

Design Optimization of Geometric-Confined Cardiac Organoids Enabled by Machine Learning Techniques

Andrew Kowalczewski^{1,2}, Shiyang Sun^{1,2}, Nhu Y. Mai^{1,2}, Yuanhui Song^{1,2}, Plansky Hoang^{1,2}, Xiyuan Liu³, Huaxiao Yang⁴, Zhen Ma^{1,2*}

¹Department of Biomedical & Chemical Engineering, Syracuse University, Syracuse NY, USA

²BioInspired Syracuse Institute for Material and Living Systems, Syracuse University, Syracuse NY, USA

³Department of Mechanical & Aerospace Engineering, Syracuse University, Syracuse NY, USA

⁴Department of Biomedical Engineering, University of North Texas, Denton, TX

MOTIVATION

With the significant advancement in organoid technology, there is a critical need of optimization pipeline that can evaluate the designs of organoid systems under controllable engineering parameters. Integrating micropatterning-based organoid engineering and machine learning techniques, we can quickly analyze the total functional outputs of a single organoid and evaluate the impact of engineering inputs on organoid physiology from a cardiac organoid library with 7 geometric designs. AI-driven analytical approaches were applied to not only understand functional heterogeneity from single cardiac organoids, but also provide mechanistic insights of physiological enhancement associated with different geometric designs.

SUMMARY

Stem cell organoids are powerful models for studying organ development, disease modeling, drug screening, and regenerative medicine applications. The convergence of organoid technology, tissue engineering, and artificial intelligence (AI) could potentially enhance our understanding of the design principle for organoid engineering. In this study, we utilized micropatterning techniques to create a designer library of 230 cardiac organoids with 7 geometric designs. We employed manifold learning techniques to analyze single organoid heterogeneity based on 10 physiological parameters. We clustered and refined our cardiac organoids based on their functional similarity using unsupervised machine learning approaches, thus elucidating unique functionalities associated with geometric designs. We also highlighted the critical role of calcium rising time in distinguishing organoids based on geometric patterns and clustering results. This innovative integration of organoid engineering and machine learning enhances our understanding of structure-function relationships in cardiac organoids, paving the way for more controlled and optimized organoid design.

Keywords: Cardiac Organoids, Human induced pluripotent stem cells, Machine Learning, Artificial Intelligence, Design Optimization

Corresponding Author (Lead Contact): Zhen Ma zma112@syr.edu
318 Bowne Hall, Room 303G
Syracuse, NY 13244, United States

INTRODUCTION

Stem cell organoids have served as valuable models for elucidating the intricate interplay between structural architecture and physiological functionality essential for proper tissue development. Various engineering techniques are employed to generate organoids, such as cell aggregation, surface patterning, and microfluidics, aiming to gain precise control over organoid tissue patterning, structure, and functions¹⁻³. Heart organoids demonstrated great potential in recapitulating both morphogenic events and functional outputs. Notably, heart-forming organoids were developed by embedding human induced pluripotent stem cell (hiPSC) aggregates in Matrigel and differentiating through modulation of the canonical WNT pathway⁴. These cardiac organoids showed distinctive spatial organization of myocardial, endocardial, and septum-transversum-like tissue layers. More importantly, cardiac organoids showed nascent chamber formation, resembling ventricular and atrial structures akin to embryonic hearts^{5,6,7}. It is noteworthy that most cardiac organoids hitherto have been developed from 3D hiPSC aggregates, which typically lack precise control over size and geometry. Our cardiac organoids were produced using micropatterning techniques that allow geometric control of hiPSC differentiation and organoid formation⁸.

Despite the advances in organoid technology, there is still a lack of optimization pipeline that can evaluate the design of organoid systems under controllable engineering parameters. Understanding organoid design principle necessitates a comprehensive analysis of multiple variables concurrently, thereby elevating the dimensionality of data analysis and visualization. Many toolsets like dimensionality reduction (DR) and machine learning (ML) techniques have been widely utilized in bioinformatics to process, analyze, and disseminate large quantities of complex high-dimensional data, while their applications remain largely untapped when it comes to the analysis of physiological properties of organoids. Particularly in cardiovascular research, ML has emerged as a powerful tool for risk prediction by detecting abnormal cardiac events from physiological data, such as electrocardiogram (ECG)⁹⁻¹². For example, deep learning with embedded autoencoders exhibited remarkable accuracy in predicting arrhythmias from patient ECG data¹³. Beyond clinical applications, ML techniques are gaining momentum in hiPSC-based *in vitro* model systems and drug screening platforms^{14,15}. Integrating ML with traditional peak-analysis methods could mitigate data dimensionality, thereby enhancing predictions related to cardioactive effects and the mechanistic actions of cardioactive drugs. For example, our recent work demonstrated that nonlinear data processing techniques increased the accuracy of ML models to predict drug cardiotoxicity¹⁶. Recently, a convolutional neural network was trained to encode the membrane potential data collected from hiPSC-derived cardiomyocytes (hiPSC-CMs) for classification purposes. This deep learning-based approach enabled the calculation of EC50 value for tested drugs based on their proarrhythmic possibility and derivation of torsadogenic safety margin to classify the cardiotoxicity risk¹⁷.

Despite the promising progress in the field of cardiotoxicity, there remains limited studies that apply AI techniques to the engineering design of cardiac organoids. The efficient contraction of cardiac muscles hinges on the proper tissue organization that facilitates their function. These concepts are increasingly applied in tissue engineering applications, shedding light on how biomaterial design can influence tissue structure and function^{18,19}. However, reconceptualization of the design principles of cardiac organoids might require a full cycle of design, build, test, and learn from a relatively large organoid library created under controlled engineering parameters. With increasing sophistication of engineered heart models, the integration of ML approaches becomes increasingly valuable for evaluating how engineering parameters affect cardiac function. In this work, we took the advantages of micropatterning technique to create a cardiac organoid library with 7 geometric designs from circles, rectangles, to stars with varied aspect ratios, aiming to elucidate how organoid's geometry would affect their physiological functions. To achieve this, we introduced an ML-based workflow to analyze over 200 cardiac organoids with 10 unique cardiac function variables from contractile motion and calcium transient analysis performed at Differentiation Day 20. Different AI-driven approaches (**Figure 1a**), including manifold learning, unsupervised data clustering, and ensemble learning, were further applied to not only understand functional heterogeneity from single cardiac organoids, but also provide mechanistic insights of physiological enhancement

associated with different geometric designs. Our unsupervised organoid refinement process based on network analysis could reduce the organoid heterogeneity without human bias and identify the shape-determining physiological properties. This AI-based workflow for organoid analysis and refinement can be potentially used for organoid optimization to tailor cardiac organoid morphogenesis and contractile functionality, providing unbiased guidelines for future “organoid-by-design”.

RESULTS

Generation of cardiac organoids from different geometric designs

Using PEG-based surface micropatterning techniques²⁰, we were able to micropattern hiPSCs and generate cardiac organoids with different 2D geometric designs: circles with varying sizes (200 μm , 600 μm and 1000 μm in diameter); rectangles and stars with same area as 600- μm diameter circles. Rectangle designs were varied based on length-width aspect ratio: 1:1 and 1:4. Star designs varied based on area ratio between center pentagon and triangular vertices. For star 1:1, both center pentagon and triangular vertices have the same area, giving large center area with short and blunt vertices, while for star 1:4, triangular vertices have the total area 4-times larger than the center pentagon, giving small center area with long and sharp vertices. The cardiac organoids were stained with cardiac troponin T (cTnT) and actin, which showed the cardiac tissues localized at the central region of the organoids (**Figure 1b**). Cardiac organoids were generated from micropatterned hiPSCs with calcium fluorescent reporter (GCaMP6f), enabling calcium transient analysis without using calcium dye (**Figure 1c**). Meanwhile, brightfield beating videos were taken from the same cardiac organoids for contractile motion analysis (**Figure 1d & 1e**). These two measurements gave us total 10 different physiological variables, including beat rate, contraction velocity (*Cont_Vel*), relaxation velocity (*Relax_Vel*), contraction-relaxation interval (*Cont-Relax Interval*), fluorescent amplitude (*Ca Amp*), calcium rising time (t_0), calcium decay time (t_{50} and t_{75}), pulse duration (*FWHM*), and area ratio between the projected area of GCaMP fluorescence) and the area of the original 2D pattern design (*Area Ratio*).

From 7 geometric designs (circle-200, circle-600, circle-1000, rectangle 1:1, rectangle 1:4, star 1:1, and star 1:4), we collected 230 cardiac organoids, and each of them contained 10 physiological variables measured at Differentiation Day 20. First, we performed a hierarchical clustering of these organoids to evaluate the functional closeness of the organoids produced from different geometric designs (**Figure S1a**). Overall, circle-200 organoids showed a distinct functional profile compared to all the other organoids. The organoids from star shapes (1:1 and 1:4) and rectangle 1:1 had a relatively closer relationship. Next, we examined the correlation between different physiological parameters for the organoids from different shapes (**Figure S1b**). In general, Ca decay time (t_{50} and t_{75}), FWHM, and Cont-Relax Intervals had a strong positive correlation among themselves, but a negative correlation with beat rate. Interestingly, the area ratio had a positive correlation with calcium dynamics (FWHM, t_{50} and t_{75}) for the large circle (600 and 1000) and rectangle 1:1 organoids, but a negative correlation for the rectangle 1:4 and small circle-200 organoids, indicating that geometric inputs altered the correlations between physiological functions of the cardiac organoids. Though these bulk analytical methods demonstrate the overall similarity among various organoid shapes, they are limited in capturing the inherent heterogeneity within the organoid population.

Profiling single organoid heterogeneity using manifold learning

To understand the functional heterogeneity from single organoids, we used unsupervised ML techniques to reduce 10 physiological variables into 2D space for data visualization and clustering. Four manifold learning algorithms were used to visualize high-dimensional data and investigate the relationships within the same cardiac physiological dataset, including t-distributed stochastic neighbor embedding (t-SNE), uniform manifold approximation and projection (UMAP), triplet manifold approximation and projection (TriMAP), and pairwise controlled manifold approximation and projection (PaCMAP). These four algorithms have different strengths in regard to preserving the global or local structure of higher dimensional data (**Figure 2a**).

Both t-SNE and UMAP algorithms have been widely used in single-cell transcriptomics analysis. t-SNE assigns probability distributions over the high-dimensional data points in the low-dimensional map but skews more toward preserving local data structures²¹. UMAP improves upon t-SNE by using non-normalized exponential probability distribution that is not dependent upon Euclidean distances²². These two manifold learning algorithms showed similar organoid distribution with different geometric designs. Distinct separation was observed from small circle-200 organoids from the other larger organoids (**Figure S2a and S2b**). In addition, there is an obvious separation of star (1:1 and 1:4) organoids away from circle (600 and 1000) and rectangle 1:4 organoids. TriMAP improves the preservation of global data by using triplet constraints to embed into a lower dimensional space. Instead of just comparing two unique measures in the pairwise analysis of t-SNE and UMAP, TriMAP conducts a pairwise analysis of two possible neighbors with an additional third point outside of the nearest neighbors. This ensures that globally distance points remain distant at the expense of less accurate local pairings. TriMAP helped the separation of larger organoids (circle-600, circle-1000, and rectangle 1:4) over the smaller ones (**Figure S2c**). We also found that circle-200 organoids had a closer relationship to star-shaped organoids in TriMAP, in comparison to t-SNE and UMAP. Furthermore, we tested a manifold learning algorithm, PaCMAP, which conducts an adaptive pairwise analysis where a single point is trained on both a local neighbor and a distant neighbor within the same iteration, aiming to better preserve global higher dimensional data structures. In our data, PaCMAP sacrificed the clear separation among organoids that was observed from other learning algorithms (**Figure S2d**). However, it illustrated a gradient change of organoid functions cross different geometries: starting from the circle-200 organoids at the bottom-left of the graph to the star 1:1 organoids, next to a mixture of star 1:4 and rectangle 1:1 organoids, and then rectangle 1:4 organoids, and lastly branching into circle-1000 and circle-600 organoids at the top-right of the graph. Overall, different algorithms demonstrate their capability of mapping organoid heterogeneity, t-SNE and UMAP perform better on separation of organoid clusters, while TriMAP and PaCMAP perform better on elucidation of global relationship of organoids with different geometric designs.

Unsupervised organoid clustering based on physiological functions

Next, we clustered the organoids into 7 clusters using the k-means clustering algorithm based on the low-dimensional data extracted from each manifold learning method (**Figure 2b**). This approach allows us to identify the organoids in close proximity to one another and to explore variations in clusters produced by different manifold learning algorithms. From the composition tables (**Table S1**), circle-200 organoids only grouped into one cluster across different methods (cluster 4 for t-SNE, UMAP, TriMAP, and cluster 5 for PaCMAP). Circle-600 organoids had a better clustering efficiency (>60% coverage across different methods) than the circle-1000 organoids, which showed higher heterogeneity in TriMAP and PaCMAP. In contrast, star-shaped organoids were grouped into several different clusters, indicating that complex geometries produced the cardiac organoids with highest functional heterogeneity. In addition, different manifold learning methods showed different efficiency in clustering rectangle-shaped organoids: t-SNE (cluster 2) and UMAP (cluster 6) had >90% coverage of rectangle 1:4 organoids, while TriMAP (cluster 1) had 93% coverage of rectangle 1:1 organoids. This suggests that further analysis should be performed to investigate the overlapping organoids between the clusters produced from different manifold learning methods.

To identify the key physiological differences associated with organoid clustering, we then performed statistical comparison of the physiological variables for different organoid clusters (**Figure 2c**). The circle-200 organoid cluster showed a significantly longer calcium transient duration (*decay time* and *FWHM*) and contractile cycle duration (*Cont-Relax Interval*). We also identified the organoid clusters with highest contractile motion velocity (*Cont_Vel* and *Relax_Vel*) from t-SNE (cluster 7), UMAP (cluster 2), and TriMAP (cluster 6) are mostly comprised of circle-600 organoids. The organoid clusters with largest area ratio were cluster 2 (t-SNE), cluster 6 (UMAP), cluster 2 (TriMAP) and cluster 4 (PaCMAP), from which more than 60% organoids were generated from rectangle 1:4 shape. Notably, these clusters also had shortest calcium rising time (t_0), indicating fast calcium handling capability. Surprisingly, a cluster that clearly stood out from all the dimensionality reduction (DR) maps (cluster 1 in t-SNE, cluster 7 in UMAP, cluster 7 in TriMAP, and cluster 6 in PaCMAP) was comprised of mostly star 1:4 organoids, which showed relatively faster contractile velocity (*Cont_Vel* and *Relax_Vel*) and longer contractile duration (*FWHM* and *Cont-Relax Interval*).

Unsupervised organoid refinement through network analysis on shared organoid communities.

We generated a total of 28 organoid clusters from 4 manifold learning algorithms that re-grouped these shape-patterned cardiac organoids based on functional similarity (**Figure 3a**). To further elucidate the interconnectivity among these organoid clusters, we performed a network analysis by identifying co-presenting organoids across the clusters from different DR maps (**Figure 3b**). For example, UMAP cluster 3 shared the organoids of star 1:1 shape that also presented in the t-SNE cluster 5, TriMAP clusters 4 and 5, PACMAP clusters 2 & 5. Next, we computed the number of connecting clusters (**Figure 3c**) for each cluster to rank their interconnectivity. From this network analysis, we can easily conclude that circle-200 organoids are self-contained within the clusters that have least connectivity with other geometries. In contrast, PaCMAP had 3 clusters ranked in the top 6 clusters, including the top 1 cluster with the highest connectivity, indicating that PaCMAP method resulted in relatively poor organoid clustering due to its preservation of global organoid relationships. To better understand the community structure within our organoid cluster network, we performed fast greedy optimization that can gather the nodes into the communities where there are more connections within the community but fewer connections between them (**Figure 3d**). The organoid cluster network can be divided into 6 geometry-representing communities. Rectangle 1:1 community is the centroid of the network, while circle-200, rectangle 1:4, and star 1:4 communities are distant satellites. Interestingly, only 6 organoid communities were grouped in comparison to 7 geometric designs. Circle 1000 organoids were split into rectangle 1:4 communities and circle 600 communities, but not presented as the dominant organoid population in either community.

From these organoid cluster communities, we refined our organoid population by only retaining the organoids presented within its own community and meanwhile pruning the organoids shared by two or more communities (**Figure S3a**). This network-based unsupervised organoid refinement can reduce the organoid heterogeneity observed in the clustering results and elevate the core organoid representatives for each geometric design (**Figure S3b**). This process improved the correlation between the physiological variables for the organoids (**Figure S3c & S3d**). For example, area ratio and contraction velocity showed much stronger positive correlation with contractile duration parameters (*FWHM*, *Cont-Relax Interval*, t_{50} , and t_{75}) after refinement than before. Interestingly, improved correlation indicated that area ratio had a positive relationship with contraction velocity but a negative relationship with relaxation velocity. More importantly, this unsupervised organoid refinement did not significantly alter the overall data distribution within each organoid design (**Figure S4**) but enhanced the statistical power for the comparison between different organoid shapes (**Figure S5**). It allows us to identify the unique physiological properties associated with geometric designs of the organoids (**Figure 3e**). Compared to the entire organoid population, refined circle-200 organoids had most physiological variables (6 variables), followed by refined rectangle 1:4 organoids with 5 variables, which can distinguish them from the other organoids. For example, refined circle-200 organoids had longer time durations (e.g. *FWHM*, t_{75}), while refined rectangle 1:4 organoids had shorter ones. For the other shapes, refined star 1:4 organoids intended to associate with smaller area ratio, refined star 1:1 organoids with longer Ca rising time t_0 , and refined rectangle 1:1 organoids with faster *beat rate*. Interestingly, refined circle-600 organoids were not significantly different from the entire organoid population, which makes them the baseline organoids for future comparison.

Identifying shape-determining physiological variables using ensemble machine learning classifiers

Finally, we used ML techniques as a hypothesis-generating platform to provide biological basis underlying the phenotypic changes of cardiac organoids attributed to different geometric conditions, as well as the driving force behind the apparent organoid clusters in different DR maps. We employed two ensemble machine learning techniques: Random Forest (RF) and Extreme Gradient Boosting (XGB), which construct hierarchical decision models in which each node makes a decision based on a conditional test. The resulting model comprises multiple decision nodes in a flowchart structure that loosely resembles a tree structure with multiple branches. Ensemble learning models also allow us to trace every optimized decision through the hierarchy of decisions and investigate the importance of each feature on the model's predictions. To assess feature importance, we employed the permutation feature importance method, which quantifies the reduction in impurity or uncertainty at each branch of the model when evaluated on a held-out testing group.

These two classifiers RF and XGB were trained in parallel using 10 physiological variables from 230 cardiac organoids (**Figure 4a**). The classification of cardiac organoids into different geometric shapes resulted in comparable accuracy for both algorithms (79% for RF and 76% for XGB). For the classification of cardiac organoids into corresponding clusters in different DR maps, XGB outperformed RF for the organoid clusters in t-SNE and UMAP (accuracy: GB 95% vs RF 88%), but both models shared the same accuracy for the organoid clusters in TriMAP and PaCMAP (accuracy: 86%). These evaluations verified that our ML models can successfully classify the cardiac organoids based on their geometric shapes and functional clusters, thus we can use them to assess the feature importance from the classifiers trained to make corresponding predictions. We found that both RF and XGB classifiers trained to distinguish the organoids with different geometric shapes used Ca^{2+} rising time (t_0) as the most important feature at ~20% of the model (**Figure 4b**), followed by *contraction velocity* and *area ratio* as second and third features. Surprisingly, Ca^{2+} decay time (t_{50} and t_{75}), which generally are general indicators of arrhythmic behaviors of cardiac tissues, are not deemed as important features associated with geometry prediction. For classification of organoid clusters, both RF and XGB classifiers used the same feature as the geometry classification, Ca^{2+} rising time (t_0), as the most important feature for model prediction (**Figure S6**). These suggested the consistency of the most important physiological function of cardiac organoids between geometric design and functional clustering. However, the second feature in classifying the organoid clusters was beat rate, indicating that organoid clustering based on manifold learning is more dependent on beat rate, instead of contraction velocity.

Experimental validation of AI discovery.

Based on the results of unsupervised machine learning, we found that rectangle 1:4 organoids were significantly enriched in the organoid population during the organoid refinement process due to their unique functionality with enhanced calcium handling capability. In comparison, the physiological properties of circle-600 organoids showed no significant difference from the entire organoid population, which makes them the baseline organoids to represent overall cardiac organoid physiology. Therefore, we compared rectangle 1:4 organoids (function-enhanced organoids) and circle 600 organoids (baseline organoids) using an ACTN2-reporter hiPSC line to visualize the sarcomere structures. We tracked the morphological changes of the same living organoids during the cardiac differentiation (Day 0 – Day 20) (**Figure 5a & 5b**). Overall, we observed that the onset expression of sarcomere α -actinin started from Day 8/9 for both rectangle and circle organoids. On Day 20, we fixed the ACTN2-organoids for confocal microscopy and reconstructed 3D organoid images of rectangle and circle organoids (**Figure 5c**). From zoom-in individual images of these organoids, we observed better alignment of sarcomere structures in the rectangle 1:4 organoids, but random sarcomere organization from the circle 600 organoids due to the lack of anisotropic geometric confinement (**Figure 5d**). Quantitative sarcomere analysis showed no significant difference in sarcomere distance (**Figure 5e**), but a higher sarcomere orientation score from rectangle 1:4 organoids, indicating better sarcomere alignment (**Figure 5f**). Lastly, we observed that the height of the circle 600 organoids was taller than the rectangle 1:4 organoids (**Figure 5g**), which indicated anisotropic patterns might limit the organoid growth in 3D.

Next, we performed bulk RNA sequencing on rectangle 1:4 and circle 600 organoids to elucidate the transcriptomics difference between these two distinct organoid populations. Principle component analysis (**Figure 6a**) and hierarchical clustering (**Figure 6b**) showed a clear separation between rectangle and circle organoids and significant difference in gene expression patterns. Overall, we observed that rectangle 1:4 organoids showed enhanced expression of sarcomere-related genes (MYL2, MYL4, MYH6, MYH7, TNNT2, and TNNC1), calcium handling (RYSR2, CACNA1C), and epicardial differentiation (TBX18, WT1, TCF21) (**Figure 6c**). The gene associated with gap junction (GJA1), ion exchanges (ATP1A1, ATP1B1, ATP2A2), and stromal cell differentiation (TAGLN, ACTA2, SMTN) were found no significant difference between these two cardiac organoid populations. Surprisingly, rectangle 1:4 organoids showed an increased expression of transcriptional factors related to the first heart field (TBX5 and NKX2.5), but a decreased expression of ISL1 which is related to the second heart field. Rectangle 1:4 organoids showed up-regulation of cardiac-specific genes (MYL2, NPPA, GATA4), while down-regulation of early developmental genes (SOX2, MSX2, WNT3A, FOXD3) (**Figure 6d**).

Based on all the up-regulated genes in rectangle 1:4 organoids, we performed gene enrichment analysis and found rectangle organoids showed enhanced genetic patterns in cardiac development and contraction (*GO biological processes*), calcium signaling and cardiac contraction (*KEGG*), and calcium regulation and cardiac progenitor differentiation (*WikiPathways*) (**Figure 6e**).

Last, we validated whether our ML models can be used to predict these two organoid shapes in a blinded fashion. A separate testing dataset was created from a new batch of cardiac organoid differentiation from rectangle 1:4 and circle 600 patterns. We projected the new data from rectangle 1:4 organoids and circle 600 organoids into our previously learned UMAP embeddings based on the original dataset, which allows us to quickly visualize the new dataset relative to the original dataset (**Figure S7a**). We observed that rectangle 1:4 organoids were neatly projected into an expected portion of the UMAP embedding, indicating that newly differentiated rectangle 1:4 organoids had similar behaviors to the same population of the original dataset. New circle 600 organoids also cluster close to the original circle 600 organoids, though 3 of the new circle 600 organoids were projected to the region of the star geometries. Next, we utilized the two ensemble machine learning models (RF and XGB) that were built based on the original dataset to predict the geometries (rectangle 1:4 and circle 600) of the new dataset (**Figure S7b**). For RF, the accuracy for the classification of geometric designs dropped from 76% with the original validation dataset to 61% with the new testing dataset. For XGB, the classification accuracy dropped from 79% with the original validation dataset to 68% with the new testing dataset. The precision of predicting rectangle 1:4 organoids (0.75 for RF and 0.93 for XGB) was much higher than the precision of predicting circle 600 organoids (0.6 for RF and 0.55 for XGB), indicating a higher reproducibility of rectangle 1:4 organoids resisting to the differentiation batch variations. Overall, the model performance is acceptable for evaluating a completely new dataset produced with batch variations from hiPSC differentiation. These results showed the generalizability of our models for incorporating future new datasets and learning from a larger organoid library.

DISCUSSION

In this work, we generated an organoid library with various geometric designs and integrated an AI-driven workflow to enhance organoid analysis in an unsupervised manner without human bias (**Figure 7**). We were able to visualize and cluster the cardiac organoids based on their physiological outputs based on different manifold learning techniques. Particularly, PaCMAP-based dimensionality reduction better preserves the global organoid relationship among different geometric designs in a continuous way, highlighting a gradient change in physiological properties from star 1:4 organoids to circle 1000 organoids, although this technique has its limitation on poor clustering results (**Figure 2**). More importantly, we developed a network-based organoid refinement process that can reduce the organoid heterogeneity and feature the organoid representatives for each geometric design (**Figure 3**). Using supervised machine learning models, we identified that calcium handling capability, particularly calcium upstroke, is the most important physiological parameter to distinguish different organoid shapes (**Figure 4**). Additionally, we compared the organoid development, sarcomere organization, and transcriptomics profile between rectangle 1:4 organoids with enhanced calcium handling and circle 600 organoids as baseline organoids (**Figure 5 & 6**), further confirming the importance of calcium signaling and regulation in organoid optimization predicted by the ML models.

Organoid engineering with geometric constraints

Cell patterning methods have been widely used to guide 2D stem cell differentiation via varying geometric constraints^{23–28}. For instance, micropatterned human embryonic stem cells (hESCs) could self-organize into distinct radial domains, expressing germ layer-specific markers reminiscent of early gastrulation^{29,30}. Previous research conducted in our laboratory highlighted the impact of micropattern geometry on the architecture of cardiac organoids generated from simple circular patterns with varying sizes³¹. We posited that variations in mechanosensitive responses could be attributed to cellular localization within patterned cell colonies due to cell-edge contact. To delve deeper into this phenomenon, we designed more complex geometric patterns to generate a library of shaped cardiac organoids to study how changes in these intricate geometries influenced tissue

functionalities. Notably, a recent study demonstrated that cells exhibited higher initial growth rates when subjected to complex geometries, such as cross- and star-shaped pores, in contrast to square-shaped ones³². Computational simulations revealed that the branches of cross-shaped pores were rapidly populated by cells, whereas the convex points remained minimal cell growth until local curvature became positive. These findings suggest that our micropatterned hiPSCs in star shapes underwent mesoderm induction at different cell cycles, leading to heterogeneous differentiation across the various regions (center, edge, or vertex) of the cell colonies. Early differentiation heterogeneity results in significant variations in contractile functions within star-shaped cardiac organoids, as evidenced by our organoid clustering results.

To study the biological tissues with varied designs, integration of experimental platforms with microfabricated posts and computational simulation based on finite element analysis (FEA) have been instrumental in understanding how geometric constraints influence tissue morphogenesis and mechanical properties^{33–35}. For example, these *in vitro* models were developed to determine mechanical properties of microtissues under geometric constraints³⁶ and predict the necking and failure of the biological tissues due to high local stresses³⁷. With the same effort, we also established a similar experimental platform to explore how geometric designs influenced the remodeling of hiPSC-derived stromal tissues³⁸. To complement this effort, we integrated computational simulation based on volumetric contraction to predict the deformations of these constrained tissues. Although these experiment-simulation integrated models offer mechanistic insights into tissue mechanics and remodeling under geometric constraints, computational simulations face substantial challenges when modeling organoid development, which involves dynamic cell differentiation and tissue morphogenesis. Consequently, this study adopts data-driven approaches to illustrate how geometric constraints influence organoid development.

AI applications in organoid research

AI techniques have emerged as a popular solution for imaging enhancement and analysis across a broad organoid landscape, especially when considering that manual screening is both labor-intensive and susceptible to human error. In organoid research, analysis of the vast volume of images and videos has been one of the most demanding and time-consuming challenges to extract insights regarding organoid morphology, size, quantity, and function, serving as vital indicators of organoid development and their responses to environmental changes. For example, a convolutional neural network (CNN) architecture, specifically ResNet, has been effectively employed to trace the dynamic evolution of organoid cultures, thereby profiling their growth and morphological transformations during the development³⁹. In tumor organoids, ML-based imaging analysis was applied to monitor the dynamic responses of patient-derived tumor organoids by acquiring and segmenting images at distinct timepoints to classify organoid resistance to various drugs⁴⁰. Recently, several open-source packages have surfaced to facilitate the training of deep learning models tailored for organoid analysis, including MOrgAna⁴¹, OrganoID⁴², and D-CryptO⁴³, each offering unique capabilities and enhancing the accuracy of organoid segmentation, tracking, and phenotyping.

Particularly in the field of hiPSC-based cardiac models, the physiological functions of beating hiPSC-CMs can be quantified through time-series imaging datasets, which are then analyzed using traditional peak detection methods^{15,44,45}. More recently, an ML-based analytical pipeline has been employed to automate and enhance the assessment of Ca^{2+} transient abnormalities obtained from hiPSC-CMs. Similarly, different AI-based approaches have been successfully applied in categorizing hiPSC-CMs based on their drug responses and genetic deficiencies with high accuracy⁴⁶. In our study, we integrated feature extraction commonly used in the field of cardiac physiology and ML algorithms used to analyze multi-dimensional datasets, thus gaining a better understanding of design optimization of cardiac organoids. Based on single organoid analysis, we profiled the heterogeneity in organoid functionality in association with original engineering designs. More importantly, we demonstrated that ML classifiers can provide interpretable explanations for their predictions on organoid geometry, so they can be used as a hypothesis-generating tool to identify key physiological changes, specifically calcium rising time (t_0) and contraction velocity (Cont_Vel), associated with organoid's design variations.

Calcium upstroke for cardiac organoid quality control.

Ca^{2+} handling is a fundamental function in heart excitation-contraction coupling, linking electrical and mechanical properties to complete cardiomyocyte contraction. During the action potential depolarization, L-type Ca^{2+} channels open for Ca^{2+} influx, which triggers Ca^{2+} release from the sarcoplasmic reticulum (SR) via ryanodine receptor (RyR2) channels, leading to cytosolic Ca^{2+} increase (Ca^{2+} upstroke measured by Ca^{2+} rising time). This initiates the crossbridge cycle for cardiomyocyte contraction until Ca^{2+} returns to baseline, mainly facilitated by the SERCA-2 pump (Ca^{2+} decay measured by Ca^{2+} decay time). Compared to adult cardiomyocytes, hiPSC-CMs have been long recognized as embryonic-like immature cells with inefficient calcium handling, featured as reduced Ca^{2+} transient amplitude and slower Ca^{2+} upstroke⁴⁷. In contrast, growing hiPSC-CMs into engineered 3D aligned tissues could improve calcium signaling and responsiveness to caffeine, as an indicator of maturation⁴⁸. Previous studies showed that hiPSC-derived cardiac microtissues had higher calcium amplitude, as well as longer upstroke and decay times on the stiffer substrates, indicating that cardiomyocytes can sense and respond to the changes in the local microenvironment⁴⁹. In our study, Ca^{2+} rising time was used as the most important feature by the ML models to classify both pattern geometries and organoid clustering, indicating that the extracellular environment could alter the cardiac organoid physiology mainly through calcium handling. It is likely that calcium upstroke might serve as a critical link between tissue geometric designs and its corresponding physiological functions. As a result, ML models repeatedly used this variable to further delineate the heterogeneity within rectangle and star-shaped organoids, which was further confirmed by our bulk RNA sequencing results. If calcium upstroke can be modulated by overexpressing RyR2 channels or promoting t-tubule formation, we might be able to optimize cardiac organoid development, promote cardiac maturity, and reduce organoid heterogeneity.

In conclusion, integration of hiPSC technology, organoid engineering, and AI holds the potential to enhance our understanding and optimization of functional outcomes of *in vitro* engineered organoids. Our approaches open up possibilities for uncovering the relationships between organoid geometric designs and their physiological functions that traditional analytical methods might overlook. It can also serve as a valuable tool in guiding the design optimization for engineered cardiac organoids with specifically tailored physiological functions. This early success in utilizing machine learning for cardiac organoid analysis suggests broader applicability in cardiac-related research, such as predicting organoid pathology and responses to drug toxicity.

Limitations of the study

One limitation of this study is that the organoids were generated exclusively from a single hiPSC line (WTC11), though our previous research has demonstrated the successful generation of cardiac organoids from other hiPSC lines³¹. To enhance the robustness and generalizability of ML algorithms in classifying organoid phenotypes, future studies should incorporate a more extensive and diverse range of hiPSC lines from healthy donors with varying backgrounds, encompassing factors such as sex, ethnicity, and race. The inclusion of organoid data generated from other hiPSC lines will be essential to address population-level variability in organoid physiology. Another limitation in experiments is that only contractile motion and Ca^{2+} transient measurements were utilized in this study, while other physiological characterizations (e.g., action potential) might be critical to obtaining a comprehensive understanding of organoid properties. In this work, structural parameters were only limited to the Area Ratio as an indicator of cardiomyocyte differentiation, while the inclusion of other structural information, such as sarcomere organization or organoid volume, might alter the refinement outcomes and better profile the structural heterogeneity. Especially, the 2D measurement of Area Ratio based on the projected area of GCaMP fluorescence could not 100% represent the variations in the volume of the organoids. In the future, we aim to perform organoid volumetric measurement and calcium transient measurement simultaneously from each cardiac organoid using ACTN-GFP hiPSC line coupled with RFP or NIR calcium dye to further enrich our physiological readouts. As our dataset expands through the collection of other physiological parameters from cardiac organoids, we anticipate the ability to integrate and establish cardiac physiomics datasets for further mechanistic studies.

An essential consideration when training ML algorithms is the dataset size, which, in our study, was constrained due to the time-consuming and labor-intensive process of generating cardiac organoids. Moreover, the efficiency of generating cardiac organoids from smaller patterns (circle-200) was relatively low, resulting in an imbalanced

dataset with fewer organoids from specific pattern geometries. Using an unbalanced dataset in ML models often leads to a potential bias towards the majority class and poor generalization to the minority class. Though we have tried to address this issue by proper selection of ML models and fine-tuning of hyperparameters, enlarging the sample pool of organoids for specific pattern geometries will still be critical to achieving balanced and accurate predictions. In addition, testing a completely new dataset dropped the accuracy of ML prediction, which indicates the batch variability from hiPSC differentiation has a significant impact on the ML model establishment and generalizability. Creating a larger organoid library with different hiPSC lines and different differentiation batches might potentially address this issue.

Ultimately, it is plausible that a deep learning model, trained directly on raw calcium transient signals or fluorescent videos, could emerge as a more effective approach. While deep learning methods have garnered significant attention and hold potential for organoid applications, they typically demand extensive training data to establish a robust and broadly applicable model. Additionally, deep learning models could minimize data loss and maximize information retention, while it's worth noting that this approach might compromise interpretability without feature extraction processes based on traditional quantitative peak analysis.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
 - Cell line
- METHOD DETAILS
 - Micropatterning of tissue culture surface
 - Generation of cardiac organoids
 - Immunofluorescence staining and morphological characterization
 - RNA sequencing
 - Functional characterization of cardiac organoids
 - Manifold learning and dimensionality reduction
 - Organoid clustering
 - Machine learning classification and feature importance
 - Network analysis
 - Organoid refinement
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTARY INFORMATION

Supplementary information can be found online.

ACKNOWLEDGEMENTS

This work was supported by the NIH [R01HD101130 and R15HD108720], the NSF [CMMI-2130192 and CBET-1943798], Research Seed Grants (2021 and 2023) from UNT Research and Innovation Office (H.X.Y.), and Syracuse University intramural CUSE grant [II-3245-2022] (Z.M.).

AUTHOR CONTRIBUTIONS

A.K. and Z.M. conceived the study. A.K. performed the computational work with different machine learning models and data visualization. S.S. and P.H. performed all experiments, collected raw data, and processed the recorded videos. N.Y.M. produced a separate testing dataset from newly differentiation of cardiac organoids. N.Y.M. performed confocal microscopy and sarcomere analysis. Y.S. performed RNA sequencing and bioinformatics analysis. X.L. advised on the applications of machine learning models. Z.M. and H.Y. supervised the project development and funded the study. A.K. and Z.M. wrote the manuscript with discussion and improvements from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

FIGURE

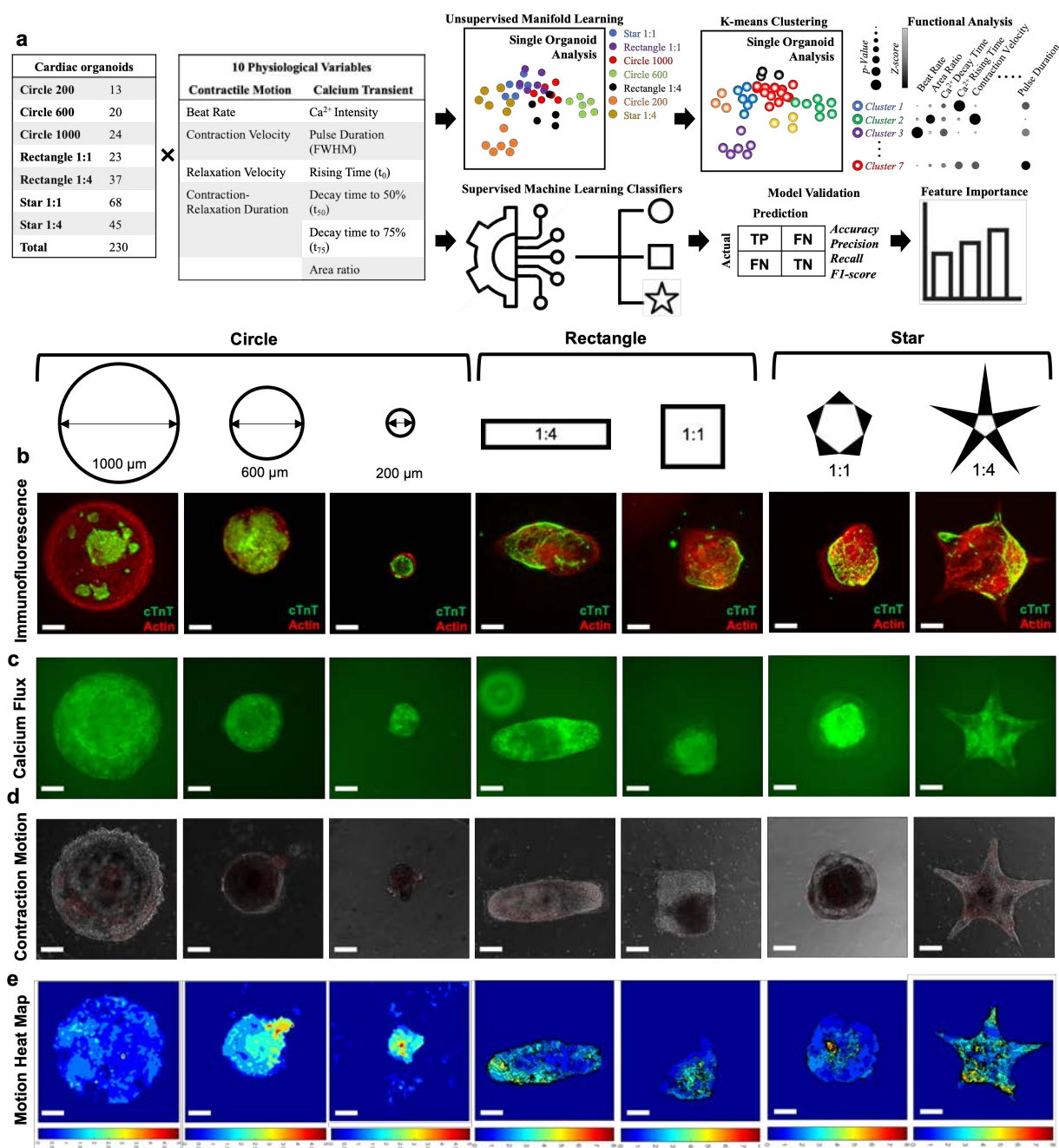


Figure 1. The data-driven single organoid analysis workflow for the cardiac organoid design library. (a) Total 230 cardiac organoids were generated from 7 geometric designs and analyzed for 10 physiological variables. These organoids were analyzed using both supervised and unsupervised machine learning. The cardiac organoids were generated from circle shapes (1000 μm , 600 μm , and 200 μm in diameters), rectangle shapes (1:1 and 1:4 aspect ratio), and star shapes (1:1 and 1:4 area ratio between central pentagram and five vertices). (a) Representative fluorescent images of cardiac organoids stained by cardiac marker (cardiac troponin T) and F-actin. (b) Representative fluorescent images of cardiac organoids for calcium transient analysis, and (c) their corresponding brightfield images for contractile motion analysis, (d) which can generate motion heat map. All the images/videos were taken and analyzed at Differentiation Day 20.

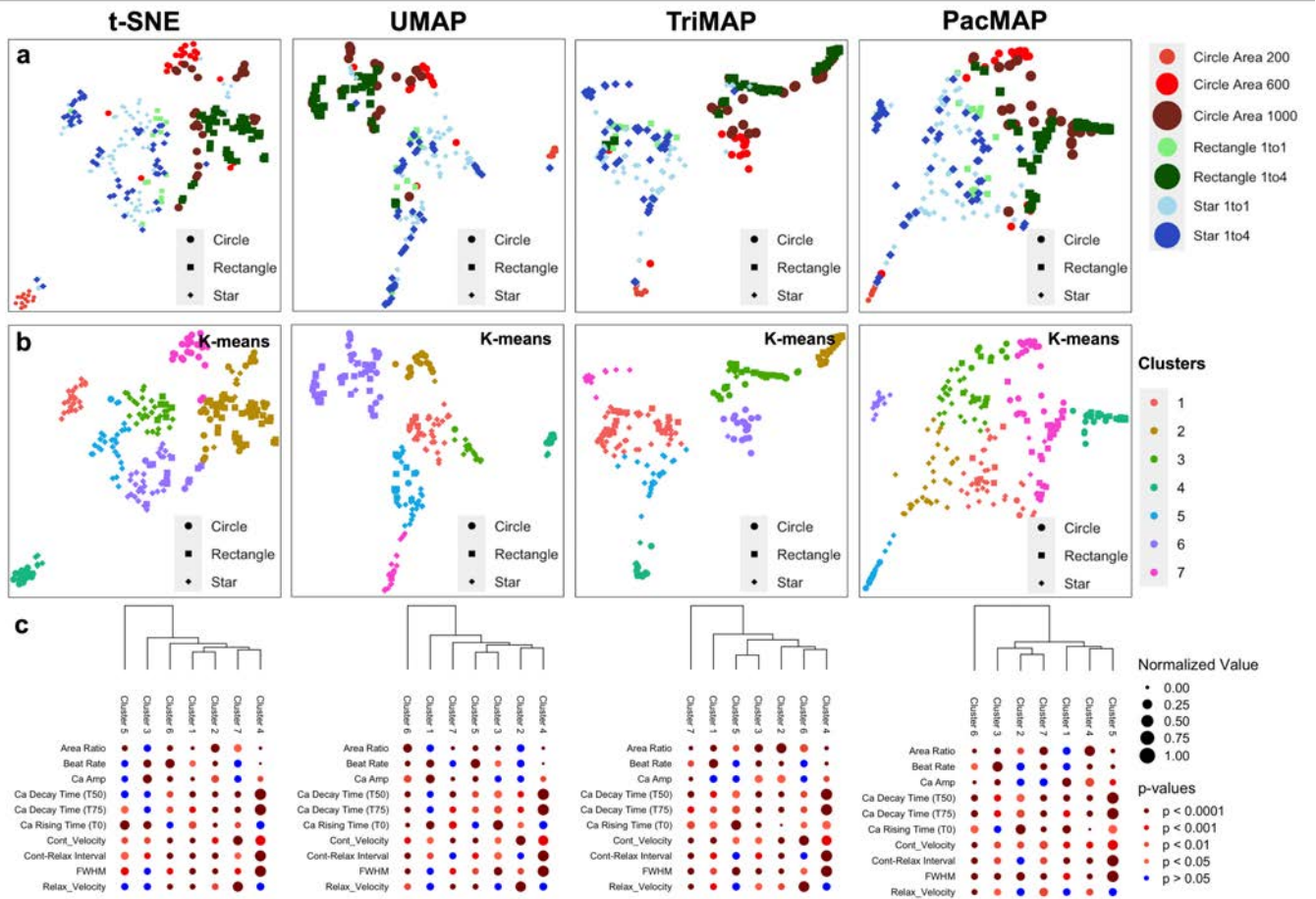


Figure 2. Organoid clustering based on different manifold learning algorithms. (a) The cardiac organoids were analyzed using t-SNE, UMAP, TriMAP, and PacMAP algorithms to generate dimensionality reduction graphs to visualize the spatial relationship of organoids generated from different geometries. (b) The cardiac organoids were re-grouped based on functional similarity using K-means clustering, which gave 7 organoid clusters for each dimensionality reduction graph. (c) Statistical analysis showed that each organoid cluster had unique contractile properties based on 10 physiological variables.

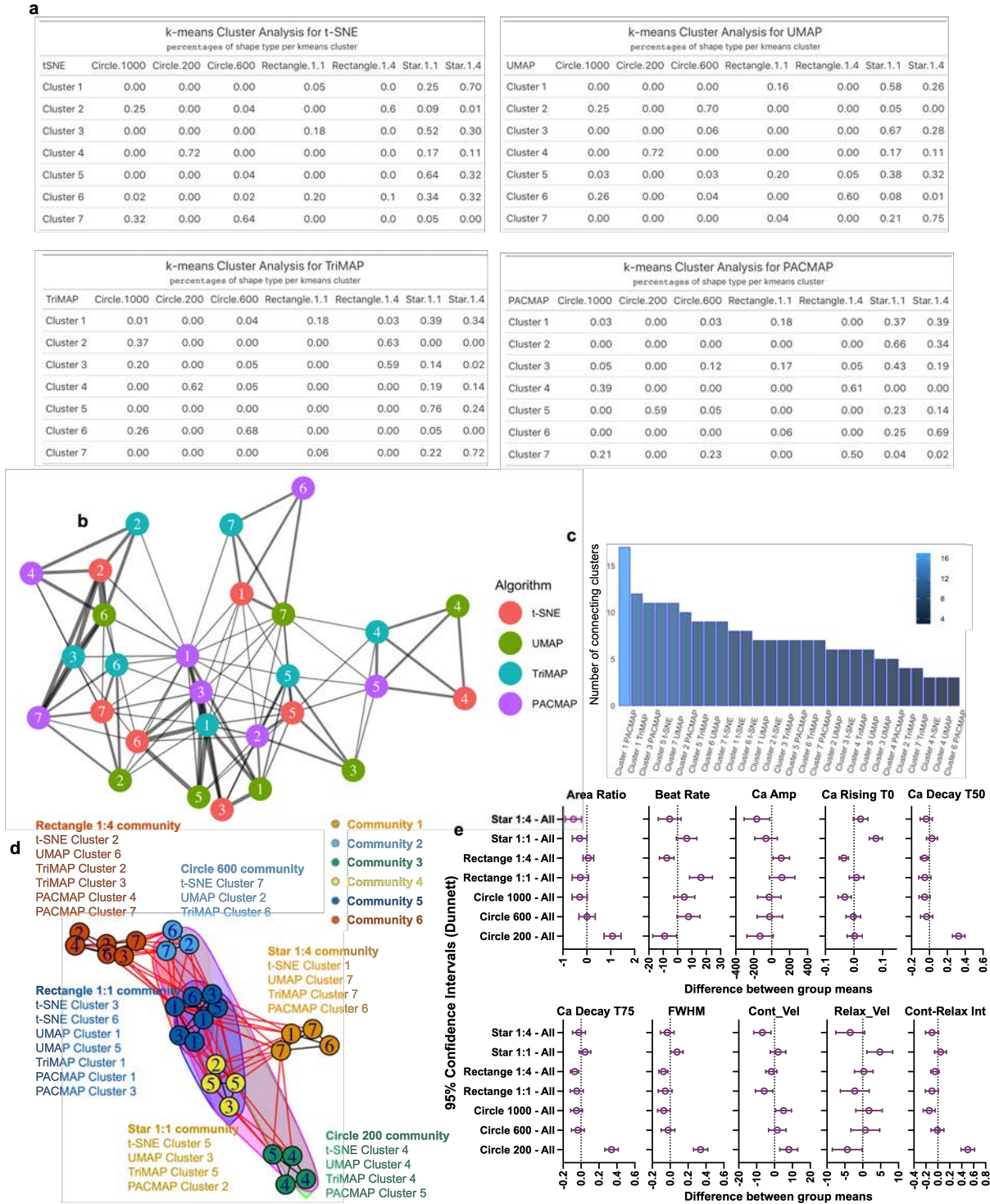


Figure 3. Network analysis on organoid clusters. (a) The organoid composition for each cluster determined by different manifold learning algorithms. (b) Cluster interconnectivity among the clusters across four manifold learning algorithms. (c) Quantifying the heterogeneity of organoid composition within each cluster based on its connections to the other clusters. (d) Organoid cluster network was regrouped into 6 shape-representing communities using Greedy optimization technique. (e) Statistical analysis on refined organoids showed unique physiological functions related to specific geometry.

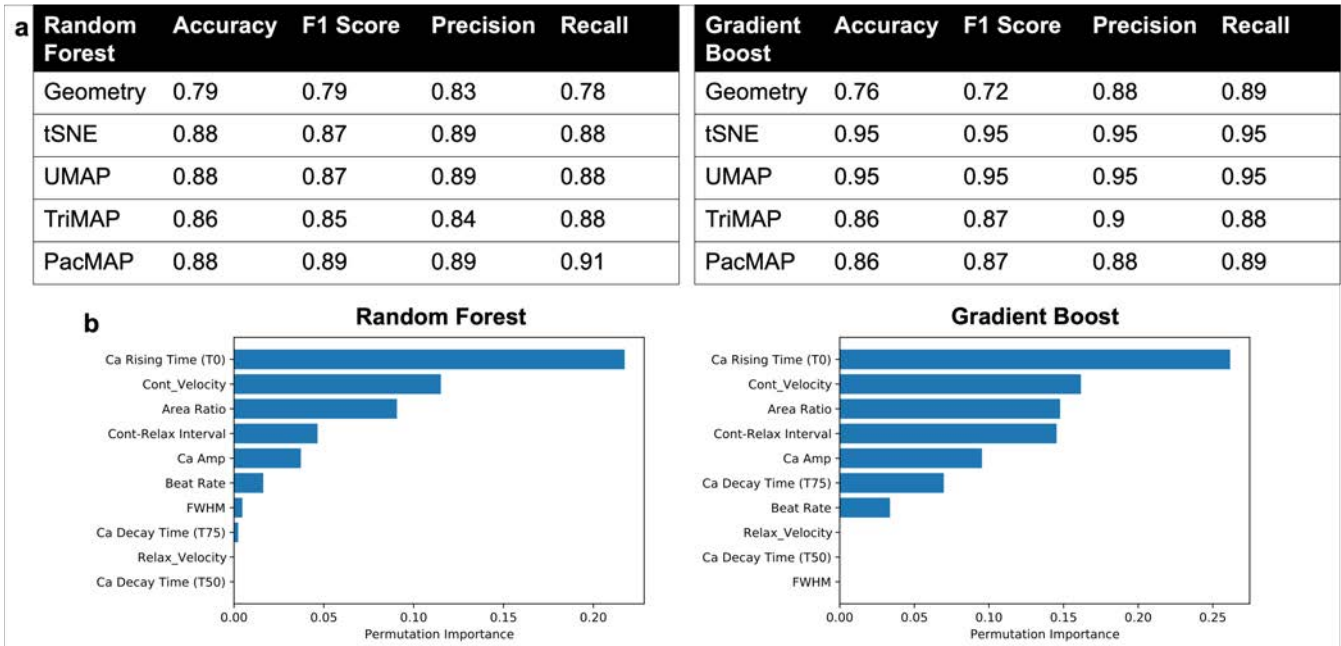


Figure 4. Feature importance analysis based on ensemble learning algorithms. (a) Two ensemble learning models, random forest, and gradient boost, were able to classify the organoids based on their geometric designs or their organoid clusters in each dimensionality reduction graph with relative high accuracy (>75%). (b) Feature importance analysis showed the determining physiological variables that were used by the machine learning models to make predictions on geometric designs of the cardiac organoids.

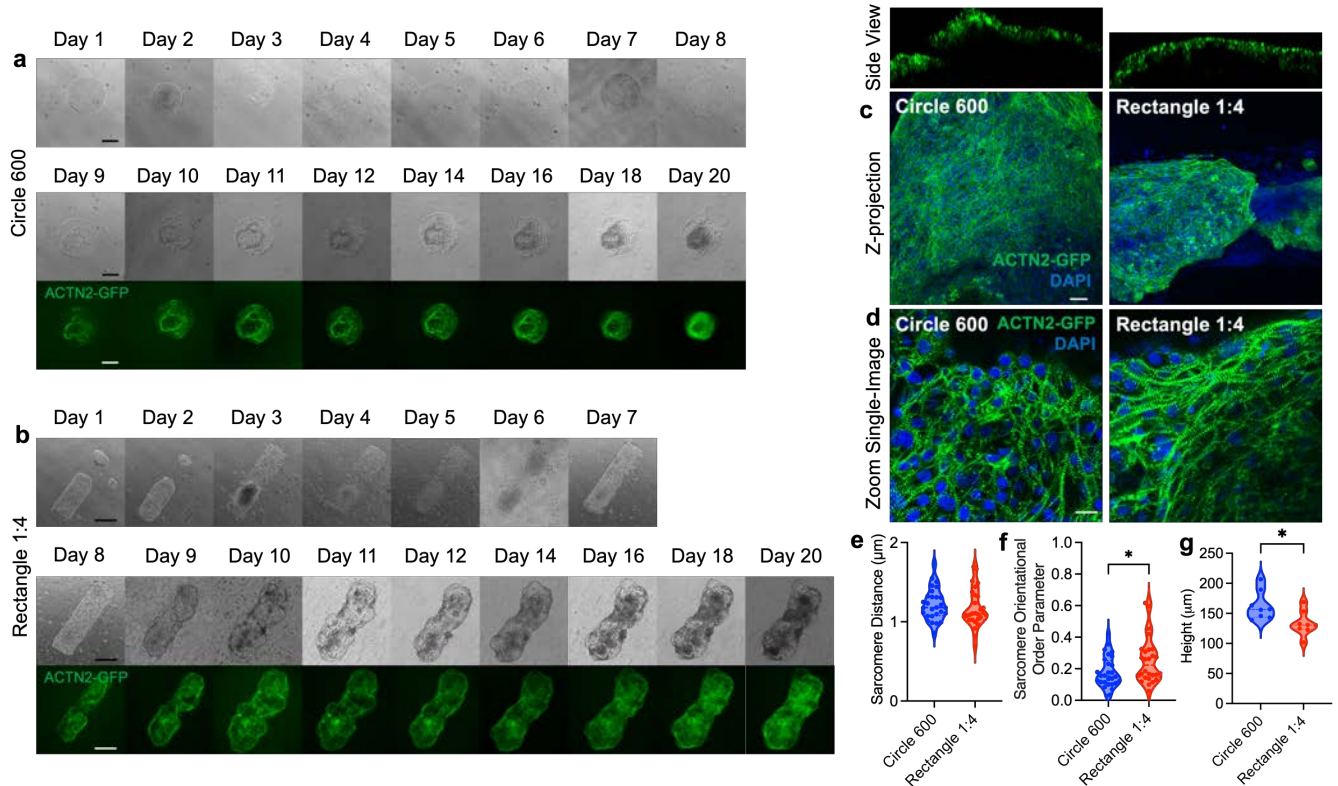


Figure 5. Structural comparison between rectangle 1:4 and circle 600 organoids. Kymographs of organoid differentiation over time for (a) one circle 600 organoid and (b) one rectangle 1:4 organoid generated from the ACTN2-reporter hiPSC line. (c) Confocal microscopy images of a circle 600 organoid and a rectangle 1:4 organoid with corresponding orthogonal views. (d) Representative images of sarcomere structures of a circle 600 organoid and a rectangle 1:4 organoid. (e) No significant difference was found for the sarcomere distance between circle 600 and rectangle 1:4 organoids. (f) Rectangle 1:4 organoids had better sarcomere alignment than the circle 600 organoids, (g) while circle 600 organoids were higher than the rectangle 1:4 organoids. Statistics: two-sided Student t-test. * $p < 0.05$.

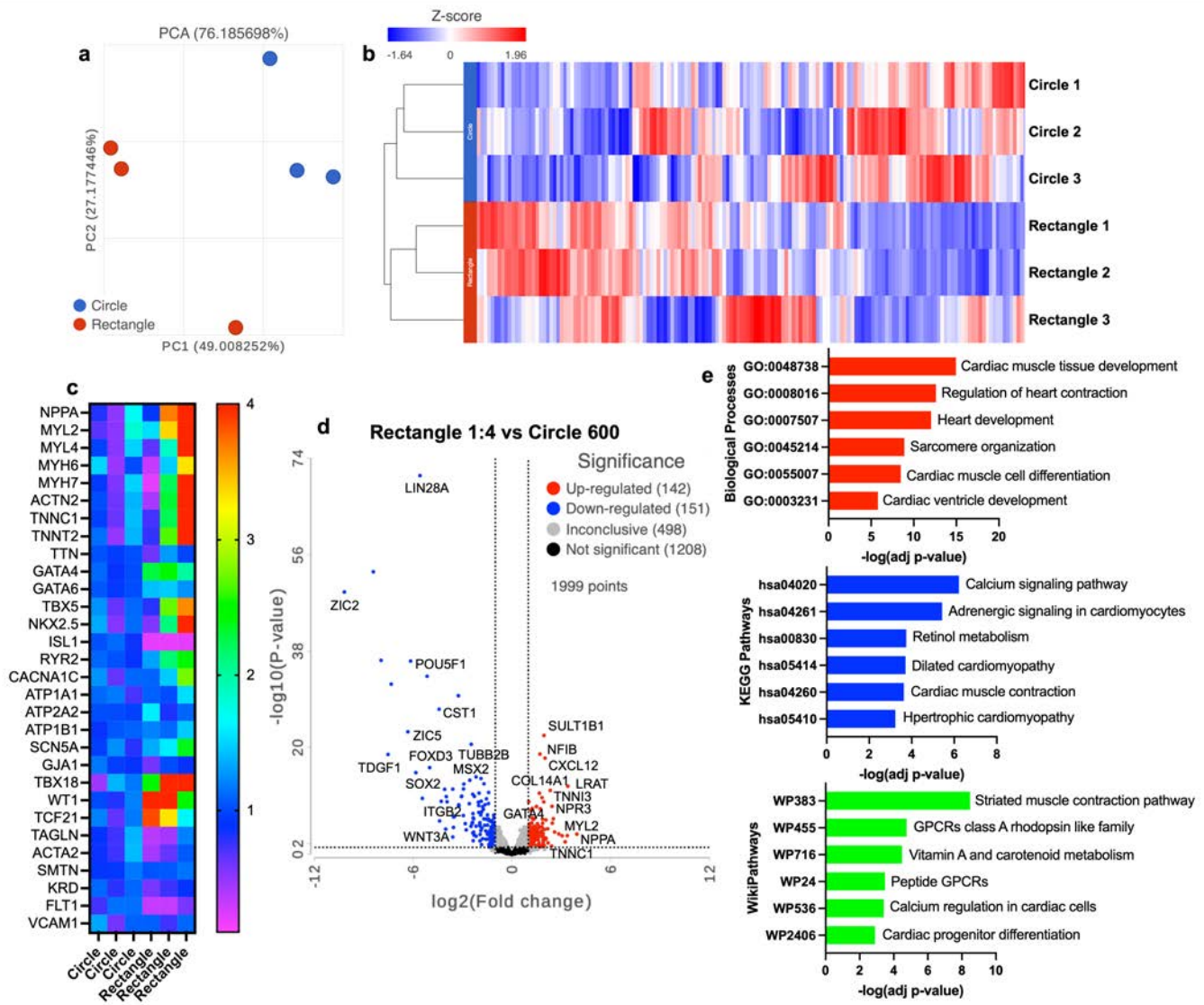


Figure 6. Transcriptomics comparison between rectangle 1:4 and circle 600 organoids. (a) Principal component analysis (PCA) showed the separation of circle 600 organoids and rectangle 1:4 organoids. (b) Heatmap and hierarchical clustering showed distinct transcriptome profiles between circle 600 organoids and rectangle 1:4 organoids. (c) Selected genes associated with sarcomere structure, transcription factors, calcium handling, ion channels, and endocardial/stromal differentiation showed differential expression between circle 600 organoids and rectangle 1:4 organoids. (d) Volcano plots showed rectangle 1:4 organoids upregulated the genes associated with cardiac differentiation and sarcomere structures, and downregulated the genes associated with early embryonic development and WNT signaling. (e) Gene enrichment analysis showed rectangle 1:4 organoids highly enriched the genes associated with heart development, calcium signaling and contractile functions.

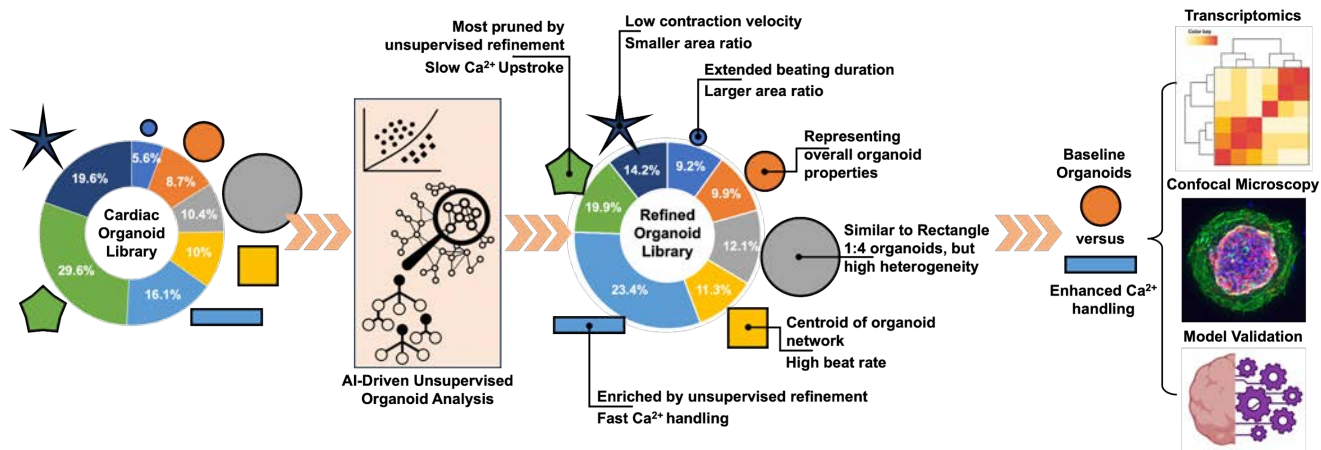


Figure 7. Summary on AI-driven unsupervised organoid analysis. The integration of a design library of cardiac organoids and an AI-driven workflow to achieve unsupervised organoid refinement. Refined cardiac organoids showed shape-specific physiological properties, which allows us to determine the geometric designs with improved cardiac functionality and elucidate mechanistic insights with biological investigations.

STAR METHODS

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Zhen Ma (zma112@syr.edu).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

- The RNA sequencing data is available at GEO with identifier listed in the key resources table. The functional data of cardiac organoid library have been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the key resources table. All relevant data reported in this paper can be shared by the lead contact Dr. Zhen Ma upon request.
- The code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell Lines

Two human induced pluripotent stem cell (hiPSC) lines were used in this study: GCaMP6f-GFP hiPSCs and ACTN2-GFP hiPSCs. GCaMP6f-GFP hiPSC line was obtained from Dr. Conklin lab at Gladstone Institute of Cardiovascular Diseases. ACTN2-GFP hiPSC line was purchased from Coriell Institute with cell line ID (AICS-0075-085). Both cell lines were genetically engineered based on parent hiPSC line (WTC11, GM25256) derived from adult skin (leg) fibroblasts (race: Asian, ethnicity: Japanese, sex: male, age: 30 years old). All the hiPSCs are maintained using Essential 8 (E8) medium (Life Technologies, cat. no. A1517001) in the 6-well plates coated with diluted Geltrex hESC-qualified matrix (Life Technologies, cat. no. A1413302).

METHOD DETAILS

Micropatterning of tissue culture surface

Surface micropatterning on tissue culture polystyrene was carried out using the selective etching approach described previously⁸. Poly(dimethyl siloxane) (PDMS) was cast from SU8 wafers with designed features to produce thin elastomeric stencils with clear-through holes. Non-fouling poly(ethylene glycol) (PEG) solution (PEGMMA 1000 and PEGDA 400 mixture) was grafted onto 6-well tissue culture plates and cured under UV light exposure (Dymax UV Illuminator; model no. 2000EC). Micropatterns were fabricated by selective oxygen plasma etching (Oxygen plasma treatment system, PlasmaEtch PE50XL) of the PEG layer using the PDMS stencils. Micropatterned tissue culture plates were sterilized by immersing in 70% ethanol for 1 hour and subsequent washing with sterile phosphate buffered saline (PBS).

Generation of cardiac organoids

Micropatterned surfaces were coated with diluted Geltrex hESC-qualified matrix (Life Technologies, cat. no. A1413302) and seeded with GCaMP6f-GFP hiPSCs (calcium reporter) or ACTN2-GFP hiPSCs (sarcomere

reporter) in Essential 8 (E8) medium (Life Technologies, cat. no. A1517001). Cardiac differentiation was initiated (Day 0) when the micropatterns reached confluency and performed via small molecules^{50,51} of GSK3 inhibitor CHIR99021 (Day 0) (Stem Cell Technologies, cat. no. 72054) and WNT pathway inhibitor IWP4 (Day 2) (Stem Cell Technologies, cat. no. 72554) in RPMI 1640 medium (Life Technologies, cat. no. 11875093) supplemented with B27-minus insulin (RPMI/B27 minus insulin) (Life technologies, cat. no. A1895601). On Day 6 onward, organoids were maintained in RPMI 1640 medium supplemented with complete B27 supplement (RPMI/B27 Complete) (Life Technologies cat. no. 17504044) until Day 20 for contractile and structural analysis.

Immunofluorescence staining and morphological characterization

Organoids were characterized based on immunofluorescence staining of cardiac tissue with cardiac troponin T (cTnT) and entire organoid structure with actin (Phalloidin). Samples were fixed with 4% (vol/vol) paraformaldehyde (PFA) for 10 minutes, permeabilized with 0.2% (vol/vol) Triton X-100 and blocked with 2% (wt/vol) bovine serum albumin (BSA). Samples were incubated with primary antibody against cTnT (dilution 1:200) for 1 hour and then incubated with secondary antibody, together with Phalloidin for 1 hour at room temperature. Leica Thunder upright fluorescent microscope with a 40X water immersion objective was used to capture z-stacks of the organoids for height measurements and Z-projection in ImageJ. The *Area Ratio* was measured by using the circular or elliptical tool to approximate the area of fluorescence of GCaMP flux for cardiac tissue and normalizing this area relative to the entire pattern area of original geometric design. Confocal microscopy (Zeiss U880) was used to characterize the organoid's morphology and sarcomere structure. Z-stack with 6.5 μ m spacing between slices was employed to reconstruct the 3D organoid for height measurement⁸. Sarcomere analysis was performed on the selected slices from image z-stacks for sarcomere distance and organization measurements⁵².

RNA sequencing

On day 20, cardiac organoids were sacrificed to extract mRNA. The organoids were collected by the cell scraper, centrifuged, and washed with PBS several times to remove all the remaining medium. mRNA was extracted using Trizol (Invitrogen, #15596026), Chloroform (Serva, #39553.01), and isopropanol (Fisher, BP2618). The RNA pellet was dissolved into 40 μ L UltraPure™ DEPC treated water (Invitrogen, #750023) and reserved at -80 °C. The RNA was quantified using a NanoDrop Microvolume UV-Vis Spectrophotometer (Thermo Scientific). The RNA quality was evaluated using Agilent 2100 Bioanalyzer at Molecular Analysis Core, SUNY Upstate Medical University. The samples with RIN $\geq$ 8.0 and concentration $\geq$ 50 ng/ μ L were used for bulk RNA sequencing. The Illumina Ribo-Zero rRNA removal kit was used for rRNA depletion of all the samples. The samples were sequenced using Illumina HiSeq with 2 x 150 bp configuration, single index, paired end reads per lane.

The raw FASTQ files were analyzed using the Partek Flow software, courtesy of a shared license provided by SUNY Upstate Medical Genomics Core. The unaligned reads were trimmed for bases to obtain a Phred quality score > 20, and then aligned using the Spliced Transcripts Alignment to a Reference (STAR) to the human genome (hg38). The post-alignment assessment was conducted for quality assurance (QA) and quality control (QC), which showed the percentage of alignment for each sample was > 75%. The average base quality score per read was between 35.8 and 38.7, indicating good quality reads. Post-alignment quantification was applied to an annotation model and normalized based on recommended parameters of counts per million (CPM). The downstream analysis included principal component analysis (PCA), differential gene expression (DESeq), hierarchical clustering, gene ontology (GO) and pathway analysis.

Functional characterization of cardiac organoids

GCaMP6f hiPSC-derived cardiac organoids were imaged in an onstage microscope incubator (OkoLab Stage Top Incubator, UNO-T-H-CO2) at 37 °C and 5% CO2 to maintain standard physiological conditions on a Nikon Ti-E inverted microscope with Andor Zyla 4.2+ digital CMOS camera. Videos of contracting cardiac organoids were recorded at 50 frames per second for ten seconds in brightfield and exported as a series of single frame image files. Contraction physiology was assessed using video-based motion tracking software that computes motion vectors based on pixel movement⁵³. The motion vectors were assimilated into a contraction motion waveform

representative of contractile physiology. Contraction physiology was also assessed by recording the calcium transient of organoids. Videos were taken under 488 nm excitation at 40 milli-seconds exposure time with 25 frames per second. Calcium flux signals were exported as Z-axis profiles in ImageJ. The fluorescence bleaching decay was corrected and related parameters t_0 , t_{50} , t_{75} were computed using in-house MATLAB scripts⁵³. The rising time t_0 is defined as the time it takes for the calcium flux to reach peak fluorescence intensity, whereas decay time t_{50} and t_{75} represent the time it takes for the calcium flux to decay 50% and 75% of the peak fluorescence, respectively. Data pre-processing was accomplished in RStudio utilizing the base function for scaling⁵⁴ and the dplyr library⁵⁵. We used standard z-score scaling to center the data to improve both our machine learning models and manifold learning algorithms. For newly added testing dataset, we standardized the new measurements according to the mean and standard deviations to correctly match the scale of the original dataset. We identified and removed 4 outliers which returned negative values for time measurements.

Manifold learning and dimensionality reduction

Manifold learning algorithms were used to embed higher dimensional data into a 2D space, thus allowing to visualize the relationships of many parameters. t-Stochastic Neighbor Embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP) were conducted with the statistical programming language R⁵⁴. t-SNE graphs were created using a Barnes-Hut implementation with the R package Rtsne⁵⁶ with a max perplexity of 15. UMAP analysis was conducted utilizing the package uwot⁵⁷ with the nearest neighbor of 15. TriMAP analysis was conducted in Python⁵⁸ with 15 nearest neighbors selected using the TriMAP library⁵⁹. PaCMAP analysis was also conducted in Python with PaCMAP library⁶⁰ with a nearest neighbor of 15 for a max of 5,000 iterations. Projecting the newly acquired testing dataset into the learned manifold embeddings was accomplished based on the learned UMAP embedding from the original dataset.

Organoid clustering

Unsupervised machine learning algorithm K-means was used to explore apparent clustering patterns within each nonlinear dimensionality reduction graph. In essence, K-means aims to partition the available data into a predefined 'K' number of clusters. In our case, every data point in the reduced 2D space is assigned to the nearest centroid within each cluster. Initially, centroids are randomly chosen and iteratively adjusted to minimize average geometric distance to the centroid of each cluster, resulting in the formation of 'k' clusters⁶¹. K-means clustering was performed using the base stats library within R. With a selection of 7 clusters for k we utilized Hartigan and Wong's algorithm which continually updates the centroid for each cluster⁶².

Machine learning classification and feature importance

Random forest (RF) algorithm is used for classification by constructing a number of decision trees, where each decision at a branch is determined from a specific feature⁶³. The randomness within the forest is introduced through a bootstrapping method, where a random subset of the training data is left out during the model's training by using a random vector of the sampled data. This reduces overfitting, a common issue in hierarchical decision models for classification. Extreme gradient boosting (XGB) builds upon RF by boosting the gradient for decision trees to improve the decision-making based on the ensemble of decisions created previously within the model⁶⁴. RF tends to classify data with a degree of randomness, whereas XGB generates more precise classification predictions by leveraging the information within its decision trees. The open-source library scikit-learn was used to implement and optimize both RF and XGB algorithms⁶⁵. Random train-test splits allocated 2/3 of the data to the training set and reserved 1/3 for a validation set for classification tasks. A completely new testing dataset unique to the initial training and validation dataset was also introduced to gauge the generalizability of our proposed model. This additional testing set was obtained from a new batch of cardiac organoid differentiation ~2 years after the original dataset was acquired. The differentiation, culture, and characterization techniques are the same. For hyperparameter tuning, the default configuration builds 100 trees within a forest with no limitations on the forest depth. The maximum number of considered features was set to the standard square root of the total features available. Gini impurity was utilized as the criterion for tree splitting, as well as the basis for calculating feature importance. A fixed learning rate of 0.1 was employed for gradient boosting. The model performance was evaluated using a confusion matrix to compute accuracy, F1 score, precision, and recall.

Confusion matrix	Prediction	
Actual	True positive (TP)	False negative (FN)
	False positive (FP)	True negative (TN)
$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$		
$F1\ score = \frac{2 \times TP}{2 \times TP + FP + FN}$		
$precision = \frac{TP}{TP + FP}$		
$Recall = \frac{TP}{TP + FN}$		

Network analysis

To investigate the relationships between clusters across the four DR algorithms, we utilized the network analysis package *igraph* in R⁶⁶. While network analysis is commonly used to find user similarities within social networks, we used this analysis to investigate not only the functional similarities of our organoids with different geometries, but also the relations within the data that are captured by each DR + K-mean clustering. In the network, each node within the graph represents a cluster from one DR technique, and each connecting edge represents a shared organoid between the clusters across four DR algorithms. More edges between two nodes mean more shared organoids tagged by both clusters. Next, we performed fast greedy clustering on our organoid cluster network. It is a hierarchical agglomeration algorithm available within the R package *igraph*. It is widely utilized to find communities of similar nodes within a network. It seeks to greedily increase the modularity score, related to the number of edges within dense clusters of all nodes. This is accomplished by merging nodes into communities to maximize the local optima for the modularity of the entire network⁶⁷. This rapid method does not require hyperparameter tuning and gives a good approximation for communities existing within the network.

Organoid refinement

Organoids were refined to ensure that each clustered community effectively represents each geometric shape. We refined the organoid measurements using the *dplyr* library⁵⁵. For example, circle-600 organoids contained within network greedy community #2 were retained as the refined circle-600 organoids, while the other 600 organoids were pruned. The organoids for the other geometric designs were refined in the same manner.

QUANTIFICATION AND STATISTICAL ANALYSIS

For single comparisons between two individual groups, a two-sided Student's t-test was used, and $p \leq 0.05$ was considered significant. For comparisons between more than two groups, one-way analysis of variance (ANOVA) was performed and $p \leq 0.05$ was considered significant. ANOVA analysis was supplemented with Tukey's multiple comparison test to determine significance between every two groups, or with Dunnett's comparison test to determine significance between one group and all the others.

REFERENCES

1. Hoang, P. & Ma, Z. Biomaterial-guided stem cell organoid engineering for modeling development and diseases. *Acta Biomater.* **132**, 23–36 (2021).
2. Unagolla, J. M. & Jayasuriya, A. C. Recent advances in organoid engineering: A comprehensive review. *Appl. Mater. Today* **29**, 101582 (2022).
3. Hofer, M. & Lutolf, M. P. Engineering organoids. *Nat. Rev. Mater.* **2021** *65* **6**, 402–420 (2021).
4. Drakhlis, L. *et al.* Human heart-forming organoids recapitulate early heart and foregut development. *Nat. Biotechnol.* **2021** *396* **39**, 737–746 (2021).
5. Lewis-Israeli, Y. R. *et al.* Self-assembling human heart organoids for the modeling of cardiac development and congenital heart disease. *Nat. Commun.* **2021** *121* **12**, 1–16 (2021).
6. Lee, J. *et al.* In vitro generation of functional murine heart organoids via FGF4 and extracellular matrix. *Nat. Commun.* **2020** *111* **11**, 1–18 (2020).
7. Lee, S. G. *et al.* Generation of human iPSCs derived heart organoids structurally and functionally similar to heart. *Biomaterials* **290**, (2022).
8. Hoang, P., Wang, J., Conklin, B. R., Healy, K. E. & Ma, Z. Generation of spatial-patterned early-developing cardiac organoids using human pluripotent stem cells. *Nat. Protoc.* **2018** *134* **13**, 723–737 (2018).
9. Patel, R. B. *et al.* Characterization of cardiac mechanics and incident atrial fibrillation in participants of the Cardiovascular Health Study. *JCI Insight* **5**, (2020).
10. Ribeiro, A. H. *et al.* Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat. Commun.* **2020** *111* **11**, 1–9 (2020).
11. Alfaras, M., Soriano, M. C. & Ortín, S. A Fast Machine Learning Model for ECG-Based Heartbeat Classification and Arrhythmia Detection. *Front. Phys.* **7**, 472268 (2019).
12. Siontis, K. C., Noseworthy, P. A., Attia, Z. I. & Friedman, P. A. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat. Rev. Cardiol.* **18**, 465–478 (2021).
13. Ebrahimi, Z., Loni, M., Daneshtalab, M. & Gharehbaghi, A. A review on deep learning methods for ECG arrhythmia classification. *Expert Syst. with Appl.* **X** **7**, 100033 (2020).
14. Juhola, M., Penttinen, K., Joutsijoki, H. & Aalto-Setälä, K. Analysis of Drug Effects on iPSC Cardiomyocytes with Machine Learning. *Ann. Biomed. Eng.* **49**, (2021).
15. Lee, E. K. *et al.* Machine Learning of Human Pluripotent Stem Cell-Derived Engineered Cardiac Tissue Contractility for Automated Drug Classification. *Stem Cell Reports* **9**, 1560–1572 (2017).
16. Kowalczewski, A. *et al.* Integrating nonlinear analysis and machine learning for human induced pluripotent stem cell-based drug cardiotoxicity testing. *J. Tissue Eng. Regen. Med.* (2022) doi:10.1002/TERM.3325.
17. Serrano, R. *et al.* A deep learning platform to assess drug proarrhythmia risk. *Cell Stem Cell* **30**, 86-95.e4 (2023).
18. Kelly, C. N., Miller, A. T., Hollister, S. J., Guldberg, R. E. & Gall, K. Design and Structure–Function Characterization of 3D Printed Synthetic Porous Biomaterials for Tissue Engineering. *Adv. Healthc. Mater.* **7**, 1701095 (2018).

19. Collins, M. N. *et al.* Scaffold Fabrication Technologies and Structure/Function Properties in Bone Tissue Engineering. *Adv. Funct. Mater.* **31**, 2010609 (2021).
20. Ma, Z. *et al.* Self-organizing human cardiac microchambers mediated by geometric confinement. *Nat. Commun.* **2015 61** **6**, 1–10 (2015).
21. Van Der Maaten, L. & Hinton, G. *Visualizing Data using t-SNE. Journal of Machine Learning Research* vol. 9 <http://jmlr.org/papers/v9/vandermaaten08a.html> (2008).
22. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. (2018).
23. Eline, C. ' *et al.* Control of Human Embryonic Stem Cell Colony and Aggregate Size Heterogeneity Influences Differentiation Trajectories. *Stem Cells* **26**, 2300–2310 (2008).
24. Bauwens, C. L. *et al.* Geometric Control of Cardiomyogenic Induction in Human Pluripotent Stem Cells. <https://home.liebertpub.com/tea> **17**, 1901–1909 (2011).
25. Bosch-Fortea, M. *et al.* Micropattern-based platform as a physiologically relevant model to study epithelial morphogenesis and nephrotoxicity. *Biomaterials* **218**, 119339 (2019).
26. Théry, M. Micropatterning as a tool to decipher cell morphogenesis and functions. *J. Cell Sci.* **123**, 4201–4213 (2010).
27. Kilian, K. A., Bugarija, B., Lahn, B. T. & Mrksich, M. Geometric cues for directing the differentiation of mesenchymal stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 4872–4877 (2010).
28. Song, W., Lu, H., Kawazoe, N. & Chen, G. Adipogenic differentiation of individual mesenchymal stem cell on different geometric micropatterns. *Langmuir* **27**, 6155–6162 (2011).
29. Warmflash, A., Sorre, B., Etoc, F., Siggia, E. D. & Brivanlou, A. H. A method to recapitulate early embryonic spatial patterning in human embryonic stem cells. *Nat. Methods* **2014 118** **11**, 847–854 (2014).
30. Deglincerti, A. *et al.* Self-organization of human embryonic stem cells on micropatterns. *Nat. Protoc.* **2016 1111** **11**, 2223–2232 (2016).
31. Hoang, P. *et al.* Engineering spatial-organized cardiac organoids for developmental toxicity testing. *Stem Cell Reports* **16**, 1228–1244 (2021).
32. Bidan, C. M. *et al.* Geometry as a Factor for Tissue Growth: Towards Shape Optimization of Tissue Engineering Scaffolds. *Adv. Healthc. Mater.* **2**, 186–194 (2013).
33. Kim, J., Mailand, E., Ang, I., Sakar, M. S. & Bouklas, N. A model for 3D deformation and reconstruction of contractile microtissues. *Soft Matter* **17**, 10198–10209 (2021).
34. Legant, W. R. *et al.* Microfabricated tissue gauges to measure and manipulate forces from 3D microtissues. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 10097–10102 (2009).
35. Liu, A. S. *et al.* Matrix viscoplasticity and its shielding by active mechanics in microtissue models: experiments and mathematical modeling. *Sci. Reports* **2016 61** **6**, 1–10 (2016).
36. Bose, P., Eyckmans, J., Nguyen, T. D., Chen, C. S. & Reich, D. H. Effects of Geometry on the Mechanics and Alignment of Three-Dimensional Engineered Microtissues. *ACS Biomater. Sci. Eng.* **5**, 3843–3855 (2019).
37. Wang, H. *et al.* Necking and failure of constrained 3D microtissues induced by cellular tension. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 20923–20928 (2013).

38. Winston, T. S. *et al.* Controlling Mesenchyme Tissue Remodeling via Spatial Arrangement of Mechanical Constraints. *Front. Bioeng. Biotechnol.* **10**, 833595 (2022).
39. Bian, X. *et al.* A deep learning model for detection and tracking in high-throughput images of organoid. *Comput. Biol. Med.* **134**, 104490 (2021).
40. Spiller, E. R. *et al.* Imaging-Based Machine Learning Analysis of Patient-Derived Tumor Organoid Drug Response. *Front. Oncol.* **11**, 771173 (2021).
41. Gritti, N. *et al.* Morgana: Accessible quantitative analysis of Organoids with machine learning. *Dev.* **148**, (2021).
42. Matthews, J. M. *et al.* OrganoID: A versatile deep learning platform for tracking and analysis of single-organoid dynamics. *PLOS Comput. Biol.* **18**, e1010584 (2022).
43. Abdul, L. *et al.* D-CryptO: deep learning-based analysis of colon organoid morphology from brightfield images. *Lab Chip* **22**, 4118–4128 (2022).
44. Lee, E. K., Kurokawa, Y. K., Tu, R., George, S. C. & Khine, M. Machine learning plus optical flow: a simple and sensitive method to detect cardioactive drugs. *Sci. Reports 2015 51* **5**, 1–12 (2015).
45. Heylman, C., Datta, R., Sobrino, A., George, S. & Gratton, E. Supervised Machine Learning for Classification of the Electrophysiological Effects of Chronotropic Drugs on Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes. *PLoS One* **10**, e0144572 (2015).
46. Juhola, M., Joutsijoki, H., Penttinen, K. & Aalto-Setälä, K. Detection of genetic cardiac diseases by Ca²⁺ transient profiles using machine learning methods. *Sci. Reports 2018 81* **8**, 1–10 (2018).
47. Lee, Y. K. *et al.* Calcium Homeostasis in Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes. *Stem Cell Rev. Reports* **7**, 976–986 (2011).
48. Fong, A. H. *et al.* Three-Dimensional Adult Cardiac Extracellular Matrix Promotes Maturation of Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes. <https://home.liebertpub.com/tea> **22**, 1016–1025 (2016).
49. Guo, J. *et al.* Elastomer-grafted iPSC-derived micro heart muscles to investigate effects of mechanical loading on physiology. *ACS Biomater. Sci. Eng.* **7**, 2973–2989 (2021).
50. Lian, X. *et al.* Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/ β -catenin signaling under fully defined conditions. *Nat. Protoc.* (2013) doi:10.1038/nprot.2012.150.
51. Lian, X. *et al.* Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proc. Natl. Acad. Sci.* **109**, E1848–E1857 (2012).
52. Morris, T. A. *et al.* Striated myocyte structural integrity: Automated analysis of sarcomeric z-discs. *PLOS Comput. Biol.* **16**, e1007676 (2020).
53. Huebsch, N. *et al.* Automated video-based analysis of contractility and calcium flux in human-induced pluripotent stem cell-derived cardiomyocytes cultured over different spatial scales. *Tissue Eng. - Part C Methods* (2015) doi:10.1089/ten.tec.2014.0283.
54. Team, R. C. R: A Language and environment for statistical computing. (2017).
55. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
56. Krijthe, J. H. Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation,. (2015).

57. Melville, J. uwot: The Uniform Manifold Approximation and Projection (UMAP) Method for Dimensionality Reduction. R package version 0.1.11. (2021).
58. Van Rossum, G. & Drake, F. L. Python 3 Reference Manual; CreateSpace. *Scotts Val. CA* 242 (2009).
59. Amid, E. & Warmuth, M. K. TriMap: Large-scale Dimensionality Reduction Using Triplets. (2019).
60. Wang, Y., Huang, H., Rudin, C. & Shaposhnik, Y. Understanding How Dimension Reduction Tools Work: An Empirical Approach to Deciphering t-SNE, UMAP, TriMAP, and PaCMAP for Data Visualization. (2020).
61. Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B. & Heming, J. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Inf. Sci. (Ny)*. **622**, 178–210 (2023).
62. Hartigan, J. A. & Wong, M. A. A K-Means Clustering Algorithm. *J. R. Stat. Soc. Ser. C (Applied Stat)*. **28**, 100–108 (1979).
63. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
64. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* **13-17-August-2016**, 785–794 (2016).
65. Pedregosa FABIANPEDREGOSA, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
66. Csárdi, G. & Nepusz, T. The igraph software package for complex network research. (2006).
67. Clauset, A., Newman, M. E. J. & Moore, C. Finding community structure in very large networks. *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.* **70**, 6 (2004).