# A machine learning framework for kinetic phenotypic prediction of neurological disease states in patient- derived cell models

Michael Jones[1], Chi-Chou Huang[1], Hideki Sasaki[1], Tze-Yuan Cheng[1], Tohru Sugawara[2], Yingxiao Shi[2], Justin Ichida[2], James S.J. Lee[1]

[1] DRVision Technologies LLC, 15921 NE 8th St. Suite 200, Bellevue, WA 98008, USA

[2] Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research, University of Southern California, Los Angeles, CA

## Introduction

Patient-derived cell models coupled with long term imaging microscopes and state-of-the-art live cell fluorescent reporters provide a powerful platform to study dynamic morphological and biomolecular events and infer kinetic phenotypes to predict the states of the neurological diseases.

We are developing a machine learning based kinetic informatics discovery (KID) framework. It is applied to induced motor neurons (iMN) culture of amyotrophic lateral sclerosis (ALS) patient populations, to detect subtle differences underlying disease states. A machine learning enabled soft-matching detects somas and neuritis of iMNs. A tracking function tracks all iMN-like objects. Trajectory features for tracked objects are used to filter out non-iMN tracks. The states of the remaining true iMNs are inferred by measuring the responses of fluorescent reporters. A field inference stage uses random-forests classifier to infer the disease states at the field level. Like the field inference step, a patient inference stage uses random-forests classifier to infer the disease states at the patient level. The objective of this study is to validate the effectiveness of the KID framework using iMN movies.

## KID Tool Processing Flow

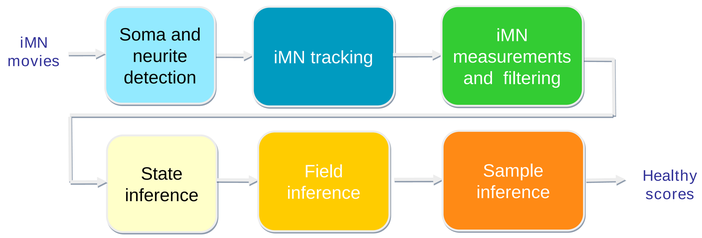


**Fig 1. KID tool processing flow:** The KID tool detects and tracks iMNs, to infer neuron states, and to score at field and sample levels. An iMN movie containing image sequence of iMN generation and survival studies. A soma and neurite detection step inputs the iMN movies to detect somas and neurites of iMN-like objects An iMN tracking step applies to the soma segmentation mask to track all iMN-like objects in the movie. An iMN measurement and filtering step measures fixed point and trajectory features and use them to filter out non-iMN tracks . The tracks remained after the iMN filtering step are mostly true iMNs. The state of the filtered iMN at each time point are inferred by a state inference step. The field inference step infers a summary score for all tracked iMNs in the entire field/movie. The sample field inference step infers a summary healthy score for all tracked iMNs in the entire patient sample.

## iMN movies

Image sequence is acquired using a Nikon BioStation CT. Motor neuron reporter HB9::RFP and neuronal firing reporter Zif-GAP43-Venus/pGG16 were imaged (see Fig 2) every 6 hours in 96 well plates for 27 days in conversion process. Neurotrophic factors were then withdrawn and the survival process was then imaged for 20 days.

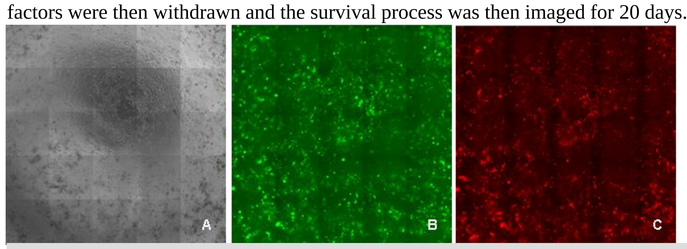


**Fig 2. Three image channels.** A) The bright field phase contrast channel showing the surface of the well. B) The neuron firing reporter channel (green) showing Zif-GAP43-Venus/pGG16+ signal. C) The motor neuron identity reporter channel (red) showing Hb9::RFP+ signal.

## Soma, neurite detection and tracking

A robust stitching and alignment module was designed to align neighboring image tiles when stitching the whole field of view at each time point, and to compensate for stage movements over time. The iMN detection accuracy and tracking accuracy were greatly improved with the correctly aligned image sequences. A machine learning enabled segmentation function, soft-matching, was applied to each time frame of the motor neuron channel (red) to detect iMN somas and neurite. A special AIVIA tracking recipe was applied to the soma segmentation mask to track all iMN-like objects in the time-lapse sequence. The example detection and tracking results of four iMNs and their morphology over time are shown in gallery view in Fig 3.

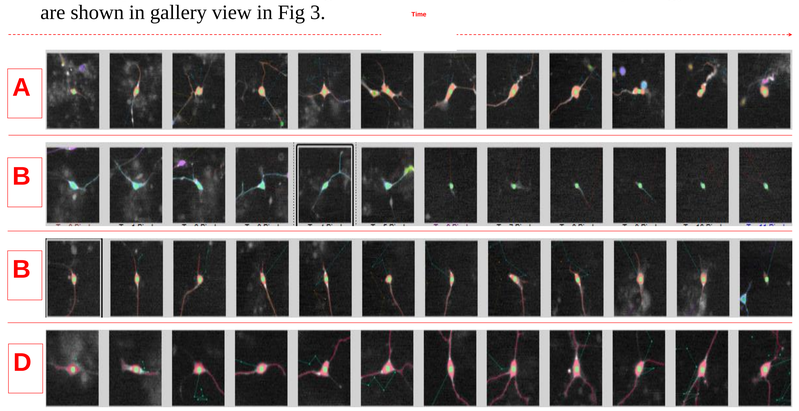


**Fig 3. Gallery view showing detection and tracking results of four iMNs and their morphology over time.** (A) Gallery view of the first iMN. B) Gallery view of the second iMN. C) Gallery view of the third iMN. D) Gallery view of the fourth iMN.

## iMN measurements, filtering and state inference

Fixed point neuron features, including soma measurements, and neurite measurements were calculated for all tracked objects. Trajectory features for each tracked object were then derived from fixed point neuron features of the whole track, and used for iMN filtering. Non-iMN tracks were filtered out if (a) the total track length is too short, or (b) the maximum neurite length is too short, or (c) not have enough track points with detectable neurites, or (d) not associate with any neurites.

The tracks remain after the iMN filtering step are mostly true iMNs. The state of the iMN at each time point could then be inferred by measuring the mean intensity within the soma area from the neuron firing reporter channel image (green). An iMN is at the neuronal firing state if its mean intensity of neuron firing channel is greater than a certain threshold; otherwise it is at the non-firing state

## Field and sample inferences

In the field inference step, field measurement summary statistics are derived from trajectory features of all tracked iMNs in the entire field/movie. Field measurement summary statistics include the minimum, maximum, average and standard deviation of the six neuron measurements: Soma area, Average intensity within soma area from the neuronal firing reporter channel, Average intensity within soma area from the motor neuron reporter channel, Maximum dendrite length, Dendrite branches and Track length. To score a field, we trained a random-forests classifier by supervised learning using 24 field measurements of a set of training fields from both healthy controls and diseased patients.

For sample inference, 24 measurements for a sample are calculated by averaging all the 24 field measurements from all available fields of the sample. Like the field inference, we trained a random-forests classifier by supervised learning using all 24 patient measurements of a set of training healthy controls and diseased patients.

## Study Materials and Methods

### Detection and tracking accuracy evaluation

We manually created truth by identifying iMNs that should be detected and tracked for 12 selected movies, 4 from healthy controls and 8 from diseased patient lines.

### State inference accuracy evaluation

We created the neuronal firing signal truth mask for the same 12 movies used for detection/tracking accuracy evaluation by thresholding the neuron firing reporter image with grey level of 50 (the maximum grey level is 255). The truth for an iMN is at the neuronal firing state if 90% of its soma area overlaps with the neuronal firing signal truth mask, otherwise it is at non-firing state.

### Field inference accuracy evaluation

We evaluated fields of ten test movies (5 healthy and 5 diseased) to determine their field scores (0.0 to 1.0). A field is considered healthy if its score is greater than 0.5, otherwise it is considered diseased.

### Sample inference accuracy evaluation

We evaluated sample inference by eight test patients (3 healthy and 5 diseased) to determine their healthy scores (0.0 to 1.0). A patient is considered healthy if its score is greater than 0.5, otherwise it is considered diseased

## Results

The accuracy results are summarized below:

- ❑ Detection accuracy: 94.6% $\pm$ 0.287%
- ❑ Tracking accuracy: 94.5% $\pm$ 1.387%
- ❑ State inference accuracy: 87.9% $\pm$ 0.408%
- ❑ Field inference accuracy: 90%
- ❑ Patient diagnosis accuracy: 87.5%

To further valid the KID tool, we used a masked test set consisting of 5 healthy and 5 patient lines to compare Z factor between KID score and survival alone. The KID tool showed Z factor of 0.5056 (good separation). The Z factor of the conventional survival assay using iMN survival metrics was -19.43 (very poor separation).

## Discussions and Conclusion

These results provide a strong evidence of analyzing dynamic molecular events using time-lapse microscopy for discovering the pathogenic mechanisms underlying disease phenotypes. We will multiplex additional biomolecular reporters for spatial-temporal functional dynamics and cell fate readouts in the future.

## References

1. **Sasaki, T. Cheng, M. Jones, Y. Li, H. Lai1 , C. Huang, J. Ichida, S.J. Lee.** Inference of survival states in induced motor neurons of neurological diseases. *Poster presented at the American Society for Cell Biology (ASCB) annual meeting, San Francisco, CA. 2016*
2. **Cheng, G.R. Linares, H. Sasaki, H. Lai, W. Donaldson, C. Huang, J. Ichida, S.J. Lee.** Machine learning-enabled analytical tool for kinetic drug screening of induced motor neuron survival in ALS patient panels. *Poster presented at the ASCB annual meeting, San Francisco, CA. 2016*

## Acknowledgments

This research was supported in part by grant no. 4R44NS097094-02 from the NINDS and grant no. 5R43MH100780-02 from the NIMH