Target-agnostic drug discovery approach for identification of a modulator of myogenic differentiation impairment in DM1.

Zahra Hanifehlou[1], Amélie Weiss[1], Florine Roussange[2], Kalina Radoynovska[1], Jacqueline Gide[2], Johana Tournois[2], Yong-Jun Kwon[1], Arnaud Ogier[1], Laurent Brino[1], Peter Sommer[1], Cécile Martinat[2]* andChristina Jacob [1]*.

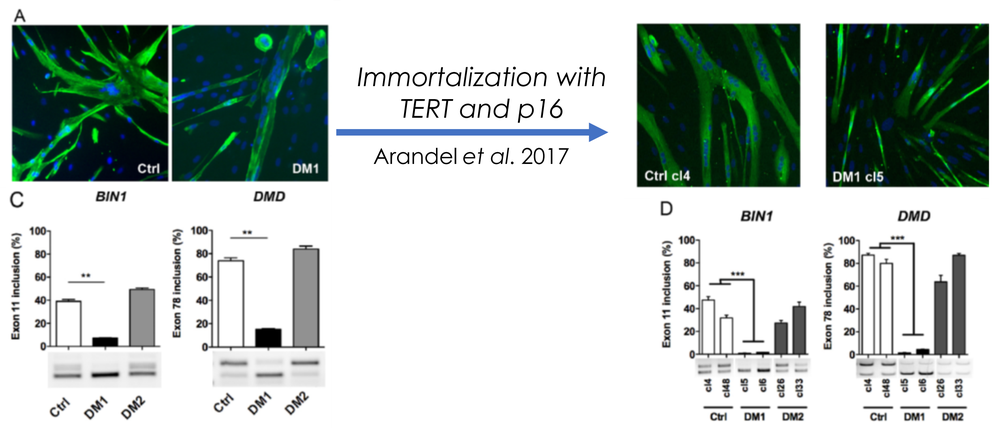
*1. Ksilink, Strasbourg, France. *christina.jacob@ksilink.com 2. I-StemInserm/UEVE-UMR861 – CECS,Evry , France. *cmartinat@istem.fr*

Introduction

Within Neuromuscular diseases (NMDs), Myotonic Dystrophy Type 1 (DM1) is the most prevalent form of adult muscular dystrophy of genetic origin. Despite the fact that causative mutation and gene were identified, the underlying mechanisms of the pathogenesis are yet to be fully understood, which in turn constitutes an obstacle for the development of disease-modifying therapies (Maury *et al*. 2019, Bassez *et al*. 2018).

In this work, we have developed **a target-agnostic assay in a DM1 patient-derived cell model**. We have successfully developed image analysis tools to **screen > 7000 compounds in two doses**. Finally, we're working on developing **Deep Learning methods to improve hit detection** to allow for large scale screening and discovery of new chemical matter.

Model development

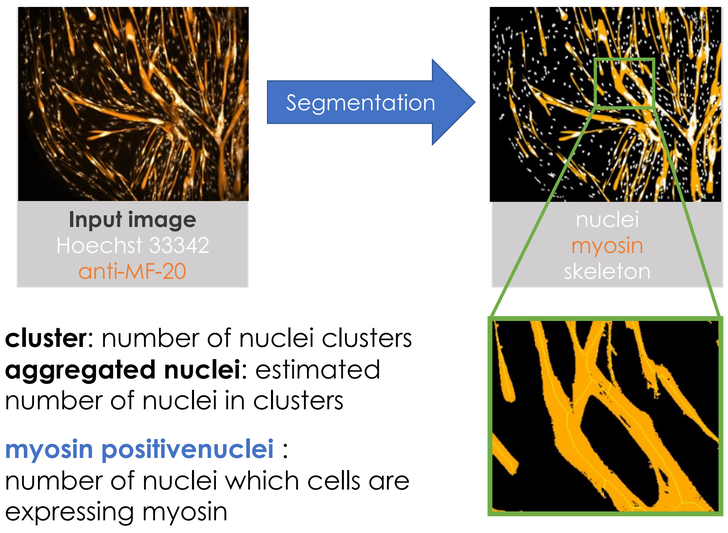


Primary myoblasts from a DM1 patient (2600 CTG)

Immortalized myoblasts from a DM1 patient (around 2600 CTG)

Assay validation and screening

❖ Image Analysis workflow:



cluster: number of nuclei clusters
aggregated nuclei: estimated number of nuclei in clusters

myosin positivenuclei :
number of nuclei which cells are expressing myosin

fusion index myosin:

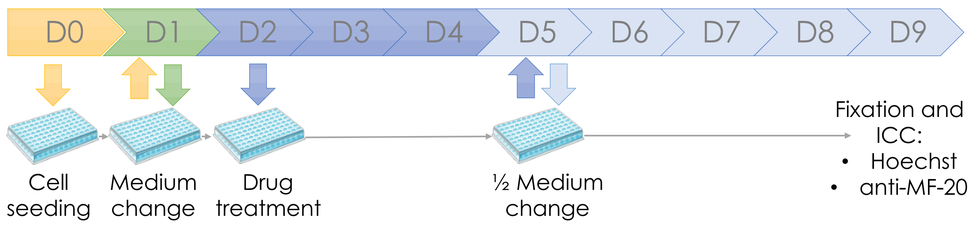
$$\frac{area\ myosin\ pos.nuclei}{total\ nuclei\ area}$$

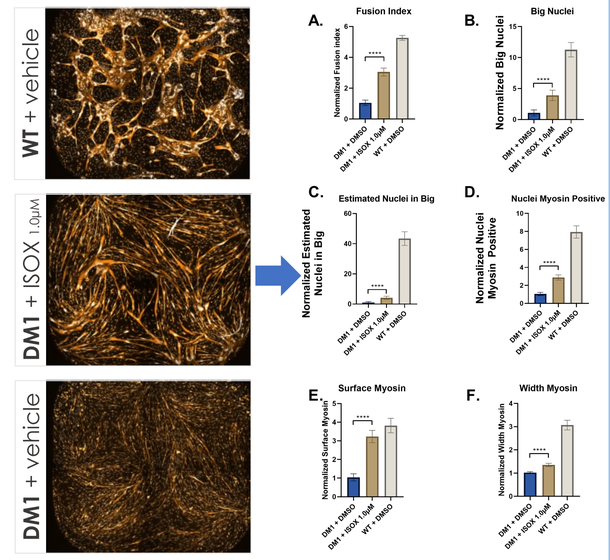
area myosin: total myosin area

width myosin:

$$\frac{myosin\ area}{skeleton\ length}$$

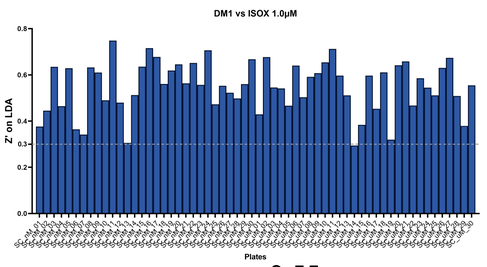
❖ Pharmacological validation:





❖ Screening of >7000 bioactive compounds in 2 doses:

Z'



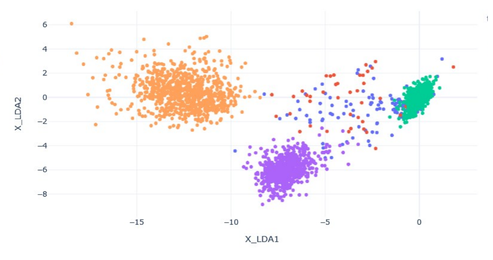
$\mu_{Z'}$ = 0.55

LDA: 6descriptors , 3 classes

Primary actives

2.5µM: 1.6% hit rate
250nM: 0.6% hit rate

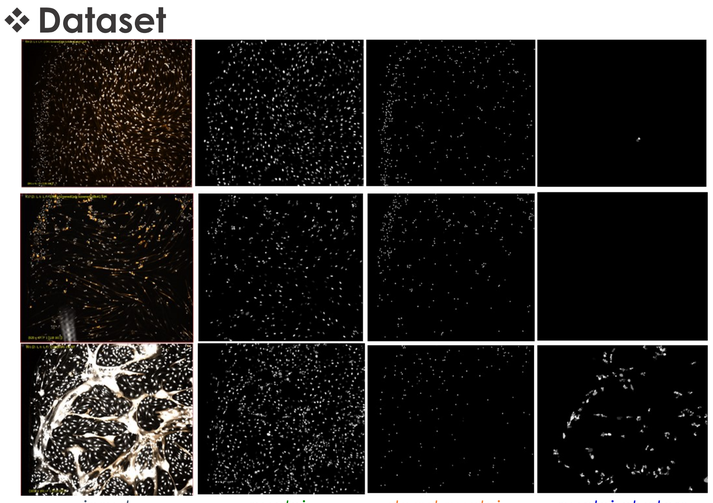
Confirmation



2.5µM: 55% confirmation rate
250nM: 90% confirmation rate

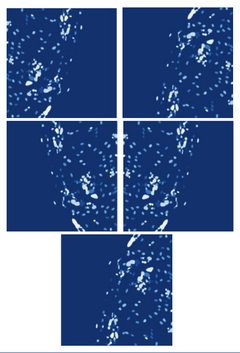
Automatic Nucleus Segmentation

❖ Dataset



Data Augmentation

Rotation (90, 180, 270 degrees), Flipping(left-to-right, up-and-down), Gaussian blurring.



Data set size

- 150 manually segmented images of DM1 cells treated with DMSO or ISOX, and control WT cells.
- Train image size is 256 × 256-pixel.

To increase the amount of data we used data augmentation techniques.

Training

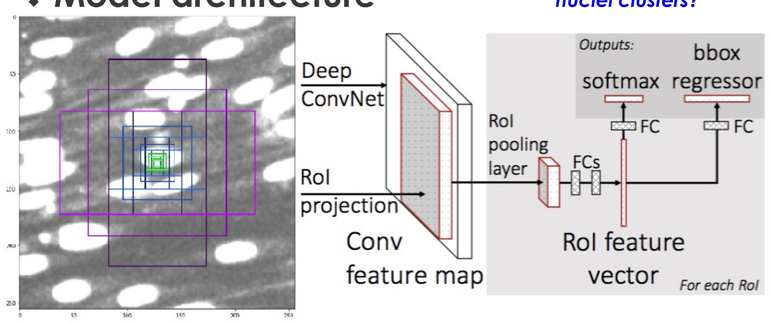
1. Initialize model parameters using the COCO pre-trained model.
2. Train the model with Kaggle's 2018 Data Science Bowl. First *the head is trained, and then all layers are trained.*
3. Fine-tune the model with our data and train *all layers.*

Evaluation

- Mean average precision (mAP) at different thresholds of the IoU between the ground truth and predicted segmentation.

❖ Model architecture

nuclei ?
dead nuclei
nuclei clusters?



Fast R-CNN (Regions with CNN features)

R-CNNs consist of three main types of networks:

1. the model uses few layers of a pre-trained ResNeXt101 to identify promising features from an input image.
2. The convolutional feature passed through the Region Proposal Network (RPN) which uses a series of convolutional and fully connected layers to produce ROIs
3. ROIs passed through a classification network which learns to classify the object contained in each ROI.

Model loss functions:

$$l = l_{cls} + l_{box}$$

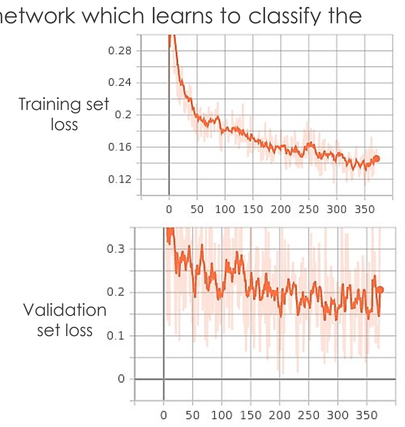
Where l_{cls} is classification loss and l_{box} is bounding box loss.

The overall loss function is:

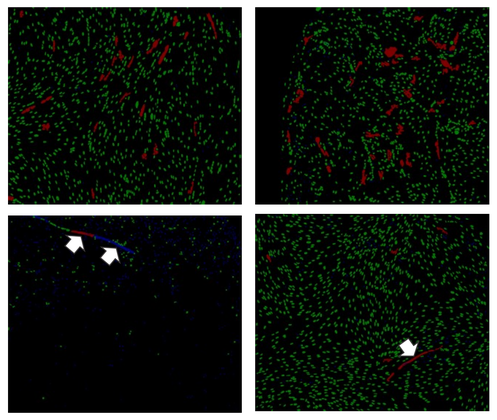
$$l(p, u, t^u, v) = l_{cls}(p, u) + 1[u \geq 1] l_{box}(t^u, v)$$

$$l_{cls}(p, u) = -log p_u \text{ Type equation here.}$$

$$l_{box}(t^u, v) = \sum_{i \in \{x,y,w,h\}} L_1^{smooth}(t_i^u - v_i)$$



❖ Results



Conclusion

After validation of the disease relevant cell lines and development of our in-house image analysis algorithm, we were able to miniaturize and optimize a phenotypic assay recapitulating DM1 phenotype and its pharmacological modulation by a tool compound. Screening of >7000 bioactive compounds in 2 doses and then performing LDA on the identified 6 descriptors gave excellent HTS performance with Z' > 0.5 and good confirmation rate (1 dose, 3 replicates). We successfully developed a DL method to automatically segment nuclei and are now working on automatic Myosin fiber segmentation. To do so, we will use a UNET based deep learning model. Ultimately, these DL methods will allow for a fully automated platform from image generation to hit picking, adaptable to different NMDs and at a larger scale, to uncover new potent therapeutic small molecules.

References

Maury *et al*. "Pluripotent stem cell-based drug screening reveals cardiac glycosides as modulators of myotonic dystrophy 1" **2019** *iSCience*. Arandel *et al*. "Immortalized human myotonic dystrophy muscle cell lines to assess therapeutic compounds" **2017** *Dis. Model Mech*. Kaiming *et al*. "Mask R-CNN" **2018** *arXiv e-prints*. https://www.kaggle.com/c/data-science-bowl-2018.