Supplementary Information Methods for:

Transfer learning enables predictions in network biology

Christina V. Theodoris*[1-4], Ling Xiao[2,5], Anant Chopra[6], Mark D. Chaffin[2], Zeina R. Al Sayed[2], Matthew C. Hill[2,5], Helene Mantineo[2,5], Elizabeth M. Brydon[6], Zexian Zeng[1,7], X. Shirley Liu[1,7,8], Patrick T. Ellinor*[2,5]

Affiliations:
[1] Department of Data Science, Dana-Farber Cancer Institute, Boston MA, USA.
[2] Cardiovascular Disease Initiative and Precision Cardiology Laboratory, Broad Institute of MIT and Harvard, Cambridge, MA, USA.
[3] Division of Genetics and Genomics, Boston Children's Hospital, Boston MA, USA.
[4] Harvard Medical School Genetics Training Program, Boston, USA.
[5] Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA.
[6] Precision Cardiology Laboratory, Bayer US LLC, Cambridge, MA, USA.
[7] Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA.
[8] Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA, USA.
* Co-corresponding authors

Correspondence to: christina.theodoris@gladstone.ucsf.edu, ellinor@mgh.harvard.edu

Supplementary Information Methods

Robustness of gene embeddings to batch-dependent technical artifacts

To quantify the impact of common batch-dependent technical artifacts on Geneformer gene embeddings, we compared 1) the cosine similarity of embeddings from two randomly selected genes from the same cell (which we expect to have low cosine similarity), 2) the cosine similarity of embeddings from the same gene from two different cells of the same cell type from the same batch (which we expect to have high cosine similarity), and 3) the cosine similarity of embeddings from the same gene from two different cells of the same cell type from different batches (which we expect to have high cosine similarity if the gene embeddings are robust to batch-dependent technical artifacts). We performed the above procedure to quantify A) platform-related effects using 500 cells iPSCs assayed in parallel on the Drop-seq (single-cell) or DroNc-seq (single-nucleus) platform[1], B) preservation-related effects using 330 fresh vs. frozen natural killer (NK) cells from the same donor[2,3], and C) individual patient variability-related effects using 344 aortic ECs cells from different control patients[4]. Distributions were statistically compared by Wilcoxon rank sums with Bonferroni correction.

Of note, although we found that Geneformer was robust to the common batch-dependent technical artifacts that we tested, Geneformer was not designed specifically for scRNA-seq batch integration. As such, users may elect to preprocess their limited task-specific data with alternative batch integration methods prior to using that data for fine-tuning Geneformer towards their downstream task if they find their dataset to be persistently affected by such batch-dependent artifacts.

Robustness of cell embeddings to batch-dependent technical artifacts

Geneformer cell embeddings were visualized in two dimensions with scanpy's[5] implementation of Uniform Manifold Approximation and Projection[6] (UMAP). To contrast the robustness to batch-dependent technical artifacts compared to the original data, this UMAP of the Geneformer cell embeddings was compared to the UMAP of the original data generated according to the procedures outlined in the scanpy clustering tutorial, with or without normalization by ComBat[7] or Harmony[8] methods as indicated. All cells were utilized for the UMAPs of the aortic dataset[4]; UMAPs for the iPSC to cardiomyocyte differentiation dataset[1] were generated by randomly sampling 500 cells for memory efficiency.

All Drop-seq cells without downsampling were used to train the cell type classification fine-tuning application, and all DroNC-seq cells were used to evaluate this trained model. The DroNc-seq (single-nucleus) platform detected significantly fewer genes than the Drop-seq (single-cell) platform ($p<0.05$, Wilcoxon rank sum test) in cells undergoing iPSC to cardiomyocyte differentiation[1]. Additionally, cardiomyocyte type 1 was significantly overrepresented in the Drop-seq data compared to the DroNC-seq data, where cardiomyocyte type 2 was significantly overrepresented ($p<0.05$, X^2 test) (cell types from annotations by original authors). This may indicate that the distinction between these cardiomyocyte types is a technical artifact. Of note, the fine-tuned cell type classifier was least confident in distinguishing between these potentially artifactual two cardiomyocyte subtypes, and most inaccuracies were due to preferentially predicting the cardiomyocyte type 1, which was overrepresented in the Drop-seq training data and underrepresented in the DroNC-seq evaluation data.

Context-dependence of gene embeddings

Context dependence of gene embeddings was quantified by measuring standard error of embeddings of the same gene across variable cell types (fibroblasts, mesenchymal stem cells (MSCs), T cells, NK cells, monocyte/macrophage/dendritic cells (MonoMaphDC), smooth muscle cell (SMC) 1 and 2) within aortic tissue from control patients[4]. *GAPDH*, a known housekeeping gene, was used as a control gene expected to have low context-dependence compared to known highly context-dependent NOTCH receptors 1, 2, and 3, ligand DLL1, and RBPJ. Context dependence was also quantified by measuring the effect on the embeddings of the remaining genes in the transcriptome in response to in silico reprogramming fibroblasts[9] by artificially adding *OCT4*, *SOX2*, *KLF4*, and *MYC* (*OSKM*) to the front of their rank value encodings. Gene embeddings were shifted from their original fibroblast state to be more cosine similar to that gene's embedding in the iPSC state[5,9]. We also quantified context dependence of gene embeddings in response to in silico myoblast differentiation. Guo et al. *eLife* 2022 reported generation of iPSC-derived myoblasts that pass through an initial PAX3+/MYOD- stage (S1) towards the PAX3+/MYOD+ myogenic state (S2B)[10]. We tested in silico differentiation via artificially adding *MYOD* to the front of the rank value encodings of PAX3+/MYOD- (S1) cells and found the embeddings of the remaining genes significantly shifted towards the MYOD+ (S2B) myogenic cell state as measured by cosine similarity.

Of note, in silico reprogramming/differentiation is only modeling the first step in the cell state transitions that occur in response to addition of the reprogramming/differentiation factors as the remainder of the gene context within the rank value encoding is equivalent to the original cell aside from the addition of these factors (e.g. *OSKM* or *MYOD*). The relevant measurement that is quantified in this approach is the directionality of the shift that occurs in response to addition of these factors. For example, while addition of random genes shifts the remaining genes' embeddings in a random direction, addition of *OSKM* significantly shifts the remaining genes' embeddings towards their embeddings within the context of the iPSC state, indicating that the directionality of the predicted shift is concordant with the true biological cell state transition from fibroblast to iPSC in response to the experimental addition of OSKM.

Geneformer attention weight analysis

When examining the mean attention weights of each gene in aortic ECs from control patients[4], we found that 20% of attention heads significantly attended transcription factors[11] more than other genes ($p<0.05$, Wilcoxon rank sum test, Benjamini-Hochberg (BH)-corrected), indicating that specific attention heads learned, in an entirely self-supervised manner, the relative importance of transcription factors in distinguishing cell states. We also found that specific attention heads attended the top 50 most central genes significantly more than the 50 most peripheral genes in the N1-dependent network[12] ($p<0.05$, Wilcoxon rank sum test, BH-corrected). Concordantly, these centrality-driven attention heads consistently attended to a significantly greater degree the highest ranked genes in each cell's unique rank value encoding in aortic ECs, smooth muscle cells, T cells, and macrophage/monocyte/dendritic cells from control patients[4] ($p<0.05$, Wilcoxon rank sum test, BH-corrected). Only cells with a full 2048 input size were included for this analysis so that ranks would be comparable. Interestingly, attention heads in the earliest layers were consistently the most diverse in terms of gene ranks they attended, suggesting the model initially orients to the observed cell state through a joint

survey of distinct portions of the input space. The middle layers were most broad in terms of gene ranks they attended, and the final layers focused on the highest ranked genes that uniquely define each cell state.

In silico perturbation analysis

We tested the impact on fetal cardiomyocytes from the Fetal Cell Atlas[13] of in silico deleting genes known to be involved in cardiomyopathy, structural heart disease, and hyperlipidemia from the relevant clinical genetic testing panels from Blueprint Genetics. Genes in the hyperlipidemia panel that were also present in the cardiomyopathy or structural heart disease panels were excluded from the hyperlipidemia analysis. Impact of in silico deletion was correlated with that gene's rank within the rank value encoding, which is expected given genes that uniquely distinguish cell state are encoded at higher ranks. However, impact of in silico deletion was not correlated with the number of detections of that gene, indicating single cell RNA-seq dropout rates were not significantly impacting the predicted deleterious effect of in silico deletion. Gene set enrichment of the top 500 genes whose in silico deletion in fetal cardiomyocytes was predicted to have the most deleterious effect was determined using ToppFun with a cut-off of 0.05 for p-values adjusted for multiple hypotheses.

We tested the impact on fetal cardiomyocytes from the Fetal Cell Atlas[13] of in silico deleting *GATA4*, a known congenital heart disease gene, on housekeeping genes[14] compared to genes known to be significantly dysregulated by *GATA4* variants in a previously reported iPSC disease model of *GATA4*-related heart defects[15]. Target confidence was defined as false discovery rate (FDR) significance reported in the original study[15] stratified into four bins (from least to most confident: <500, 500-1000, 1000-1500, >1500). Cosine similarity of gene embeddings with or without in silico deletion of *GATA4* was compared between housekeeping genes[14] and GATA4 targets in each of the aforementioned bins by Wilcoxon rank sums test with BH correction. We tested the impact of in silico deletion of *GATA4* by the same method on GATA4 direct versus indirect target genes and NKX2-5 and NOTCH1 direct target genes. GATA4 direct target genes were defined as any gene significantly dysregulated in response to the *GATA4* variant where GATA4 bound within 20kb of the gene's TSS by ChIP-seq in the iPSC disease model[15], whereas indirect target genes were genes significantly dysregulated in response to the *GATA4* variant without GATA4 binding within 20kb of their TSS. NKX2-5 direct target genes were defined as genes significantly activated and bound by NKX2-5 in analysis of *NKX2-5*$^{-/-}$ vs. *NKX2-5*$^{+/+}$ human embryonic stem cell-derived cardiomyocytes by bulk RNA-seq and endogenous NKX2-5 ChIP-seq[16], removing those genes that were also GATA4 targets as defined above. NOTCH1 direct target genes were defined as genes significantly dysregulated in *NOTCH1*$^{+/-}$ vs. *NOTCH1*$^{+/+}$ iPSC-derived ECs and bound by NOTCH1 in endogenous NOTCH1 ChIP-seq as previously reported[12].

Similarly, we tested the impact on fetal cardiomyocytes from the Fetal Cell Atlas[13] of in silico deleting *TBX5*, a known congenital heart disease gene, on housekeeping genes[14] compared to known TBX5 direct and indirect target genes as previously reported in an iPSC disease model of *TBX5*-related heart defects[17]. TBX5 targets were defined as those significantly dysregulated in *TBX5*$^{-/-}$ vs. *TBX5*$^{+/+}$ human iPSC-derived cardiomyocytes by single cell RNA-seq; direct targets were also bound by TBX5 by ChIP-seq whereas indirect targets were not[17].

Then, we tested the effects of in silico deletion of *GATA4*, *TBX5*, or *GATA4* + *TBX5* on housekeeping genes[14] compared to genes significantly dysregulated in response to the *GATA4* variant that had both GATA4 and TBX5 binding by ChIP-seq within 20kb of their TSS in the iPSC disease model[15]. Because the *GATA4* variants studied in this model[15] disrupt its co-binding with TBX5, we compared the sum of the effect of individual deletion of *GATA4* or *TBX5* to the effect of combination deletion of *GATA4* and *TBX5* to determine the model's interpretation of cooperativity.

Alternative models

Geneformer fine-tuning performance for gene classification was compared to common alternative modeling methods from sklearn including support vector machines (SVM), random forest, and logistic regression trained with input features being either the ranks from the rank value encoding or transcript counts (normalized by total transcripts within that cell) of each labeled gene across the task-specific dataset as indicated in Supplementary Table 2. Default sklearn parameters were utilized aside from max depth of 2 for the random forest models. Performance was quantified as AUC +/- standard deviation and F1 score calculated based on a 5-fold cross-validation strategy as described in the Methods section *Geneformer fine-tuning*.

For the dosage sensitive vs. insensitive fine-tuning application, we also compared to non-pretrained models with the same architecture as Geneformer (described in the Methods section *Geneformer architecture and pretraining*) or smaller versions of Geneformer with retained width-to-aspect ratio (as prior research[18] indicates that the optimal width-to-aspect ratio is application-specific). The four layer model had 256 embedding dimensions, 4 attention heads per layer, and feed forward size of 512; the three layer model had 128 embedding dimensions, 2 attention heads per layer, and a feed forward size of 256; and the one layer model had 64 embedding dimensions, 2 attention heads per layer, and a feed forward size of 128. Fine-tuning of these non-pretrained models was accomplished by initializing the model with random weights (for the six layer model, we used the same initialization weights as Geneformer prior to pretraining), adding a final task-specific transformer layer, and fine-tuning all layers with the task-specific data as indicated in Supplementary Table 2 (no layers were frozen given the weights were randomly initialized).

We also quantified the impact of pretraining with smaller and less diverse pretraining corpuses than Genecorpus-30M. A six layer model with the same architecture as Geneformer was initialized with the same weights as Geneformer prior to pretraining and then pretrained with each of the downgraded corpuses. Smaller corpuses with retained diversity were generated by randomly downsampling Genecorpus-30M to 1 million, 100,000, or 10,000 cells. Corpuses with reduced diversity were generated from an esophagus tissue dataset[19] using either the total ~850K cells or randomly downsampling to 100,000 or 10,000 cells. Rank value encodings for the esophagus corpus were generated using gene transcript count distributions aggregated only from that dataset. These models pretrained with downgraded corpuses were all fine-tuned identically to Geneformer for the dosage sensitive vs. insensitive fine-tuning application as described in the Methods section *Geneformer fine-tuning* with performance quantified via AUC calculated based on a 5-fold cross-validation strategy.

Cell type annotation fine-tuning application

Although Geneformer is most focused on understanding network dynamics rather than cell-level annotations, we investigated Geneformer's performance in cell type annotation given it is a common application for previously published models. We compared Geneformer to alternative XGBoost (CaSTLe[20]) and deep neural network-based (scDeepSort[21]) models. These methods train a new model from scratch for each separate tissue using the same supervised learning objective as is used for the final cell type predictions in that specific tissue. Therefore, these approaches do not take advantage of the large amounts of data available more broadly that are not specifically labeled for that task. Additionally, their learning objective is not meant to gain widely generalizable knowledge during the initial training, and they do not transfer knowledge to new tasks.

Labeled training data for cell type annotation was provided by the authors of scDeepSort ("human_cell_atlas.7z" at https://github.com/ZJUFanLab/scDeepSort/releases). Data from a diverse set of tissues (placenta, brain, spleen, blood, liver, kidney, pancreas, large intestine, and lung) with varying numbers of cell type classes were used for method comparisons. Both adult and fetal tissues were used within each category. Large intestine samples included those labeled as fetal intestine, ascending colon, transverse colon, sigmoid colon, and rectum; and blood samples included those labeled as peripheral blood and bone marrow. Following scDeepSort's method, cells numbering more than 5% of the total cells in each tissue were included. Data from the provided samples was shuffled and randomly divided into training and evaluation data at a ratio of 80:20. All cell types present in the evaluation data were represented within the training data.

The pretrained Geneformer was fine-tuned with zero frozen layers for 10 epochs, otherwise using the learning hyperparameters described in the Methods section *Geneformer fine-tuning*. scDeepSort was trained for each tissue using the provided default parameters for DeepSortClassifier except for setting the validation fraction to 0 so that the full training data was used. The default number of epochs for scDeepSort is 300. CaSTLe was trained for each tissue using the provided methodology except that the feature selection was performed only using the training data to avoid data leakage from the evaluation dataset (the default methodology uses both the training data and evaluation data for feature selection). After features are selected, CaSTLe expands them based on expression level bins to generate a one-hot encoding of the input data. Because this requires the same training features as will be used for inference, CaSTLe uses both the training and evaluation data to expand the features based on expression level bins. We used CaSTLe's provided methodology for this feature expansion, but it is important to note that this method limits the use of the trained model to make predictions in data that was not available at the time of training as it is not possible to add new expression level bins for the selected features after the initial training.

Predictive performance was evaluated using accuracy and macro F1 score, which averages the F1 score for each of the classes such that each class is equally weighted for multiclass predictions.

Disease modeling

We fine-tuned Geneformer using single-nuclei transcriptomic data to distinguish cardiomyocytes from non-failing hearts or hearts affected by hypertrophic or dilated cardiomyopathy using

single-nuclei transcriptomic data from affected patients. Samples were randomly assigned as training or out-of-sample data by patient so that no single cells from the out-of-sample data were used for training. We only included non-failing heart samples that had a documented normal ejection fraction. Training data included 93,589 cardiomyocytes (non-failing n=9, hypertrophic n=11, dilated n=9); out-of-sample data included 39,006 cardiomyocytes (non-failing n=4, hypertrophic n=4, dilated n=2). Hyperparameters (max learning rate, learning scheduler, warmup steps, weight decay, seed, frozen layers) were tuned using the RayTune[22] implementation HyperOpt[23], which uses a Tree-structured Parzen Estimators algorithm for parallel optimization over the search space. Tuning was distributed across 3 nodes each with 4 Nvidia A100 40GB GPUs (total 12 GPUs). The final hyperparameters utilized for training were: max learning rate: 0.000804; learning scheduler: polynomial with warmup; optimizer: Adam with weight decay fix[24]; warmup steps: 1812.678558; weight decay: 0.258828; seed: 73.152431, batch size: 12, frozen layers: 2, epoch 0.9.

Predictive accuracy, precision, and recall were then quantified on the out-of-sample data as defined above, using a straight average of the predictions on all classes without weighting by occurrence. Having established the model had a high predictive accuracy (90%) on out-of-sample data, we defined the non-failing versus hypertrophic or dilated cardiomyopathy states as the mean of the embeddings of cardiomyocytes from each of those conditions from the training data. We then investigated the genes that defined each state by performing in silico deletion or activation of each gene expressed within the transcriptome of each cardiomyocyte from the non-failing hearts or hearts affected by hypertrophic or dilated cardiomyopathy from the training data.

We first determined genes whose in silico deletion or activation in non-failing cardiomyocytes significantly shifted the non-failing embeddings towards the hypertrophic or dilated cardiomyopathy state in the 256-dimensional embedding space. For the in silico deletion analysis, we first delete random genes in each non-failing cardiomyocyte and determine where in the embedding space this deletion pushes the cell embedding, which defines the random distribution. Then, we determine genes whose deletion in every cardiomyocyte in which they are expressed statistically significantly shifts the embeddings towards the average hypertrophic or dilated cardiomyopathy position compared to the random distribution ($p<0.05$, Wilcoxon rank sum test, BH-corrected) (Fig. 6a). For the in silico activation analysis, we perform an analogous procedure except that instead of deleting each gene, we move the gene to the front of the rank value encoding to model overexpression of that gene.

We then performed in silico treatment analysis in cardiomyocytes from hypertrophic cardiomyopathy patients to identify genes whose in silico deletion or activation in hypertrophic cardiomyocytes significantly shifted cell embeddings back towards the average non-failing embedding position compared to the distribution of deleting or activating random genes in hypertrophic cardiomyocytes ($p<0.05$, Wilcoxon rank sum test, BH-corrected) (Fig. 6a). We only considered genes that both shifted cell embeddings towards the non-failing state while also shifting embeddings away from the average dilated cardiomyopathy embedding position. We then performed the analogous in silico treatment analysis in cardiomyocytes from dilated cardiomyopathy patients to determine genes whose in silico deletion or activation significantly shifted cell embeddings towards the non-failing state while also shifting embeddings away from the average hypertrophic cardiomyopathy embedding position.

Gene set enrichment was determined using the Gseapy implementation of Enrichr[25] with a cut-off of 0.05 for p-values adjusted for multiple hypotheses. Gene and cell embeddings from the last layer of the model were used for disease modeling given the model was fine-tuned for the relevant objective.

Visualization

Seaborn[26], Matplotlib[27], and scanpy[5] were used to generate data visualizations.

Cardiomyopathy single nuclei RNA-seq sample collection and preprocessing

Human tissue samples

Adult human myocardial samples of European ancestry were collected from organ donors by the Myocardial Applied Genetics Network as previously described[28]. Non-failing samples were obtained from organ donors with no history of heart failure and hypertrophic or dilated cardiomyopathy samples were obtained from explanted hearts of donors receiving a heart transplant. Transmural cardiac tissue was collected from the left ventricular free wall of a non-infarcted region excluding the septum (predominantly representing the anterior wall midway between the apex and base). Written informed consent for research use of donated tissue was obtained from next of kin in all cases. Research use of tissues was approved by the relevant institutional review boards at the Gift-of-Life Donor Program, the University of Pennsylvania, Massachusetts General Hospital, and the Broad Institute.

Single nuclei RNA-seq and data processing

Single nucleus suspensions were generated as previously described[28]. Cells were loaded into the 10x Genomics microfluidic platform (Single cell 3' solution, v3) for an estimated recovery of 5000 cells per device. Libraries were generated according to manufacturer protocols with modifications as previously described[29]. Libraries were multiplexed at an average of 4 libraries per flow cell on an Illumina Nextseq550 in the Broad Institute's Genomics Platform. Single nuclei RNA-seq data were processed with quality control, alignment, cell type identification, and differential expression analysis as previously described[29].

Experimental testing of Geneformer-predicted targets in engineered cardiac microtissues

We tested whether targeting *TEAD4*, a top Geneformer-predicted dosage-sensitive gene in cardiomyocytes, would damage the ability of *WT* iPSC-derived cardiomyocytes to generate contractile stress. We also tested whether targeting Geneformer-predicted candidate therapeutic candidates for dilated cardiomyopathy would restore normal levels of contractile stress in engineered cardiac microtissues generated from iPSC-derived cardiomyocytes harboring a *TTN* truncating mutation in the A-band known to cause dilated cardiomyopathy in humans. Experiments measured contractile force and stress (force per unit area) in engineered cardiac microtissues derived from $TTN^{+/+}$ iPSCs exposed to empty lentivirus, isogenic CRISPR-edited $TTN^{+/-}$ iPSCs ($cN22577fs^{+/-}$ as previously described[30]) exposed to empty lentivirus or lentivirus expressing CRISPR-CAS9 and guides with a control target sequence (AGAACCGACCCGGCAATCCG), $TTN^{+/+}$ iPSCs exposed to lentivirus expressing CRISPR-CAS9 and guides with a *TEAD4* (TTGCCCGCTACATCAAGCTC) target sequence, or $TTN^{+/-}$ iPSCs exposed to lentivirus expressing CRISPR-CAS9 and guides with

a *PLN* (TGCTTGTTGAGGCATTTCAA), *GSN* (TGCAGTATGACCTCCACTAC), *ESRRG* (TGAT CCTTGGTGTCGTATAC), or *HMGB1* (ATTTGAAGATATGGCAAAAG) target sequence for knockout of the respective genes by CAS9 co-expressed in each iPSC line.

For CRISPR-mediated knockout of target genes, lentiviral vectors (pLentiCRISPR_V2) carrying CRISPR-CAS9 and single guide RNAs targeting the aforementioned genes were purchased from GenScript and were packaged into lentiviruses with Dharmacon Trans-Lentiviral packaging kits (Horizon, TLP5913). Lentiviruses were transduced into iPSC-derived ventricular cardiomyocytes and transduced cells were purified with 5-day puromycin (1 μg/ml) selection on day 3 post-transduction. After 14 days, cells were dissociated for generating engineered cardiac microtissues as described below. CRISPR-mediated knockout was quantified by QPCR in the same batch of cells used for generating engineered cardiac microtissues for each condition (with 3 technical replicates). Total RNA was extracted with the QIAGEN RNeasy Mini Kit (74106) according to the manufacturer's instructions. cDNA was synthesized with the iScript kit (Bio-Rad 1708891). qPCR was performed with the SsoAdvanced Universal probes supermix (Bio-Rad 1725284) on a Bio-rad CFX384 Real-time system with TaqMan probes (Life Technologies 4331182) of *TEAD4* (Hs01125032_m1), *PLN* (Hs01848144_s1), *GSN* (Hs00609272_m1), *ESRRG* (Hs00976243_m1), or *HMGB1* (Hs01923466_g1).

Cardiac microtissues were generated as previously described[31]. Briefly, PDMS microfabricated tissue gauges (μTUG) substrates were molded from the SU-8 masters with embedding fluorescent microbeads (Fluoresbrite 17147; Polysciences, Inc.) onto the cantilevers. μTUG substrates were treated with 0.2% pluronic F127 for 30 minutes. iPSC-CMs were dissociated after trypsin digestion and mixed with stromal cells (human ventricular fibroblasts (FBs)) to enable tissue compaction. The amount of stromal cells added was 6% of the total myocyte population. A suspension of ~1,300,000 cells within reconstitution mixture, consisting of 2mg/mL liquid neutralized collagen I (BD Biosciences) and 0.5mg/mL human fibrinogen (Sigma-Aldrich), was added to the substrate. The device was centrifuged to drive the cells into the micropatterned wells. Within 24 hours, FBs compacted the tissues. Force readouts were taken after 4 days in culture.

The displacement of fluorescent microbeads at the top of the cantilevers was then tracked using the SpotTracker plug-in in ImageJ (NIH). Displacement values were run through a custom MATLAB script to compute twitch force (dynamic tension), static force, and instantaneous velocity. Electrical field stimulation of biphasic square pulses of 1ms was given by placing two carbon electrodes (Ladd Research laboratories) separated by 2cms on the sides of the samples connected through platinum wires to a stimulator. Effects of each condition were quantified by contractile force and stress (force per unit area) generated by the engineered cardiac microtissues.

References

1. Selewa, A. *et al.* Systematic Comparison of High-throughput Single-Cell and Single-Nucleus Transcriptomes during Cardiomyocyte Differentiation. *Sci Rep* **10**, 1535 (2020).
2. 10X Genomics Datasets. *https://www.10xgenomics.com/resources/datasets/fresh-68-k-pbm-cs-donor-a-1-standard-1-1-0*.
3. 10x Genomics Datasets. *https://www.10xgenomics.com/resources/datasets/frozen-pbm-cs-donor-a-1-standard-1-1-0*.
4. Li, Y. *et al.* Single-Cell Transcriptome Analysis Reveals Dynamic Cell Populations and Differential Gene Expression Patterns in Control and Aneurysmal Human Aortic Tissue. *Circulation* **142**, 1374-1388 (2020).
5. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15 (2018).
6. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. (2018).
7. Zhang, Y., Parmigiani, G. & Johnson, W. E. ComBat-seq: Batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform* **2**, (2020).
8. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289-1296 (2019).
9. Xing, Q. R. *et al.* Diversification of reprogramming trajectories revealed by parallel single-cell transcriptome and chromatin accessibility sequencing. *Sci Adv* **6**, 463-474 (2020).
10. Guo, D. *et al.* iMyoblasts for ex vivo and in vivo investigations of human myogenesis and disease modeling. *Elife* **11**, e70341 (2022).
11. Lovering, R. C. *et al.* A GO catalogue of human DNA-binding transcription factors. *Biochim Biophys Acta Gene Regul Mech* **1864**, 194765 (2021).
12. Theodoris, C. V. *et al.* Human disease modeling reveals integrated transcriptional and epigenetic mechanisms of NOTCH1 haploinsufficiency. *Cell* **160**, 1072-1086 (2015).
13. Cao, J. *et al.* A human cell atlas of fetal gene expression. *Science* **370**, 808 (2020).
14. Eisenberg, E. & Levanon, E. Y. Human housekeeping genes, revisited. *Trends Genet* **29**, 569–574 (2013).
15. Ang, Y. S. *et al.* Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis. *Cell* **167**, 1734-1749 (2016).
16. Anderson, D. J. *et al.* NKX2-5 regulates human cardiomyogenesis via a HEY2 dependent transcriptional network. *Nat Commun* **9**, 1373 (2018).
17. Kathiriya, I. S. *et al.* Modeling Human TBX5 Haploinsufficiency Predicts Regulatory Networks for Congenital Heart Disease. *Dev Cell* **56**, 292-309 (2021).
18. Henighan, T. *et al.* Scaling Laws for Autoregressive Generative Modeling. *CoRR* **abs/2010.14701**, (2020).
19. Madissoon, E. *et al.* ScRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. *Genome Biol* **21**, 1 (2019).
20. Lieberman, Y., Rokach, L. & Shay, T. CaSTLe - Classification of single cells by transfer learning: Harnessing the power of publicly available single cell RNA sequencing experiments to annotate new experiments. *PLoS One* **13**, (2018).

21. Shao, X. *et al.* ScDeepSort: A pre-trained cell-type annotation method for single-cell transcriptomics using deep learning with a weighted graph neural network. *Nucleic Acids Res* **49**, e122 (2021).
22. Liaw, R. *et al.* Tune: A Research Platform for Distributed Model Selection and Training. (2018).
23. Bergstra, J., Yamins, D. & Cox, D. D. Making a Science of Model Search. 155-123 (2012).
24. Loshchilov, I. & Hutter, F. Decoupled Weight Decay Regularization. (2017).
25. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
26. Waskom, M. seaborn: statistical data visualization. *J Open Source Softw* **6**, 3021 (2021).
27. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput Sci Eng* **9**, 90–95 (2007).
28. Tucker, N. R. *et al.* Transcriptional and Cellular Diversity of the Human Heart. *Circulation* **142**, 466–482 (2020).
29. Chaffin, M. *et al.* Single-nucleus profiling of human dilated and hypertrophic cardiomyopathy. *Nature* **608**, 174–180 (2022).
30. Hinson, J. T. *et al.* Titin mutations in iPS cells define sarcomere insufficiency as a cause of dilated cardiomyopathy. *Science (1979)* **349**, 982–986 (2015).
31. Boudou, T. *et al.* A microfabricated platform to measure and manipulate the mechanics of engineered cardiac microtissues. *Tissue Eng Part A* **18**, 910-919 (2012).