A Deep Learning framework for the High Throughput/High Content Screening of Dilated Cardiomyopathy (DCM)

Nicolas Wiest-Daesslé[1]*, Eriona Heta[2], Angelica Roa[2], Kalina Radoynovska[1], Amandine Thibaudeau[1], Aurore Vuidel[1], Christina Jacob[1], Arnaud Ogier[1], Laurent Brino[1], Peter Sommer[1], Wolfram-Hubertus Zimmermann[2,3]* and Juyong Yoon[1]

1. *Ksilink, Strasbourg, France. *nicolas.wiestdaessle@ksilink.com 2. myriamed, Göttingen, Germany. *wolfram.zimmermann@myriamed.com*

3. Institute of Pharmacology and Toxicology, University Medical Center, Georg-August-University, Göttingen, Germany.

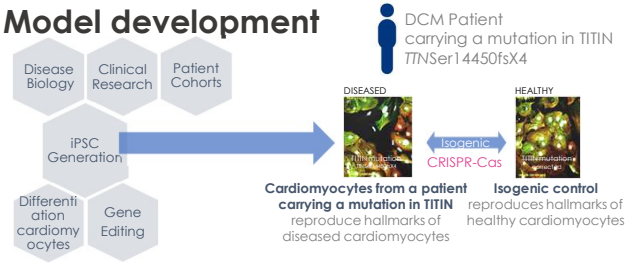
Introduction

With a prevalence of 4:10,000, dilated cardiomyopathy (DCM) is the most common hereditary heart disease, characterized by dilatation of one or both ventricles and impaired systolic and diastolic function. Numerous mutations have been identified to be causal for the development of DCM. TITIN-truncating variants account for up to 20% of the genetic DCM cases.

To recapitulate the cellular DCM-pathology in a cellular model compatible with high throughput image analysis, we generated isogenic iPSC lines with either introducing a TTN-truncating mutation (AT insertion in exon 326) in wild-type (WT → W1D) or correcting the mutated TTN from a patient (D → D1W) using CRISPR-Cas9 gene editing. These four cell lines were then functionally characterized by measurement of force of contraction (FOC) in Engineered Human Myocardium (EHM) along with assessing the calcium cycling by calcium flux assays.

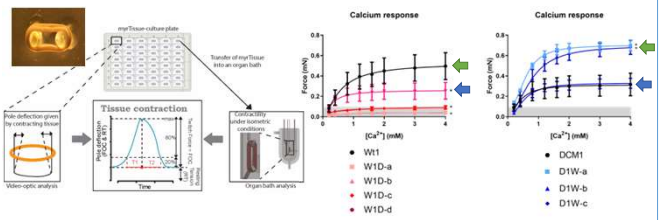
We then developed conditions enabling confocal fluorescent imaging of the cellular sarcomeric structures by immunostaining of alpha-actinin and cardiac troponin T.

Model development



DCM-patient derived				Healthy-donor derived				
DCM1 (Gramlich et al. 2015)	D1W (DCM1 TTN mutation corrected, WT-like)			WT1	W1D (engineered WT1 with TTN mutation, DCM-like)			
	D1W-a	D1W-b	D1W-c		W1D-a	W1D-b	W1D-c	W1D-d (homozygous)

3D-functional characterization (EHM)



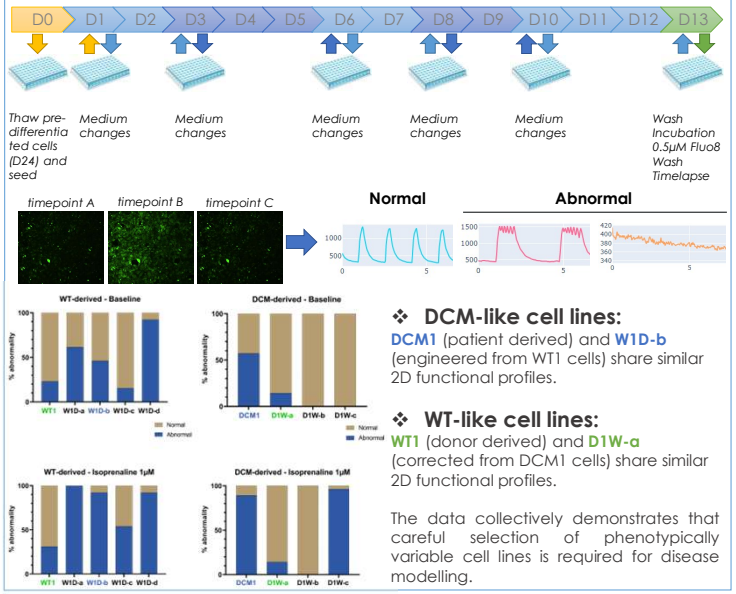
❖ DCM-like cell lines:

DCM1 (patient derived) and **W1D-b** (engineered from WT1 cells) share similar 3D functional profiles.

❖ WT-like cell lines:

WT1 (donor derived), **D1W-a** and **D1W-b** (corrected from DCM1 cells) share similar 3D functional profiles.

2D-functional characterization (Ca Imaging)



❖ DCM-like cell lines:

DCM1 (patient derived) and **W1D-b** (engineered from WT1 cells) share similar 2D functional profiles.

❖ WT-like cell lines:

WT1 (donor derived) and **D1W-a** (corrected from DCM1 cells) share similar 2D functional profiles.

The data collectively demonstrates that careful selection of phenotypically variable cell lines is required for disease modelling.

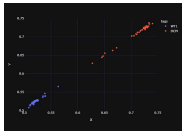
Deep Learning Model

The presented results are based on an Efficient Net B7 architecture, with minor adjustments, on the tasks.

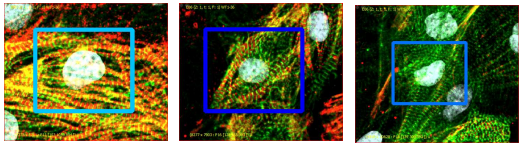
The Deep Learning tasks is split into two application targets:

a) a classification task to discriminate between cell lines,
- Activation: Softmax
- Loss: Cross Entropy

b) a 2D embedding task with selected landmarks for each cell types
- Activation: LogCosh
- Loss: Mean Squared Error

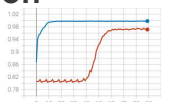


Inputs:
- small cropped areas of the acquired images (256x216),
- 3 channels Data Nuclei (Hoechst), a-actinin, t-Troponin,
- Crops are centered on nuclei
- Crops are selected with respect to the amount of expressed a-actinin and t-Troponin in the underlying area.



The training set consisted in 2000 cells per conditions picked from 5 wells out of 42 from a single plate. The remaining data from this plate is used for validation.

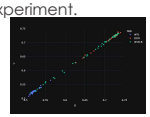
Deep Learning Training / Validation



The **training** occurred on 4 Tesla V100 GPUs
The convergence is obtained after 60 Epochs,
with a training accuracy of 99.8% and validation of 97.3%

Further testing was performed with other plates exhibiting high classification rates even with "unseen" data (i.e. from other plates and experimental noise).

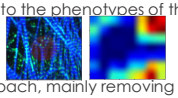
The **embedding**, was an effective game changer for the training experiment.



The observed effects, apart from a raise in validation score, are:
- A regularizing effect
- An analysis tool for engineered cells selection

We are still analyzing the use of the embedded space for distance analysis between samples in order to get a more understandable metric.
Other embedding strategies (i.e. knowledge driven and unsupervised) are under investigation, to better control the underlying spaces.

We are also investigating the neural networks saliency on the images, in order to grasp where the features come from, and how they are related to the phenotypes of the cells.
Tests were performed through GradCAM algorithm,
in order to compute heat maps from images.



Finally, we also performed an ablation studies of the approach, mainly removing the Nuclei channel from the CNN, in order to assess the approach with data originating mainly from the sarcomeres structures. The training and validations, showed similar learning patterns, but reaching lower values (Train 88.2%, Validation 90.8%).

Well Level	DCM	WT
DCM1	88.10%	11.90%
WT1	0%	100%
W1D-b	97.62%	2.38%
D1W-a	3.57%	96.43%

Tile Level	DCM	WT
DCM1	78.86%	21.14%
WT1	15.27%	84.73%
W1D-b	75.64%	24.36%
D1W-a	31.58%	68.42%

Conclusion

Among the CRISPR-Cas edited cell clones generated, 3D and 2D functional characterization guided us into selecting two pairs of comparable functional behavior. On one hand, **DCM1** (patient derived) and **W1D-b** (TTN mutation engineered in WT1 cells) exhibited similarly impaired contractility in a 3D Engineered Heart Myocardium assay, while **WT1** (healthy donor) and **D1W-a** and **D1W-b** (corrected TTN mutation from DCM1 cells) displayed comparable 3D behavior. On the other hand, calcium cycling confirmed that both **DCM1/W1D-b** and **WT1/D1W-a** were responding similarly in baseline cycling and under beta-adrenergic stress (chronic isoprenaline exposure model). We have then developed an immunofluorescence assay in 384-well plates based on the calcium imaging protocol, staining for both alpha-actinin and troponin-T. These images generated datasets on which we succesfully developed, trained and validated our Deep Learning model, leading to an accurate classification of the cell lines.

Our next steps will be to assess different pharmacological treatments and optimize the assay to allow for high throughput/high content imaging assay.

References

Tiburcy *et al.* "Defined Engineered Human Myocardium with Advanced Maturation for Applications in Heart Failure Modelling and Repair" **2017** *Circulation.*

Gramlich *et al.* "Antisense-mediated exon skipping: a therapeutic strategy for titin-based dilated cardiomyopathy" **2015** *EMBO Mol. Med.*

Tan and Le "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks" **2019** *arXiv e-prints.*

Bardes *et al.* "VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning" **2021** arXiv.