
IWP2
Feature categories containing enrichments for the search target
Target
Filter(s)
Other Features
Enrichment found in protocol:
Score
Ischaemia Modelling
IWP2
Protocol Variable
Analysis Method
Cell Profile
Study Char.
Meas. Endp.
DOI 1
95
92
89
0
DOI n
True
Not Reported
False
Search engine can be toggled, searching by enrichments to identify relevant protocols designs regardless of their intended application and reported values
3 Protocol Benchmarking
E
Compare custom protocol designs to existing protocols
Empty data row
User input using PDF template
PDF
Filled protocol details comparable to database
Maturity indicator comparison:
Contractile Force
Cell Size
Beat Rate
Suggested indicator for optimal maturity
Reference value range for user's protocol design
Q1
Q2
Q3
Q4
Q5
F
Sarcomere Length
Cell Area
T-tubule Structure
Contractile Force
Contractile Stress
Contraction Upstroke Velocity
Calcium Flux Amplitude
Time to Calcium Flux Peak
Time from Calcium Peak to Relaxation
Conduction Velocity from Calcium Imaging
Action Potential Conduction Velocity
Action Potential Amplitude
Resting Membrane Potential
Beat Rate
Max Capture Rate of Paced CMs
MYH7% (MYH6)
MYL2% (MYL7)
TNNI3% (TNNI1)
Legend
Protocol 250
Protocol 80
Predicted Data
Experimental Data (Within Range)
Experimental Data (Below Worst Bound)
Maturity Indicator Categories
Morphology
Contractile Function
Calcium Handling
Electrophysiology
Rhythmicity
Transcriptomic Markers
G
4 Enrichment Browser
Query CMPortal to inform testable hypotheses for experimental considerations and designs
Select target...
Maturation Indicators
Physiological Descri..
Cell Sex and Ancestry
Disease Modelling
Pharmacological Test
Contractile Force
Ca Flux Amplitud..
MYH7 Percentage
Action Potential ...
Query Outcomes:
Associated with BEST (Q1) maturation:
hiPSC Backbone Media - DMEM
Cell Line - BJ1
Associated with WORST (Q5) maturation:
Maturation Media - RPMI-1640
Maturation Strat. - Cell Alignment
H
Contractile Force (mN)
Sarcomere Length (μm)
ρ = -0.23
p = 0.41
n=15
J
RPMI+B27
DMEM+FA
Contractility (Ω)
Time (s)
% Against RPMI+B27
**
I
Contractile Impedance (Ω)
Gene Expression
Day
-1 0 1 2 3 4 5 6 7 8
Media
Replating Media
RPMI+B27 and DMEM+FA Maturation Media Comparison
Base Media
RPMI-1640
DMEM no glucose/10 mM HEPES
Supplements
2% B27 with insulin
2mM L-carnitine
5mM creatine
5mM taurine
1mM nonessential amino acids
1× insulin-transferrinselenium
1× linoleic-oleic acid
K
z-score 2 -2
FAO
Glycolysis
CPT1B
MLYCD
ACACB
ATP5F1A
PRKAA1
PPARGC1A
ACAA1
ACADSB
ACAT1
HADHB
LPIN1
PPARA
DGAT2
HK2
GPI
PFKL
ALDOA
TPI1
GAPDH
PGK1
PGM2
ENO1
PKM
SLC2A1
LDHA
RPMI +B27
DMEM +FA
L
DAPI
α-actinin
50μm
Sarcomere Length (μm)
ns