

# UCSF

## UC San Francisco Previously Published Works

### Title

Trans-channel fluorescence learning improves high-content screening for Alzheimer's disease therapeutics.

### Permalink

<https://escholarship.org/uc/item/48x87915>

### Journal

Nature machine intelligence, 4(6)

### ISSN

2522-5839

### Authors

Wong, Daniel R  
Conrad, Jay  
Johnson, Noah  
[et al.](#)

### Publication Date

2022-06-01

### DOI

10.1038/s42256-022-00490-8

Peer reviewed



Published in final edited form as:

*Nat Mach Intell.* 2022 June ; 4(6): 583–595. doi:10.1038/s42256-022-00490-8.

## Trans-channel fluorescence learning improves high-content screening for Alzheimer’s disease therapeutics

**Daniel R. Wong**<sup>1,2,3,4,5,6,7</sup>, **Jay Conrad**<sup>\*\*1</sup>, **Noah Johnson**<sup>\*\*1,9,10</sup>, **Jacob Ayers**<sup>1</sup>, **Annelies Laeremans**<sup>4,8</sup>, **Joanne C. Lee**<sup>1</sup>, **Jisoo Lee**<sup>1</sup>, **Stanley B. Prusiner**<sup>1</sup>, **Sourav Bandyopadhyay**<sup>4,8</sup>, **Atul J. Butte**<sup>2,6,7</sup>, **Nick A. Paras**<sup>1</sup>, **Michael J. Keiser**<sup>1,2,3,4,5,\*</sup>

<sup>1</sup>Institute for Neurodegenerative Diseases, University of California, San Francisco, CA, 94158, USA

<sup>2</sup>Bakar Computational Health Sciences Institute, University of California, San Francisco, CA, 94158, USA

<sup>3</sup>Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA, 94158, USA

<sup>4</sup>Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA, 94158, USA

<sup>5</sup>Kavli Institute for Fundamental Neuroscience, University of California, San Francisco, San Francisco, CA, 94158, USA

<sup>6</sup>Department of Pediatrics, University of California, San Francisco, CA, 94158, USA

<sup>7</sup>Center for Data-Driven Insights and Innovation, University of California, Office of the President, Oakland, CA, 94607, USA

<sup>8</sup>Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA, 94158, USA

<sup>9</sup>University of Colorado Alzheimer’s and Cognition Center, Linda Crnic Institute for Down Syndrome, University of Colorado, Anschutz Medical Campus, Aurora, CO, 80045, USA

<sup>10</sup>Department of Neurology, University of Colorado, Anschutz Medical Campus, Aurora, CO, 80045, USA

---

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:

\*Correspondence: keiser@keiserlab.org.

\*\*These authors contributed equally

### Author Contributions Statement

Conceptualization, D.R.W., J.C., N.J., N.P., and M.J.K.; Methodology, D.R.W., J.C., N.J., N.P., and M.J.K.; Software, D.R.W.; Validation D.R.W., A.L., and S.B.P.; Formal Analysis, D.R.W.; Investigation, D.R.W., J.C., N.J., J.C.L., J.A., J.L., and A.L.; Resources, S.B.P., S.B., A.J.B., N.P., and M.J.K.; Data Curation, D.R.W.; Writing – Original Draft, D.R.W.; Writing – Review & Editing, D.R.W., and M.J.K.; Visualization, D.R.W.; Supervision, S.B.P., S.B., A.J.B., N.P., and M.J.K.; Project Administration, D.R.W., and M.J.K.; Funding Acquisition, S.B.P. and M.J.K.

### Competing Interests Statement

The authors declare no competing interests. Stanley B. Prusiner is a member of the Scientific Advisory Board of ViewPoint Therapeutics and a member of the Board of Directors of Trizell, Ltd., neither of which have contributed financial or any other support to the studies discussed here.

## Abstract

In microscopy-based drug screens, fluorescent markers carry critical information on how compounds affect different biological processes. However, practical considerations, such as the labor and preparation formats needed to produce different image channels, hinders the use of certain fluorescent markers. Consequently, completed screens may lack biologically informative but experimentally impractical markers. Here, we present a deep learning method for overcoming these limitations. We accurately generated predicted fluorescent signals from other related markers and validated this new machine learning (ML) method on two biologically distinct datasets. We used the ML method to improve the selection of biologically active compounds for Alzheimer's disease (AD) from a completed high-content high-throughput screen (HCS) that had only contained the original markers. The ML method identified novel compounds that effectively blocked tau aggregation, which had been missed by traditional screening approaches unguided by ML. The method improved triaging efficiency of compound rankings over conventional rankings by raw image channels. We reproduced this ML pipeline on a biologically independent cancer-based dataset, demonstrating its generalizability. The approach is disease-agnostic and applicable across diverse fluorescence microscopy datasets.

---

## Introduction

HCS efforts generate a wealth of complex phenotypic information pivotal to the drug discovery process. Extracting biological insight from these observations, however, often requires performing multiple experiments, which can demand extensive time and resources<sup>1</sup>. We posit that ML can extract actionable information otherwise encoded within HCS, and learn the phenotypic relationships between related biological processes. As a corollary, we can potentially glean information on related but different biological processes—increasing the power of computationally mining large archival HCS datasets to gain new information.

The traditional way to gather information within HCS is simply to add biological markers to track different processes. However, we cannot introduce new markers for archival datasets because the experiments are finished. Furthermore, capturing additional markers may be impractical due to expensive or cumbersome visualization procedures, like the optimization of multi-channel fluorescent immunohistochemistry or interference across detection wavelengths<sup>2,3</sup>. We overcame these marker limitations computationally by directly *learning* the phenotypic relationships between a highly informative but cumbersome marker and other similar yet more easily accessible markers. These hidden relationships were then projected into *de novo* images displaying the desired fluorescent signal of the cumbersome marker. In this study, we focused on an archival HCS dataset that tracked the phenotypic effects of small molecules for the treatment of AD.

Despite extensive drug discovery efforts, no effective treatments exist that prevent or even slow the progression of AD<sup>4</sup>. Of the many therapeutic targets under investigation<sup>5</sup>, the pathogenic<sup>6–13</sup> misfolding and accumulation of tau protein into neurofibrillary tangles (NFTs) has emerged as a target mechanism<sup>10</sup>. Younger decedents tend to have more aggressively propagating tau prions than older decedents<sup>12</sup>. Furthermore, studies have linked this tau prion propagation<sup>14,15</sup> and aggregation<sup>16</sup> to increased tau hyperphosphorylation.

Although hyperphosphorylation is not required to form ordered assemblies of tau<sup>17,18</sup>, NFTs in the brains of deceased AD patients are highly enriched for hyperphosphorylated tau (pTau)<sup>16,19</sup>. Hyperphosphorylation inhibits the binding of tau to microtubules<sup>20,21</sup>, thus preventing normal microtubule assembly<sup>21,22</sup> and axonal transport<sup>23</sup>. It may also increase the propensity of tau to be recruited into NFTs<sup>6,23,24</sup>. These factors promote disease phenotypes including synaptic dysfunction, enhanced neuroinflammation, and neuronal cell loss<sup>21,23,24</sup>.

Accordingly, discovering compounds that inhibit this tau prion propagation may improve our understanding of the disease and provide avenues for novel therapeutics. Our study focused on inhibiting the propagation and aggregation of tau in cells exposed to tau prions. To screen for compounds, we developed a HCS procedure that utilizes a biosensor cell-line overexpressing a 0N4R isoform of tau fused to the yellow fluorescent protein (YFP-tau)<sup>25</sup>. Due to the correlation of tau oligomerization with increased hyperphosphorylation, we posited that tracking the level of pTau in response to drug treatment would provide an additional point of validation to identify molecules that inhibit prion propagation. Furthermore, we hoped that synthesizing pTau signal would eliminate spurious signal such as contamination of the YFP channel by autofluorescent cell debris. The AT8 antibody allows us to quantify these correlated hyperphosphorylation events and is traditionally used to label disease-associated paired helical filaments in human tissue samples<sup>26</sup>. This antibody is specific to the Ser202/Thr205 epitope<sup>27,28</sup>, which is one of the most hyperphosphorylated residues in AD-afflicted brains<sup>16,29</sup>. Furthermore, hyperphosphorylation of this region has been linked to increased prion propagation<sup>14,15</sup> and aggregation<sup>16</sup>. Although AT8 immunoreactivity is a useful surrogate to identify a disease-relevant phenotype and recognize clinically relevant pTau<sup>27</sup>, it was not used in our HCS because immunostaining would preclude some advantages of live-cell imaging<sup>30,31</sup>. Therefore, we turned to a different solution based on ML. We computationally synthesized a representation of the AT8 channel from existing data, rather than repeating the vast HCS with this useful antibody, which would have been laborious and cost prohibitive. This approach of synthesizing biomarker phenotypes is in contrast to other HCS studies for Alzheimer's<sup>32-34</sup>, and ML-based methods<sup>35</sup> such as predicting genetic risk factors<sup>36</sup>, small molecule structures<sup>37</sup>, and docking<sup>38,39</sup>.

Two seminal papers<sup>40,41</sup> describe a ML method for predicting fluorescence images from bright-field images. Augmented microscopy approaches like these extract latent information from images post hoc<sup>42</sup> and in a goal-driven way<sup>43,44,45</sup>. We reasoned that if two fluorescent channels were sufficiently related, we could train models to extract hidden information and generate “trans-channel” images depicting AT8-pTau from related YFP-tau images. Our logic hinged on the expectation that a ML method could both identify subtle aggregate morphology and exploit the phenotypic correlations between tau hyperphosphorylation and aggregation.

As schematized in Supplementary Figure 1, we collected a three-channel dataset—4',6-diamidino-2-phenylindole (DAPI) nuclear marker, YFP-tau, and AT8-pTau—one year after the original HCS experiment that had only used YFP-tau and DAPI. After training our generative model on this new three-channel dataset, we applied this model to the archival

HCS dataset to synthesize predicted AT8-pTau channels. We then evaluated whether the computed AT8-pTau images would better guide discovery efforts and improve the screen. To our knowledge, this is the first prospective demonstration of reliably constructing new fluorescent images *in silico* from pre-existing fluorescent markers—providing a new way to improve the drug discovery process and leverage complex biological features otherwise latent within historical HCS archives. Moreover, we assessed the generalizability of this trans-channel method by applying it to an entirely unrelated biological context of a genome-wide functional screen in U2OS osteosarcoma cells.

## Results

### We Constructed a Dataset of Tau Propagation for ML Training

We sought to improve *in vitro* screening by constructing a new image channel from abundant, common, and relatively inexpensive channels—even after a screen’s completion. The archival HCS data contained extensive information on many compounds, but it only contained live-cell YFP-tau and DAPI images and could not have contained an AT8-pTau channel, as this required a fixed cell format. Therefore, our first task was to generate a completely independent dataset to train a ML model to predict the AT8-pTau channel solely from the YFP-tau and nuclear DAPI channels.

We constructed the new training dataset using a fixed cell system and collected all three channels (YFP-tau, DAPI, and AT8-pTau) representing a range of conditions. Ideally, ML models leverage large and diverse datasets during their training. Hence, we collected varied image data so that the model could learn to capture information about different conditions and perturbation scenarios that may be found in the HCS dataset, which itself has many different compounds present. In the training dataset, we curated a range of six small molecule perturbations, each resulting in different tau aggregation and cell viability phenotypes (Methods).

We collected the new training dataset to model infectious tau propagation similar to the archival HCS dataset (Methods). We imaged three channels—DAPI, YFP-tau, and AT8-pTau (Figure 1)—to obtain 57,600 non-overlapping images that were  $3 \times 2048 \times 2048$  pixels in size. The archival HCS dataset was similar, differing predominantly in its use of live-cell imaging without AT8, and also containing thousands of compounds. Otherwise, the protocols were identical (Methods).

### We Trained a ML Model to Predict Hidden Tau Phosphorylation

To learn a mapping from the input DAPI and YFP-tau channels to the output AT8-pTau channel, we designed a convolutional neural network (CNN) model motivated by the U-Net architecture<sup>43</sup>, but for a task other than segmentation. Our design employed “skip connections”<sup>43</sup> that preserved lower complexity features from earlier layers and combined them with higher complexity features from deeper layers (Figure 2A). The model is 12% the size of U-Net, preserves image dimensions, and generates a non-binary image channel (Supplementary Figure 2). We made our PyTorch code and fully trained models openly available (Methods). We randomly stratified the three-channel images into a 70% train and

30% held-out test split. On visual inspection, the ML model removed aberrant signal and accurately mapped the phenotypes of the related channels to construct realistic AT8-pTau images with a high resemblance to the actual objectives (Figure 2B).

The ML model learned to identify and enhance regions of interest in the YFP-tau channel, effectively generating AT8-pTau images by gleaning visual information about the relationships between a hidden phosphorylation readout and the YFP and DAPI channels. It removed YFP-tau signal that was not present in the AT8-pTau image and selectively retained pertinent signal (Figure 2B). On visual inspection, there did not appear to be an intuitive set of visual criteria or easily predefined attributes like pixel intensity by which a human could accomplish the same task (e.g., see Supplementary Figure 3). The model's decision-making appeared to be non-trivial.

We evaluated the model quantitatively on the held-out test set of YFP-tau and DAPI inputs and AT8-pTau objectives. To assess image similarity between the observed and predicted AT8-pTau images, we calculated the Pearson correlation coefficient (PCC) and also the mean squared error (MSE) metric, which is more commonly used in regression tasks. A PCC=1.0 indicates exactly identical images (with a possible constant scaling factor), whereas 0.0 indicates complete pixel-wise disagreement. We obtained an average PCC=0.74±0.19 over the test set of  $n=17,280$  images (Figure 3A). This corresponded to a MSE=0.52±0.39 after normalizing to zero-mean and unit-variance (Figure 3B). To quantify its pixel-wise performance, we measured the trade-off between true positive pTau pixels and false positives via the area under the receiver operating characteristic (AUROC), as well as the tradeoff between pixel precision and recall via the area under the precision recall curve (AUPRC). The model achieved an AUROC=0.98 and an AUPRC=0.65 (Figure 3C). Cross-validation results are shown in Supplementary Figure 4. Importantly, we found that learning from either single channel alone was insufficient to glean the AT8-pTau signal (Supplementary Figure 5).

The YFP-tau and AT8-pTau channels contained substantial overlapping information. If they did not, the prediction task would be difficult or impossible. In theory, signal from the YFP-tau is a superset of the AT8-pTau signal and includes all of the introduced tau and aggregated protein (Supplementary Figure 6). Thus, given the similarity between the YFP-tau channel and AT8-pTau channel, we might theoretically achieve good "prediction" by simply extracting the input YFP-tau *as-is* and stipulating this as the model's predicted output. This trivial Null YFP Model required no ML, and yielded a PCC=0.53±0.23 (corresponding to an MSE=0.94±0.47) over the entire dataset (Figure 3A-B). The ML model exceeded this baseline with an increase in average PCC=0.21 ( $p < 0.00001$ ; corresponding to a decrease in average MSE=0.42), which is consistent with the model learning to generate an output that more closely approximated phosphorylation state than was already provided by the YFP-tau input alone. We also created a Null DAPI Model that simply returned the input DAPI image as the output (PCC=0.18±0.13, corresponding to a MSE=1.64±0.26). The ML model's test performance was consistent across all six drug perturbations (Supplementary Figure 7).

As with most multichannel imaging studies, we were wary of image bleed-through that the model might exploit as a hidden crutch. Excitation and emission plots of the different fluorophores are shown in Supplementary Figure 8A. To test the hypothesis that the ML model leveraged undetectable but pernicious hidden AT8 signal within the input channels, we performed pixel-intensity ablations on the input images at increasing intensity percentiles to eliminate potential low-intensity bleed-through signal (Supplementary Figure 9). We did not detect bleed-through as a confounder. Accordingly, we note that any confounding signal augmenting a model's performance would be of no help when applied subsequently to the archival HCS dataset, since this HCS did not undergo immunohistochemistry preparation and had neither AT8 antibody nor fluorophore.

### ML Improved Hit Rate and Compound Triaging in Tauopathy HCS

Turning to its prospective use in drug discovery, we applied the trans-channel ML model to the archival HCS dataset and found practical improvements. For most HCS pipelines, immunostaining is not normally employed due to batch variability and increased labor and cost, and must be balanced against the advantages of live-cell formats such as time-course data collection. If ML could inexpensively infer these immunohistochemistry channels, then it could enhance HCS efforts by providing previously unavailable channels and inferences—thereby improving drug discovery pipelines. Similarly, one might imagine a wealth of data could be mined in a hypothesis-guided way from substantial existing datasets of completed screens in the public and private sectors, thus advancing biological knowledge and improving medical therapies.

Compounds are conventionally ranked by priority, favoring those that lower tau aggregation for secondary dose-response testing<sup>46</sup>. Medicinal chemists also consider chemical structure. Instead, we purposely took a naive approach by having the ML model make decisions based solely on cellular phenotype—unguided by human intuition. To assess the ML method for practical use, we directly compared it to the conventional method of drug discovery unguided by ML. To our knowledge, this study is the first practical application of trans-channel ML image generation actively used in a conventional in-house screen. Using the archival HCS dataset, we prospectively generated an AT8-pTau channel for each compound using the predictive model. Due to the large dataset size, we randomly selected one run of the HCS (consisting of 1,600 unique compounds) for the analysis. We constructed a ML-derived priority queue (PQML), ranking compounds based on aggregation scores using the predicted, machine-learned AT8-pTau images. We then calculated aggregation scores using the same General Electric (GE) cellular image software (Methods) as in the original archival HCS evaluation pipeline a year before. The conventional method's priority queue (PQC) ranked each compound solely by the aggregation scores of pre-existing YFP-tau images. Hence each compound received rankings in both queues and these rankings could disagree substantially.

We prospectively collected complete dose-response profiles for the top 40 compounds from each queue. We chose 40 compounds (per Methods) for each queue due to cost and labor limitations. Despite operating solely on a computationally generated channel, the PQML's hit rate was comparable with the conventional method and effectively proposed

overlooked compounds that passed dose-response testing. Of the 40 compounds tested from PQML, 11 passed secondary testing (27.5%). Of the 40 compounds tested from PQC, 12 were active (30%). Importantly, compounds that were highly prioritized by both the ML and conventional methods obtained a much higher hit rate than either method alone. Ten compounds ranked within the top 40 of both queues and 6 out of these 10 were confirmed by dose response. Taken together, progressing compounds by this combined criterion yielded a success rate of 60%—an improvement of 30% over the conventional method and 32.5% over the ML method (Supplementary Table 1). However, our sample size was smaller for these overlap compounds, which must be considered when assessing generalizability.

Interestingly, 5 of the 11 active compounds in PQML's top 40 were effectively missed in the PQC, which ranked them 539<sup>th</sup>, 545<sup>th</sup>, 560<sup>th</sup>, 582<sup>nd</sup>, and 1,596<sup>th</sup> out of 1,600 possible compound ranks. The activity of one such rescued compound is shown in Figure 4A. The ML model eliminated YFP-tau aggregates (false-positive imaging signal) that yielded a poor ranking for the compounds in the PQC (Figure 4B). Supplementary Table 2 shows all compound dose-response profiles for reducing tau aggregation and cell count.

Additionally, the ML method achieved higher enrichment<sup>47</sup> for active compounds. We discovered a total of 17 unique dose-response confirmed compounds from both lists. For these compounds, PQML had an area under the enrichment curve of 0.93 versus PQC's area under the enrichment curve of 0.85 (Figure 5A). These active compounds achieved an average rank of 119 in PQML, which was better than nearly half their average rank of 235 in PQC (Figure 5B). Therefore, fewer compounds from PQML would need to be tested to find the same number of active compounds from PQC.

Of the 17 total active compounds from both lists, 5 fell outside of PQC's top 40, and 6 fell outside of PQML's top 40. However, to recover the active compounds missed in each case, we would need to test more than twice as many candidate compounds on average ( $n=764$ ) by the PQC rankings as would be necessary by the PQML rankings ( $n=302$ ). Hence, the ML method ranked missed-but-active compounds more than twice as well as the conventional method—including them in a smaller, more tractable search space (Supplementary Figure 10A-B).

To compare the utility of trans-channel prediction with other imaging-based profiling analysis approaches<sup>45</sup>, we investigated various alternative methods, such as adding noise to the HCS images to prioritize compounds solely by their visual robustness; the use of CellProfiler<sup>48</sup>; and finally, operating instead on feature extraction via convolutional autoencoders (Methods; Supplementary Figure 11). The ML method achieved the highest AUC (Figure 5). Furthermore, when we calculated a compound's dose-response activity from the ML-derived AT8-pTau images instead of scoring all queues solely by the conventional YFP-tau images, the ML method indeed well exceeded other methods' scores (Supplementary Figure 10C-F).

### We Tested ML Method on HCS with Different Conditions

To assess whether the success of our trans-channel ML approach was specific merely to the tauopathy HCS, we investigated whether the method of learning fluorescent signal from



related markers applied to a HCS unrelated in its biology and perturbation. We performed the same trans-channel fluorescence learning task. However, compared to the tauopathy screen, this HCS used an entirely different cell line, microscope, perturbation method, and fluorescent markers. ML models can exploit dataset-specific patterns in subtle ways; thus, our goal was to assess the approach where many of the data parameters differed from the tauopathy screen.

We performed a functional genomics HCS in a cancer cell model. In this fluorescence-based arrayed whole-genome screen (details in Methods), we plated a U2OS cell line with a perturbation scheme that systematically knocked down all coding genes with siRNA. We marked cellular DNA with Hoechst fluorescent dye and tracked the cyclin-B1 protein with a green fluorescent protein (GFP) fusion. These two markers are biologically related, with cyclin-B1 specific to the G2/M phase of mitosis and also involved in DNA damage repair<sup>49</sup>. In theory, the two markers contain shared information, so we investigated whether the ML model could learn this signal relationship. We collected a total of 324,989 images.

Unlike the tauopathy dataset, we observed subtle bleed-through signal and artifacts in the raw Hoechst channel (Figure 6A; Supplementary Figure 8B). Hence, as a preprocessing step to mitigate bleed-through, we ablated all pixels below the 95<sup>th</sup> percentile in our Hoechst images, which inevitably removed some Hoechst signal as well (Figure 6A). We performed threefold cross-validation to predict cyclin-B1 signal solely from the ablated Hoechst channel. These models achieved an average pixel-wise PCC=0.75±0.13 (corresponding to an MSE=0.50±0.26) on  $n=108,330$  test images (Figure 6B). This was 87.5% ( $p < 0.00001$ ) higher than the Null Model that used the ablated input as its prediction (PCC=0.40±0.14; MSE=1.21±0.28). Despite heavy ablations to the models' input, the models were able to learn the non-trivial cyclin-B1 phenotype. Upon visual inspection, predicting where high-intensity cyclin-B1 signal resided solely from the Hoechst signal did not appear obvious. As an unexpected benefit, the ablation procedure resulted in image predictions that were mostly free of signal artifacts in the cyclin-B1 channel (Figure 6A).

When we omitted the ablation procedure and trained a threefold cross-validation on raw images, the ML model accurately predicted cyclin-B1 signal and produced detailed and realistic images that matched the observed images to which it had been blinded (Figure 6A). This resulted in an average PCC=0.85±0.08 (corresponding to an MSE=0.30±0.15) on  $n=108,330$  test images (Figure 6C). This was 93.2% higher ( $p < 0.00001$ ) than the Null Model that stipulated the raw Hoechst image as its prediction (PCC=0.44±0.12; MSE=1.11±0.23). Training on raw, unablated images resulted in a 13.3% ( $p < 0.00001$ ) increase in average PCC performance versus the ablated training procedure; we also quantified the performance change from progressively ablated inputs (Supplementary Figure 12).

## Discussion

We developed and assessed a method to computationally augment fluorescence microscopy and HCS. This method could be especially useful for tapping into otherwise hidden information in large and information-rich archival datasets. Three observations merit

emphasis: (1) trans-fluorescence models generated biologically informative images that were drop-in replacements for existing HCS workflows; (2) the model operated on an archival HCS dataset independent from the newly created training dataset and facilitated discovery of new compounds; (3) the method generalized to a different biological environment (a functional genomics screen in U2OS cells).

We readily integrated the trans-channel predictions into a conventional compound evaluation workflow because all of the *in silico* channel images were generated as drop-in replacements for the equivalent *in vitro* experiment. Generating a predicted AT8-pTau image from existing YFP-tau and DAPI images required only fractions of a second. With the proper hardware or cloud resources, this parallelizable method scales to large, high-volume HCS datasets. Its ability to accurately predict trans-channel fluorescence can potentially serve as a biologically informed and economically practical means of hypothesis generation, especially if the desired marker in question has barriers to widespread utilization. This method is also pragmatic where information-rich bright-field images are not available<sup>40</sup>, which was the case for our screen. Furthermore, accurate *in silico* labeling decreases variability from possible experimental labeling artifacts.

As measured by dose-response confirmed compounds, the ML method performed similarly to the conventional method in the drug discovery task. We were intrigued, however, that the ML method found a comparable number of active compounds, many of which were missed in the original screening campaign. A strong comparison could not be made between the approaches and their propensity to find active compounds at the small scale of 17 unique active compounds, when the methods' hit rates differed by one hit. Rather, the demonstrated utility of the ML method was to triage compounds more efficiently and to rescue efficacious compounds that were missed due to poor rank in the primary screen. Encapsulating information on AT8-marked phosphorylation and projecting it onto the HCS dataset facilitated compound ranking and subsequent validation. Chemical information is available in Supplementary Table 3.

We acquired the three-channel fixed-cell HCS training dataset using a slightly different protocol than the archival live-cell HCS dataset. Separated by differences in protocol, cell format, experimenters, and a year, the ML model relied solely on extracting hidden signal of phenotypic phosphorylation from YFP-tau images—with no AT8 labeling in the HCS, and thus no possibility of fluorescent bleed-through. This implies that the model has value outside of the strict data regime on which it was trained. This is not always the case in ML applications, which often do not generalize<sup>50</sup>. The archival HCS test dataset consisted completely of compounds that were not part of the training dataset, yet the model was able to operate on new images and a new and larger compound space greater than a thousand compounds to facilitate discovery.

By evaluating the method also on the biologically unrelated U2OS dataset, we showed that trans-channel fluorescent learning worked across disparate data domains. Intended only as a first test in a different disease and biological domain, we did not seek to address nuanced biological questions related to accurately constructing cyclin-B1 signal solely from Hoechst signal. Hence, we did not evaluate whether this method could improve screening

here. We hope that researchers can tailor the method to their own use cases in order to augment phenotypic screening endeavors in ways we have not envisioned. Furthermore, the architecture is fully convolutional, invariant to image size, and adaptable to any input shape. Hence, we expect this method to aid development efforts across a wide range of image applications.

Several caveats limit and focus the scope of this study. The tauopathy cell line was not neuronal, and HEK cells were chosen for practicality—HEK cell lines are frequently used in drug screens since they are easy to grow and transfect. Additionally, the method, as with most ML applications, was data intensive and required many image examples for learning. This could be a pragmatic hurdle, but we hope to leverage transfer learning<sup>51</sup> as openly shared cellular image data become increasingly common<sup>52</sup>. Accordingly, we have made all of the tauopathy training dataset and one full plate of the U2OS functional genomics HCS dataset publicly available (Methods). For new biological conditions, this method requires new experimental work to produce training data, and models cannot be applied out-of-the-box on unseen marker sets. Furthermore, the AT8-pTau images collected experimentally could be missing information on tau that is phosphorylated at other epitopes of interest. Thus, what we currently interpret as YFP off-target signal—i.e., the “false positive” pixels that show fluorescence in the YFP-tau but *not* in the AT8-pTau channels—could represent disease-relevant phosphorylation at sites other than the Ser202/Thr205 epitope. Since the model is trained to only recapitulate phenotype from morphological features between channels, the “false positive” YFP-tau aggregates would seem not to have been phosphorylated at the Ser202/Thr205 epitope. As positive aggregate signal contributes to worse ranking, removing false positives lowers a compound’s aggregation score and thus increases its prioritization. A future exercise would be to assess antibodies for other commonly hyperphosphorylated epitopes. Consequently, we did not seek with this study to implicate tau hyperphosphorylation as a disease-inducing mechanism, but rather used the AT8 marker as an approximate surrogate for a relevant disease phenotype. Lastly, the model had high AUROC and AUPRC, but lower Pearson performance (Figure 3). Since the classification metrics operate on thresholded pixels, this may indicate that the model can precisely predict the presence of signal, but not always its exact continuous intensity value.

Trans-channel learning is a tool for hypothesis-guided biological discovery. When a new biologically informative channel can be reliably predicted on an archival dataset, it decodes actionable high-content signal that can guide compound prioritization and rescue missed opportunities hidden in completed HCSs for drug discovery. Conversely, we may likewise attempt to learn one marker from another in order to *falsify* hypotheses about the relatedness of biological processes (Supplementary Figure 5). We hope that the techniques we make available here, which may be attempted on any archived high-content screen, will be of broad use to the microscopy, screening, and drug discovery communities.

## Methods

### Tg2541 Mouse Line

For all of the key resources and materials used for this study see Supplementary Table 4. The Tg2541 transgenic mouse line expressed the 0N4R isoform of human tau, under

control of the murine neuron-specific Thy1.2 genetic promoter<sup>53</sup>. Tg2541 mice were originally generated on a mixed C57BL/6J×CBA/Ca background<sup>53</sup> and were then bred onto a C57BL/6J background using marker-assisted backcrossing for eight generations before intercrossing to generate Tg2541 homozygous congenic mice. Homozygous Tg2541 mice on a congenic C57BL/6J background were kindly provided by Dr. Michel Goedert (Medical Research Council, Cambridge, UK). The mice were maintained at room temperature in a facility accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International in accordance with the Guide for the Care and Use of Laboratory Animals. All procedures for animal use were approved by the University of California, San Francisco's Institutional Animal Care and Use Committee.

### Phosphotungstic Acid Precipitation of Tau in Tg2541 Mice

Pooled brain homogenate from 6- to 7-month-old Tg2541 mice (both sexes) was prepared as previously described<sup>25</sup>. Briefly, a 10% (weight/volume) homogenate in DPBS was prepared using a rotor-stator type tissue homogenizer (Omni International). Phosphotungstic acid (PTA) precipitation of the brain homogenate was then performed as previously described<sup>12,54</sup>. Briefly, 10% brain homogenate was incubated with 2% sarkosyl (Sigma Aldrich) and 0.5% benzonase (Sigma Aldrich) at 37°C with constant agitation at 1,200 rotations per minute for 2 hours on an orbital shaker. Sodium PTA was dissolved in water and the pH was adjusted to 7.0. A final concentration of 2% sodium PTA was added to the brain homogenate and incubated overnight under the same agitation conditions. The brain homogenate was then centrifuged at 16,100×g for 30 minutes at room temperature, and the supernatant was removed. The pellet was resuspended in 2% sarkosyl and 2% sodium PTA in DPBS and incubated for 1 hour at room temperature. The sample was centrifuged again under the same conditions, the supernatant was removed, and the pellet was resuspended in DPBS, using 10% of the initial starting volume, and stored at -80°C until further use.

### Training Dataset for the Cellular Tau(P301S)-YFP Assay

Tau(P301S)-YFP cells were developed by transfecting human embryonic kidney cells (HEK293T female; ATCC) by transient transfection using Lipofectamine 2000 (ThermoFisher) to overexpress the full-length 0N4R isoform of human tau containing the familial disease-linked P301S missense mutation and the yellow fluorescent protein (YFP) fused to the C-terminus for visualization. A stable monoclonal line was maintained in Dulbecco's modified enriched medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. Tau(P301S)-YFP cells in confluent flasks were collected using trypsin and resuspended in electroporation buffer at a concentration of  $8.4 \times 10^7$  cells/mL. HEK293T cells that overexpress the microtubule-binding repeat domain (RD) of 4R human tau with the P301L and V337M mutations<sup>13</sup> (TauRD(P301L/V337M)-YFP cells) were maintained in DMEM supplemented with 10% FBS and 1% penicillin/streptomycin. Tau seeds were prepared by passaging PTA-precipitated tau protein from Tg2541 mouse brain samples once through TauRD(P301L/V337M)-YFP cells and then collecting the cell lysate, as described previously<sup>25</sup>. All cells were cultured at 37°C and 5% CO<sub>2</sub> in a humidified incubator.

Cell lysate containing tau seeds was combined with the Tau(P301S)-YFP cells at 1.8 µg/mL and electroporation was performed on an Amaxa Nucleofactor instrument (Lonza Bioscience) using the program #Q-001. Seeded Tau(P301S)-YFP cells were then plated at  $5 \times 10^3$  cells/well in a 384-well plate (Greiner) coated with poly-L-ornithine (PLO, 30–70kDa; Sigma Aldrich). To coat plates, PLO was dissolved at 1 mg/mL in 0.15 M sodium borate buffer pH 8.4 and filter-sterilized, and then further diluted to 0.1 mg/ml in sterile deionized (DI) water. The PLO coating solution was added to the culture plate wells and incubated in culture plate wells for six hours at room temperature, then removed. The wells were washed three times with sterile DI water and air-dried in a biosafety cabinet.

Immediately following cell plating, small-molecule drug compounds prepared in pure DMSO were added at a fixed concentration of 10 µM. The small molecules were selected to represent the range of effects on tau aggregation and cell viability by compounds in drug libraries of interest, such as the ChemBridge CNS-Focused Research Library. We used six different compounds to capture three different phenotypic spaces: reducing intracellular tau aggregation without affecting cell viability; reducing tau aggregation and reducing cell viability; and increasing tau aggregation and reducing cell viability.

Actual compound identifiers are proprietary and are obfuscated to identifiers in the form “DRW#”. Two compounds were selected for reducing tau aggregation and also reducing cell viability (DRWT1 and DRWT2); and two compounds were selected for increasing tau aggregation and reducing cell viability (DRWT3 and DRWT4); two compounds were selected for reducing intracellular tau aggregation without affecting cell viability (DRWT5 and DRWT6). Control wells were also prepared with cells electroporated with or without tau seeds and the same volume of DMSO but no drug compound.

At four days post-seeding, cells were washed three times with sterile Dulbecco’s Phosphate (DPBS) and then fixed with 4% paraformaldehyde (PFA) in DPBS for 20 minutes at room temperature. Cells were washed with DPBS and then permeabilized with 0.1% Triton X-100 in DPBS for 20 minutes at room temperature. Cells were washed and then blocked with 3% bovine serum albumin (BSA; Millipore) in DPBS for one hour at room temperature. Cells were then incubated with mouse monoclonal antibody AT8 (1:500; Thermo Fisher Scientific) and rabbit polyclonal antibody MCM2 (1:500; Abcam) in 3% BSA in DPBS overnight at 4°C. Cells were washed with DPBS and then incubated with Alexa Fluor 594 Plus-conjugated goat anti-mouse and anti-rabbit secondary antibodies (1:500; Thermo Fisher Scientific) in 3% BSA in DPBS for 90 min at room temperature. Cells were washed with DPBS and then covered with DPBS and imaged on an InCell Analyzer 6000 High-Content Confocal Microscopy System (GE Healthcare). Twenty-five non-overlapping, 2048×2048 pixel fields were captured per well. We chose markers with minimal overlap of the excitation spectra to minimize potential image bleed-through (Supplementary Figure 8A). This full training dataset was uploaded here for open access (DOI: 10.17605/OSF.IO/XNTD6).

### Constructing the Archival HCS Dataset

The setup for the archival HCS data was the same setup used to construct the training set (the cellular Tau(P301S)-YFP aggregation assay), with the exceptions that the HCS data did not undergo any of the immunocytochemistry steps, and live-cell imaging was used

instead. We constructed this HCS dataset more than one year prior to the three-channel training dataset. Only two imaging channels were captured: nuclear DAPI and YFP-tau. No AT8 antibody was plated. Thus no AT8-pTau channel was captured, and there was no possibility of bleed-through. Furthermore, the compound library in the HCS data was much larger and more diverse than that used for model training. The original six compounds used in the training set with AT8 were not included in the random HCS subset of 1,600 unique compounds evaluated by the ML model.

The screened compound collection is from the Chembridge Central Nervous System (CNS) library with a multiple parameter optimization score greater than four<sup>55,56</sup>. This selection was driven by our desire to find compounds that would be more easily progressed and developed as a CNS drug.

### Training and Evaluation of the Models on Tauopathy Dataset

The full source code and fully trained models are available at <https://github.com/keiserlab/trans-channel-paper>. The necessary python packages are PyTorch, pandas, cv2, cupy, numpy, and sklearn. Training and testing of the model was performed on one Nvidia GeForce GTX 1080 graphics card, and completed in 30 epochs totaling 111 hours of training.

The full dataset consisted of six different drug perturbations: two reduced aggregation while maintaining cell count; two reduced aggregation while reducing cell count; and two increased aggregation while reducing cell count. We imaged a total of six 384-well plates, one plate for each drug. Each well was divided into four fields and each field was imaged, resulting in a full image dataset consisting of 57,600 TIFF images each of size 2048×2048 pixels for each of the three channels. As a preprocessing step, we linearly scaled all images from the original TIFF range of 0 to 65,535 to the more tractable range of 0 to 255 by dividing each pixel by 65,535 and then multiplying by 255. We converted the resulting images to 32-bit floating point. The images were randomly shuffled, and then split into a 70% training set and 30% test set.

We used the fully convolutional architecture in Figure 2A for training. As inputs to the model, we concatenated the transformed YFP-tau image and the transformed DAPI image into a tensor of size 2×2048×2048 pixels. The model generated a predicted AT8-pTau image of size 2048×2048 pixels.

We used a stochastic gradient descent optimizer with momentum = 0.90, a learning rate = 0.001, and a batch size of one. We trained the model for 30 epochs. We used a negative Pearson correlation loss function for training, in which we minimized the following objective function:

$$\text{NPCC}(A, B) = \frac{-\sum_i (A_i - A_\mu)(B_i - B_\mu)}{\sqrt{\sum_i (A_i - A_\mu)^2 \sum_i (B_i - B_\mu)^2}}$$

such that  $A$  and  $B$  were the images being compared,  $i$  was summed over all pixels,  $A_\mu$  was the average of image  $A$ , and  $B_\mu$  was the average of image  $B$ .

## Novel Model Design Considerations

Our architecture differed from U-Net in several important ways (Supplementary Figure 2). First, the model was tasked with assigning pixel values that resided in a greater range (0 to 65,535) compared to U-Net's binary zero or one objective. Hence we used a negative Pearson correlation loss while U-Net used a cross-entropy loss. Our architecture had fewer hidden layers than U-Net, resulting in an architecture that was 12% the size of U-Net. Also the upsampling procedure was different. We used transposed convolutions followed by bilinear interpolation, while U-Net had two convolutions followed by bilinear interpolation. We used transposed convolutions for learnable upsampling during each convolution operation. Finally, our architecture preserved the original pixel image dimensions, while U-Net returned a smaller mask. Preserving image dimensions allowed for a direct pixel-wise one-to-one mapping from the input to the output.

## Performance Metrics

When constructing the PCC performance metric, we iterated over the test set, flattened all of the images to one dimensional vectors, and then found the PCC of the flattened label image with the flattened ML-predicted image using the Numpy library's `corrcoef` function.

When constructing the MSE metric, we normalized all AT8-pTau and predicted AT8-pTau images to have zero-mean and unit-variance in order to account for differences in underlying pixel value distributions between the label images and the predicted images. Normalizing both distributions placed them in a more comparable regime for calculating MSE.

When calculating MSE, we normalized each image by subtracting the average of that image, and then dividing by the standard deviation. The MSE was determined as follows:

$$\text{MSE}(A, B) = \frac{\sum_j (A_j - B_j)^2}{n}$$

Such that  $A_j$  and  $B_j$  were the normalized images being compared, and  $j$  was the index of the image out of a total of  $n$  test images.

To construct the ROC and PRC curves, we first normalized the images to have zero-mean and unit-variance. We then compared the predicted image to the actual, and chose a pixel threshold  $t$  such that any pixel greater than or equal to  $t$  was considered as positive for signal, and anything lower than  $t$  was negative for signal. We chose an intensity threshold of 1.0 to binarize the image, which retained most of the features of interest and aggregation signal (Supplementary Figure 13).

Hence, we performed pixel-wise classification across the full test set of images. For the thresholds applied to the predicted image, we chose the pixel value range 0 (permissive) to 1,000 (non-permissive) with different increments. In the range  $-1.0$  to  $-0.1$  and the range  $0.1$  to  $1.0$ , we chose incrementing values of  $0.1$ . From the range  $-0.1$  to  $0.1$  we chose a more fine-grained increment with a step size of  $0.01$ . In the range  $1.0$  to  $4.0$  we chose a step size

of 1.0, and from the range 4.0 to 22.0 we chose a step size of 2.0. Finally, we evaluated the pixel intensity threshold equal to 1,000 (no pixels assumed this value).

### Construction and Evaluation of PQML and PQC

From the available YFP-tau fusion images and DAPI images of the HCS subset of 1,600 compounds, we constructed predicted AT8-pTau images for each pair of YFP-tau and DAPI images. We took the YFP and DAPI images and linearly scaled the pixel values to the range 0 to 255 inclusive. We then concatenated the two images together to form a tensor of dimensions  $2 \times 2048 \times 2048$ , and finally inputted this into the trained model. Next, we scaled the output of the model back to the original space by first dividing by 255 and then multiplying by the maximal possible pixel value of 65,535. We then scored these images for tau aggregation using proprietary software from GE, called the “InCell Analyzer,” which yielded a score based on puncta count and area. This same algorithm was used for previous HCS efforts, and it was applied consistently to all images. We obtained each small molecule’s aggregation score by averaging over all of that compound’s images’ aggregation scores—which consisted of averaging over four non-overlapping fields within a single well treated with the compound. We ranked all of the compounds by their aggregation score, with compounds inducing lower aggregation being higher priority than compounds with a higher aggregation score. A lower rank in the queue indicated higher priority (e.g., compounds with ranks first, second, and third were the three highest priority compounds). For constructing PQC, we used the same method as the one to construct PQML, except that we obtained aggregation scores from the original YFP-tau images instead of their corresponding ML-predicted AT8-pTau images.

To calculate enrichment curve AUCs (Figure 5), we analyzed the set of known active compounds ( $n=17$ ) discovered in the study. For each queue, we generated an enrichment curve using the rankings of the active compounds. We calculated AUC by integrating with Numpy’s composite trapezoidal function (`numpy.trapz`). We calculated average active compound ranks by averaging over each active compound’s index in the queue (Figure 5B).

### Experimental Setup for Secondary Dose-Response Experiments

For the dose-response experiments, we performed the same protocol as the cellular Tau(P301S)-YFP aggregation assay, with the exceptions being the drug doses and the compounds that we tested. We plated drug concentrations in half logs from 10 nM to 10  $\mu$ M. We tested the top 40-ranked compounds of PQML and the top 40-ranked compounds of PQC. In accordance with the cellular Tau(P301S)-YFP aggregation assay, we did not use AT8 because it would demand extensive time and labor. For each compound, we independently replicated each concentration in four separate wells. We calculated the average aggregation scores at each concentration, and used these averages to construct a dose-response plot (e.g., Figure 4A).

### Assessing Activity from Dose Response Curves

A compound that succeeded in a secondary dose-response test demonstrated a dose-dependent decrease in aggregation as a function of concentration, while maintaining a therapeutic window that preceded a noticeable drop in cell count. A compound did not need



to induce a perfect sigmoidal shape to be considered active. A compound was considered active if and only if all of the following four conditions were satisfied.

1. The best fit curve strictly had an aggregation score that decreased with increasing concentration of compound.
2. The effective response was at least 6,000 aggregation units. We required this so that only compounds were chosen that induced an effective decrease in aggregation. Medicinal chemists who worked closely on the screen chose 6,000 subjectively as a minimal threshold before we began the analyses. If the curve was monotonically increasing with increasing concentration, this condition was automatically not satisfied, and the compound was considered inactive.
3. The half maximal effective concentration (EC50) was lower than 10  $\mu$ M.
4. The concentration at which aggregation began to be ameliorated at half maximal response was lower than the concentration at which the compound decreased cell count by half. If the cell count trend increased with concentration, or was unaffected, then this condition was automatically satisfied. Otherwise, the logarithm of the EC50 of the aggregation curve minus the logarithm of the EC50 of the cell count curve must be less than  $-0.10$ . We chose  $-0.10$  as a more stringent threshold, as opposed to the difference being anything less than 0.0. This ensured that the compound's concentration at which it exhibited half of its maximum response occurred at a lower concentration than when the compound began to decrease cell count by half, and that this difference was not effectively zero.

### Compound Enrichment Comparison with Other Methods

We compared different methods against the trans-channel ML approach to assess active compound enrichment (Figure 5, Supplementary Figure 11). As a strawman enrichment baseline, we added random noise to the raw YFP-tau HCS images by assigning 10% of the pixels to the maximum pixel value 65,535 (Supplementary Figure 11A). We then reranked the compounds by their new aggregation scores using the same GE software for deriving PQC and PQML.

The next comparison method was a non-ML based image analysis platform called CellProfiler<sup>48</sup>. We set up a pipeline to extract image features from the YFP-tau channel, such as puncta count, intensity metrics, and morphological features (Supplementary Figure 11B). We include the full pipeline (YFP\_only.cpproj) in <https://github.com/keiserlab/trans-channel-paper>. Once the HCS images were featurized, we grouped images belonging to the same compound, and averaged the features to obtain a representation of each compound. We did the same for control wells with neither prion seed nor compound. We had 224 control wells (896 2048 $\times$ 2048 pixel images) evenly split among seven 384-well plates. After averaging to derive a latent representation for each compound and for the control condition, we performed a principal component analysis (PCA) with three principal components. Supplementary Figure 11C shows the 1,600 compounds plus controls plotted in PCA space. The proximity of control conditions to each other indicated a meaningful dimensionality-

reduced representation. To derive a new priority queue of compound ranks, we sorted the 1,600 compounds by their Euclidean distance to the average of the control wells in PCA space, with smaller distances being higher priority. Finally we calculated an enrichment plot for the new compound ranking (Figure 5).

In a third test, we compared the trans-channel method with feature extraction via a deep convolutional autoencoder. We designed an autoencoder similar to the trans-channel architecture (see GitHub repository). We removed skip connections and forward propagated through a bottleneck of size = 2,064,512 (approximately half of the image feature space). We trained the network on the HCS dataset until we obtained a near perfect reconstruction of the input YFP-tau (Supplementary Figure 11D). After training, we extracted the hidden state for each image of the HCS, and averaged the latent representations by compound. As with the CellProfiler analysis, we also derived a representation for each of our control wells, and averaged the latent representations. We then performed a PCA reduction with three principal components. We sorted the 1,600 compounds by their Euclidean distance to the average of the control wells in PCA space, with smaller distances being higher priority (Figure 5).

### Validation Assay of Functional Genomics Screen in U2OS Line

We generated a completely different HCS than the tauopathy study in order to assess the ML method's generalizability. This osteosarcoma dataset was subjected to different biological conditions. U2OS cells (female) expressing a stable CCNB1-GFP construct were plated into 384-well plates with 500 cells per well and reverse transfected with an esiRNA library (10 ng, Sigma Aldrich) using Hiperfect transfection reagent. The U2OS cells were cultured in DMEM medium containing 10% foetal bovine serum. The library can be found at <https://iccb.med.harvard.edu/sigma-esirna-human-1> and also <https://iccb.med.harvard.edu/sigma-esirna-human-2>. Our method was inspired by the assay presented in Bray et al.<sup>57</sup> We performed 16,194 unique functional genomic perturbations. After 72 hours, cells were stained with 5 µg/mL of Hoechst 33342 dye (ThermoFisher), incubated for 60 minutes at 37°C, washed with PBS, fixed in 4% PFA, and scanned on a Thermo Cell Insight NXT high content microscopy system using a 10X objective. We captured the two channels, resulting in 324,989 Hoechst and cyclin-B1 image pairs, each of dimension 1104×1104 pixels.

### Training and Evaluation of ML Models for the U2OS Dataset

We performed two training experiments, one using ablated Hoechst images, and one using the raw Hoechst images. For training with ablated images, we ablated Hoechst at the 95<sup>th</sup> percentile pixel-intensity threshold. Afterwards, we performed the same training procedure as the one used for the tauopathy experiment, except for the following: 1) the input to the model was one dimensional (Hoechst channel) instead of two; 2) we trained for 10 epochs instead of 30; 3) we trained a threefold cross-validation instead of a single training-test split.

For training using the unablated, raw images, we applied the same preprocessing and training procedure as the one used for training with ablated images, except for the following: 1) we left the images intact and did not perform any ablations; 2) we trained for 5 epochs

instead of 10 (the model converged faster likely because the task of learning from raw unablated images was easier).

### Quantification and Statistical Analyses

For all statistical significance tests, we used a two sample, one-sided z test. The null hypothesis stated that performance means were equal. The alternative hypothesis stated that the average performance from ML is greater than the average performance from the non-ML approach. Significance was set at  $p < 0.05$ . The values of  $n$  correspond to the number of instances in the test set (see Methods: Training and Evaluation of the Models on Tauopathy Dataset). Sample sizes were sufficiently large to perform the test. A sample is one comparison between a predicted channel and the actual channel. We assume normality due to the large sample size.

### Code Availability

The full source code and fully trained models are available at <https://github.com/keiserlab/trans-channel-paper><sup>58</sup>, DOI: 10.5281/zenodo.6336183

### Data Availability

All image data is freely available at <https://osf.io/xntd6><sup>59</sup>, DOI: 10.17605/OSF.IO/XNTD6

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgements

This work was supported by grant number 2018–191905 from the Chan Zuckerberg Initiative DAF, an advised fund of the Silicon Valley Community Foundation (MJK), the National Institutes of Health (AG002132) (S.B.P.), as well as by support from the Brockman Foundation (S.B.P.) and the Sherman Fairchild Foundation (S.B.P.).

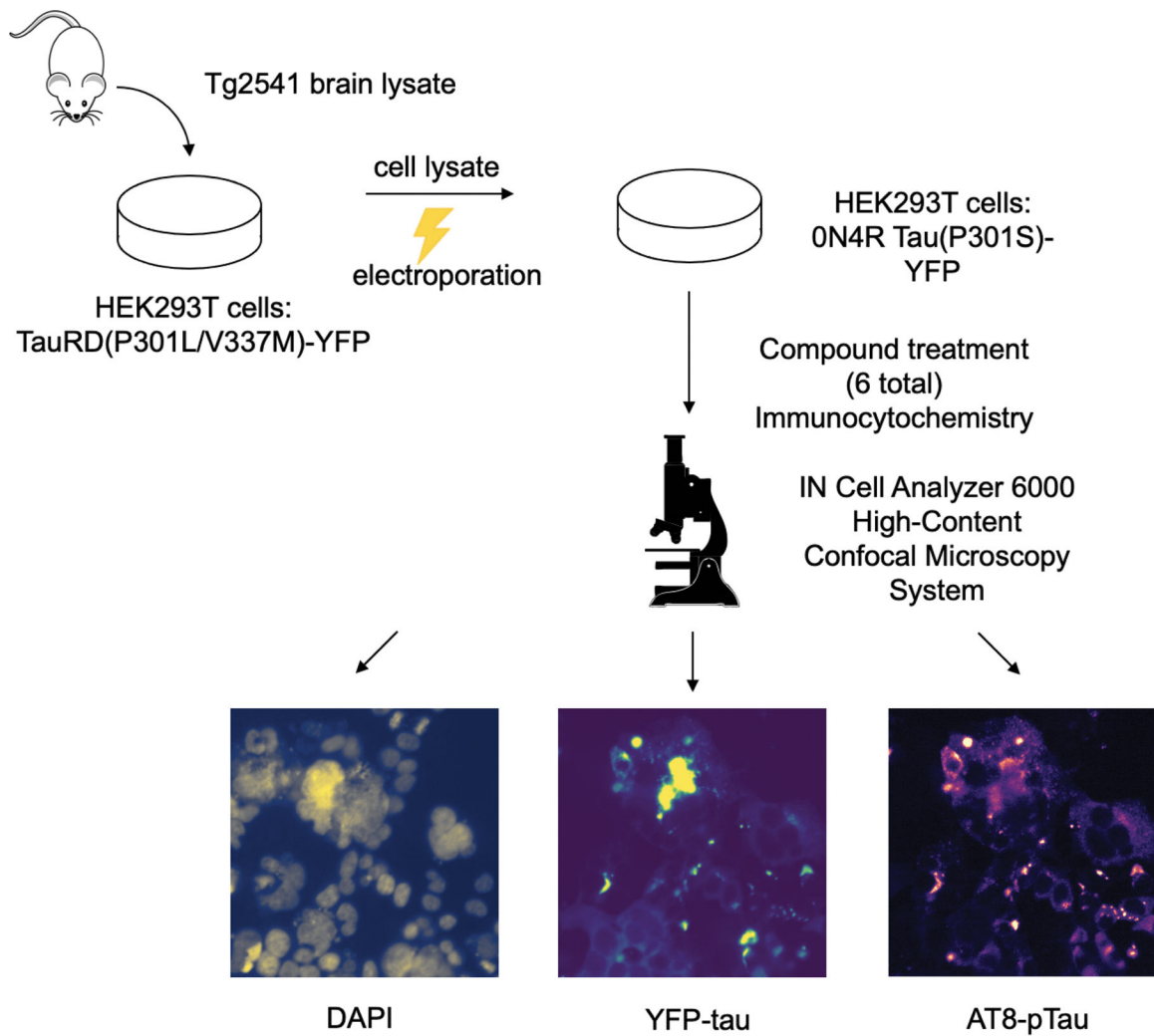
### References

1. Li Z, Cvijic ME & Zhang L 2.15 - Cellular Imaging in Drug Discovery: Imaging and Informatics for Complex Cell Biology. in *Comprehensive Medicinal Chemistry III* (eds. Chackalamannil S, Rotella D & Ward SE) 362–387 (Elsevier, 2017). doi:10.1016/B978-0-12-409547-2.12328-5.
2. Kim S-W, Roh J & Park C-S Immunohistochemistry for Pathologists: Protocols, Pitfalls, and Tips. *J Pathol Transl Med* 50, 411–418 (2016). [PubMed: 27809448]
3. Cardoso MC Fluorescence Microscopy: Spectral Imaging vs. Filter-based Imaging. in *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine* 583–586 (Springer Berlin Heidelberg, 2006). doi:10.1007/3-540-29623-9\_5560.
4. Lao K et al. Drug development for Alzheimer's disease. *J. Drug Target* 27, 164–173 (2019). [PubMed: 29732929]
5. Cummings J, Lee G, Ritter A, Sabbagh M & Zhong K Alzheimer's disease drug development pipeline: 2019. *Alzheimers. Dement* 5, 272–293 (2019).
6. Iqbal K, Liu F & Gong C-X Tau and neurodegenerative disease: the story so far. *Nat. Rev. Neurol* 12, 15–27 (2016). [PubMed: 26635213]
7. Eckermann K et al. The  $\beta$ -Propensity of Tau Determines Aggregation and Synaptic Loss in Inducible Mouse Models of Tauopathy. *J. Biol. Chem* 282, 31755–31765 (2007). [PubMed: 17716969]

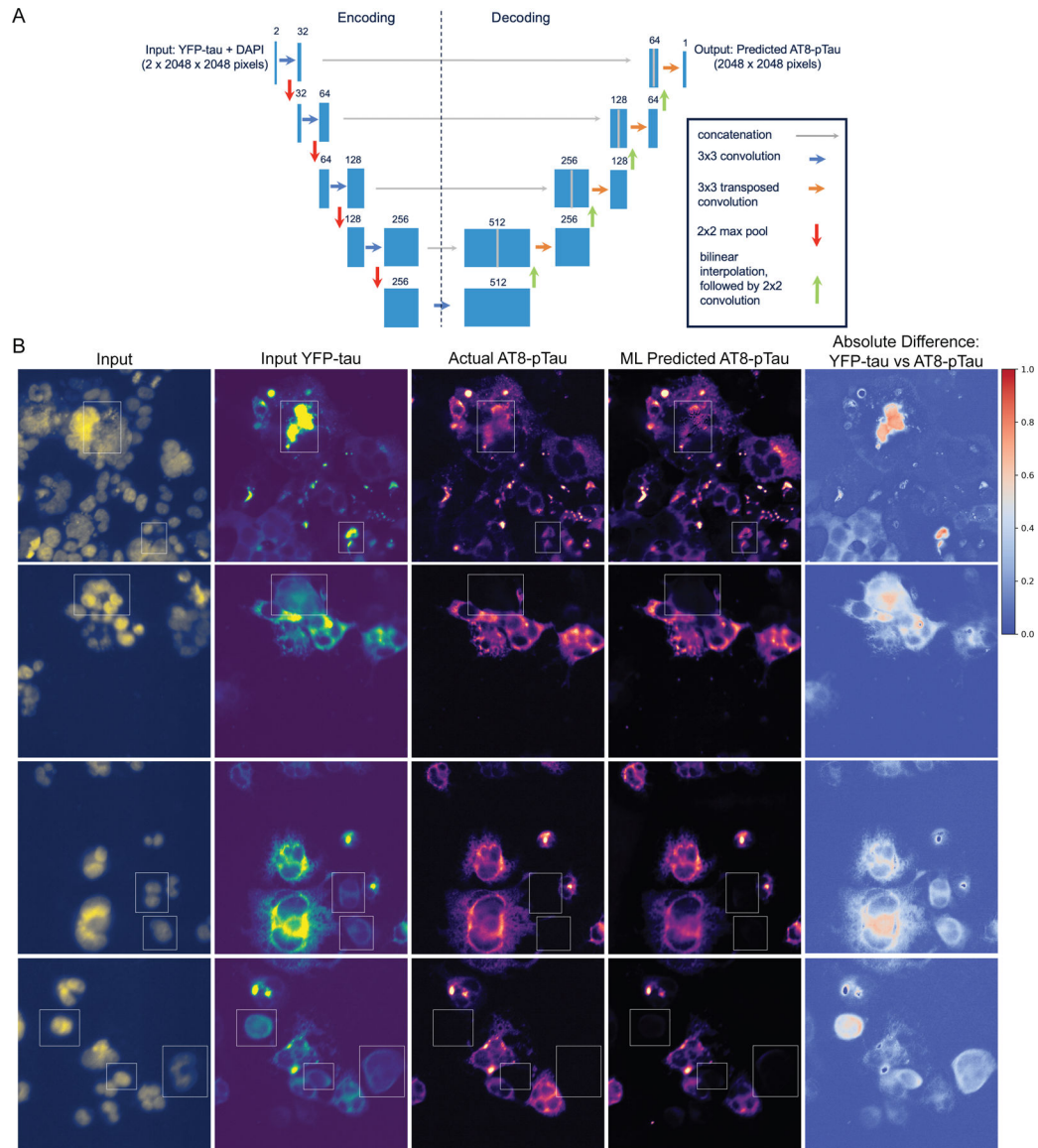
8. Fatouros C et al. Inhibition of tau aggregation in a novel *Caenorhabditis elegans* model of tauopathy mitigates proteotoxicity. *Hum. Mol. Genet* 21, 3587–3603 (2012). [PubMed: 22611162]
9. Goedert M, Clavaguera F & Tolnay M The propagation of prion-like protein inclusions in neurodegenerative diseases. *Trends Neurosci* 33, 317–325 (2010). [PubMed: 20493564]
10. Jucker M & Walker LC Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. *Nature* 501, 45–51 (2013). [PubMed: 24005412]
11. Clavaguera F et al. Transmission and spreading of tauopathy in transgenic mouse brain. *Nat. Cell Biol* 11, 909–913 (2009). [PubMed: 19503072]
12. Aoyagi A et al. A $\beta$  and tau prion-like activities decline with longevity in the Alzheimer's disease human brain. *Science Translational Medicine* vol. 11 eaat8462 (2019). [PubMed: 31043574]
13. Sanders DW et al. Distinct tau prion strains propagate in cells and mice and define different tauopathies. *Neuron* 82, 1271–1288 (2014). [PubMed: 24857020]
14. Jackson SJ et al. Short Fibrils Constitute the Major Species of Seed-Competent Tau in the Brains of Mice Transgenic for Human P301S Tau. *J. Neurosci* 36, 762–772 (2016). [PubMed: 26791207]
15. Furman JL et al. Widespread tau seeding activity at early Braak stages. *Acta Neuropathol* 133, 91–100 (2017). [PubMed: 27878366]
16. Despres C et al. Identification of the Tau phosphorylation pattern that drives its aggregation. *Proc. Natl. Acad. Sci. U. S. A* 114, 9080–9085 (2017). [PubMed: 28784767]
17. Lai R, Harrington C & Wischik C Erratum: Lai RYK; Harrington CR; Wischik CM Absence of a Role for Phosphorylation in the Tau Pathology of Alzheimer's Disease. *Biomolecules* 2016, 6, 19. *Biomolecules* vol. 6 35 (2016).
18. Goedert M et al. Assembly of microtubule-associated protein tau into Alzheimer-like filaments induced by sulphated glycosaminoglycans. *Nature* 383, 550–553 (1996). [PubMed: 8849730]
19. Grundke-Iqbal I et al. Abnormal phosphorylation of the microtubule-associated protein tau ( $\tau$ ) in Alzheimer cytoskeletal pathology. *Proc. Natl. Acad. Sci. U. S. A* 83, 4913–4917 (1986). [PubMed: 3088567]
20. Sengupta A et al. Phosphorylation of tau at both Thr 231 and Ser 262 is required for maximal inhibition of its binding to microtubules. *Arch. Biochem. Biophys* 357, 299–309 (1998). [PubMed: 9735171]
21. Alonso AC, Zaidi T, Grundke-Iqbal I & Iqbal K Role of abnormally phosphorylated tau in the breakdown of microtubules in Alzheimer disease. *Proc. Natl. Acad. Sci. U. S. A* 91, 5562–5566 (1994). [PubMed: 8202528]
22. Lindwall G & Cole RD Phosphorylation affects the ability of tau protein to promote microtubule assembly. *J. Biol. Chem* 259, 5301–5305 (1984). [PubMed: 6425287]
23. Johnson GVW & Stoothoff WH Tau phosphorylation in neuronal cell function and dysfunction. *J. Cell Sci* 117, 5721–5729 (2004). [PubMed: 15537830]
24. Gong C-X & Iqbal K Hyperphosphorylation of Microtubule-Associated Protein Tau: A Promising Therapeutic Target for Alzheimer Disease. *Curr. Med. Chem* 15, 2321–2328 (2008). [PubMed: 18855662]
25. Grandjean J-MM et al. Discovery of 4-Piperazine Isoquinoline Derivatives as Potent and Brain-Permeable Tau Prion Inhibitors with CDK8 Activity. *ACS Med. Chem. Lett* 11, 127–132 (2020). [PubMed: 32071678]
26. Preuss U, Döring F, Illenberger S & Mandelkow EM Cell cycle-dependent phosphorylation and microtubule binding of tau protein stably transfected into Chinese hamster ovary cells. *Mol. Biol. Cell* 6, 1397–1410 (1995). [PubMed: 8573794]
27. Malia TJ et al. Epitope mapping and structural basis for the recognition of phosphorylated tau by the anti-tau antibody AT8. *Proteins: Struct. Funct. Bioinf* 84, 427–434 (2016).
28. Goedert M, Jakes R & Vanmechelen E Monoclonal antibody AT8 recognises tau protein phosphorylated at both serine 202 and threonine 205. *Neurosci. Lett* 189, 167–169 (1995). [PubMed: 7624036]
29. Duka V et al. Identification of the sites of tau hyperphosphorylation and activation of tau kinases in synucleinopathies and Alzheimer's diseases. *PLoS One* 8, e75025 (2013). [PubMed: 24073234]

30. Jensen EC Overview of live-cell imaging: requirements and methods used. *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology* 296, 1–8 (2013). [PubMed: 22907880]
31. Sung M-H & McNally JG Live cell imaging and systems biology. *Wiley Interdiscip. Rev. Syst. Biol. Med* 3, 167–182 (2011). [PubMed: 20730797]
32. Scalable Production of iPSC-Derived Human Neurons to Identify Tau-Lowering Compounds by High-Content Screening. *Stem Cell Reports* 9, 1221–1233 (2017). [PubMed: 28966121]
33. Azorsa DO et al. High-content siRNA screening of the kinome identifies kinases involved in Alzheimer's disease-related tau hyperphosphorylation. *BMC Genomics* 11, 1–10 (2010). [PubMed: 20044946]
34. Assessing fibrinogen extravasation into Alzheimer's disease brain using high-content screening of brain tissue microarrays. *J. Neurosci. Methods* 247, 41–49 (2015). [PubMed: 25813427]
35. Vatansever S et al. Artificial intelligence and machine learning-aided drug discovery in central nervous system diseases: State-of-the-arts and future directions. *Med. Res. Rev* 41, 1427–1473 (2021). [PubMed: 33295676]
36. Kim S-H et al. Prediction of Alzheimer's disease-specific phospholipase c gamma-1 SNV by deep learning-based approach for high-throughput screening. *Proc. Natl. Acad. Sci. U. S. A* 118, (2021).
37. Bermudez-Lugo A, J., C. Rosales-Hernandez M, Deeb O, Trujillo-Ferrara J & Correa-Basurto J In Silico Methods to Assist Drug Developers in Acetylcholinesterase Inhibitor Design. *Curr. Med. Chem* 18, 8 (2011).
38. Basile L Virtual Screening in the Search of New and Potent Anti-Alzheimer Agents. in *Computational Modeling of Drugs Against Alzheimer's Disease* 107–137 (Humana Press, New York, NY, 2018). doi:10.1007/978-1-4939-7404-7\_4.
39. Kristy A. Carpenter XH. Machine Learning-based Virtual Screening and Its Applications to Alzheimer's Drug Discovery: A Review. *Curr. Pharm. Des* 24, 3347 (2018). [PubMed: 29879881]
40. Christiansen EM et al. In Silico Labeling: Predicting Fluorescent Labels in Unlabeled Images. *Cell* 173, 792–803.e19 (2018). [PubMed: 29656897]
41. Ounkomol C, Seshamani S, Malekar MM, Collman F & Johnson GR Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. *Nat. Methods* 15, 917–920 (2018). [PubMed: 30224672]
42. Moen E et al. Deep learning for cellular image analysis. *Nat. Methods* 16, 1233–1246 (2019). [PubMed: 31133758]
43. Ronneberger O, Fischer P & Brox T U-Net: convolutional networks for biomedical image segmentation. *MICCAI* (2015).
44. Liu G et al. Image inpainting for irregular holes using partial convolutions in Proceedings of the European Conference on Computer Vision (ECCV) 85–100 (2018).
45. Caicedo JC et al. Data-analysis strategies for image-based cell profiling. *Nat. Methods* 14, 849–863 (2017). [PubMed: 28858338]
46. Hughes JP, Rees S, Kalindjian SB & Philpott KL Principles of early drug discovery. *Br. J. Pharmacol* 162, 1239 (2011). [PubMed: 21091654]
47. Huang N, Shoichet BK & Irwin JJ Benchmarking sets for molecular docking. *J. Med. Chem* 49, 6789–6801 (2006). [PubMed: 17154509]
48. Carpenter AE et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* 7, 1–11 (2006).
49. Müllers E, Cascales HS, Burdova K, Macurek L & Lindqvist A Residual Cdk1/2 activity after DNA damage promotes senescence. *Aging Cell* vol. 16 575–584 (2017). [PubMed: 28345297]
50. Chuang KV & Keiser MJ Comment on 'Predicting reaction performance in C-N cross-coupling using machine learning'. *Science* vol. 362 (2018).
51. Soekhoe D, van der Putten P & Plaat A On the Impact of Data Set Size in Transfer Learning Using Deep Neural Networks. in *Advances in Intelligent Data Analysis XV* 50–60 (Springer International Publishing, 2016). doi:10.1007/978-3-319-46349-0\_5.
52. Williams E et al. Image Data Resource: a bioimage data integration and publication platform. *Nat. Methods* 14, 775–781 (2017). [PubMed: 28775673]

53. Allen B et al. Abundant tau filaments and nonapoptotic neurodegeneration in transgenic mice expressing human P301S tau protein. *J. Neurosci* 22, 9340–9351 (2002). [PubMed: 12417659]
54. Lee IS, Long JR, Prusiner SB & Safar JG Selective precipitation of prions by polyoxometalate complexes. *J. Am. Chem. Soc* 127, 13802–13803 (2005). [PubMed: 16201796]
55. Wager TT, Hou X, Verhoest PR & Villalobos A Central Nervous System Multiparameter Optimization Desirability: Application in Drug Discovery. *ACS Chem. Neurosci* 7, 767–775 (2016). [PubMed: 26991242]
56. Wager TT, Hou X, Verhoest PR & Villalobos A Moving beyond rules: the development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties. *ACS Chem. Neurosci* 1, 435–449 (2010). [PubMed: 22778837]
57. Bray M-A et al. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat. Protoc* 11, 1757–1774 (2016). [PubMed: 27560178]
58. Keiser M keiserlab/trans-channel-paper: v1.0.0 (2022) doi:10.5281/zenodo.6336183.
59. Wong D & Keiser M Trans-channel fluorescence learning (2020) doi:10.17605/OSF.IO/XNTD6.



**Figure 1: We constructed a cellular Tau(P301S)-YFP aggregation assay that modeled prion propagation *in vitro* with an additional AT8-pTau channel for ML training.** An overview of the experiment used to model the adverse propagation and aggregation of tau. We used brain lysate from a tauopathy mouse model to infect a human cellular model of mutant tau that has a higher propensity to aggregate. The experiment yielded three channels used for training the ML model: DAPI, YFP-tau, and AT8-pTau. Images were enhanced with ImageJ's auto-enhance feature solely for visualization purposes.



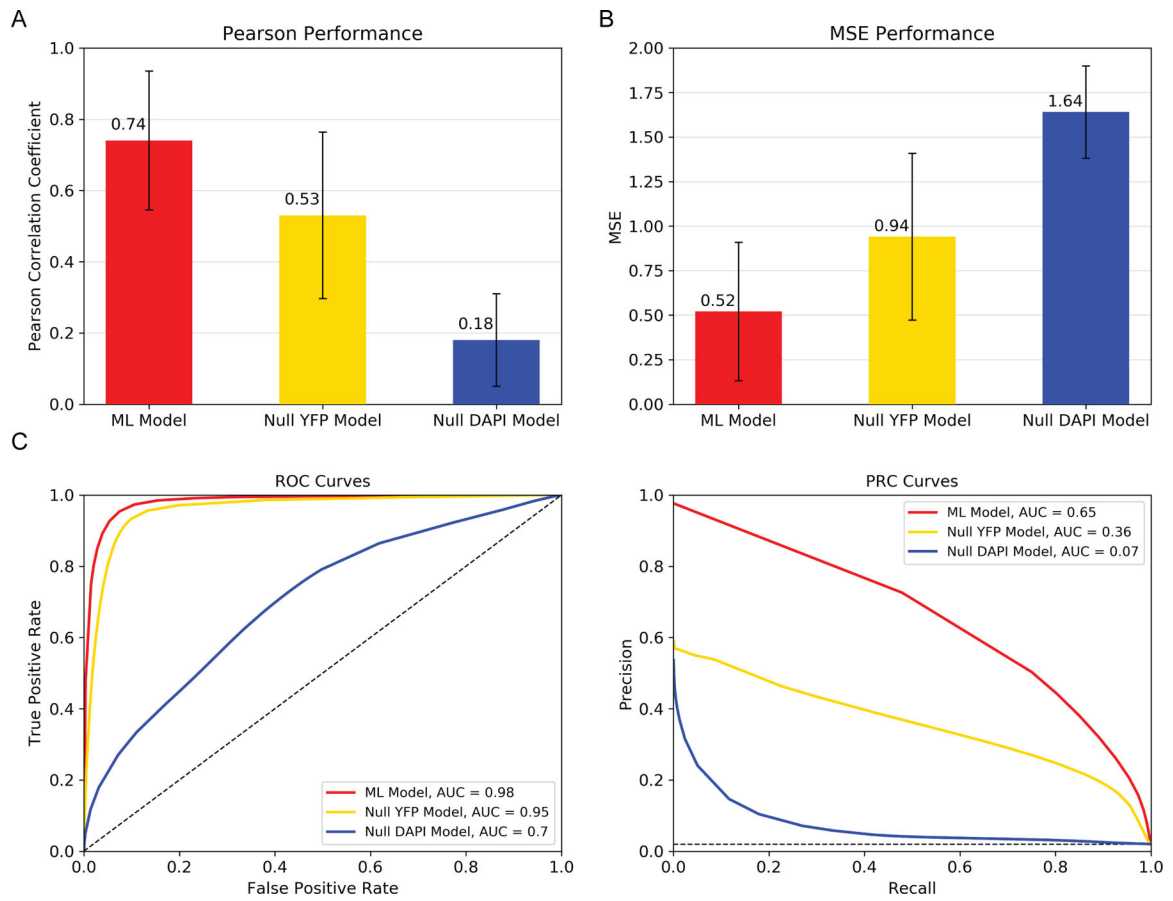
**Figure 2: The ML model generated a phosphorylated AT8-pTau channel, solely from DAPI and YFP-tau channels.**

(a) Given two image channels (DAPI and YFP-tau) as input, the neural network predicted the desired AT8-pTau image. The architecture, motivated by U-Net<sup>43</sup>, comprised two phases: an “encoding” half which increased the cardinality of convolutional filters while reducing the image dimensions, and a “decoding” half which decreased the cardinality of filters while upsampling the image dimensions with bilinear interpolation and learnable transposed convolutions. The numbers above each box indicate the number of filters used.

(b) From the input DAPI image (left column) and the input YFP-tau image (second column), the *in silico* predicted AT8-pTau image (fourth column) is shown versus the actual, hidden AT8-pTau image (third column). White boxes mark where the AT8-pTau images most diverged from YFP-tau on visual inspection. These held-out test images were not seen during model training. For visualization purposes only, images were cropped to 512×512 pixels, auto-enhanced using ImageJ, and colored using the python library Matplotlib. PCC



values between the predicted and actual images displayed are 0.88, 0.92, 0.93, and 0.86 from the top row to the bottom. The rightmost column is a heatmap of the absolute value difference between the input YFP-tau channel and the actual AT8-pTau channel after scaling both images to the range zero to one. Red regions indicate areas of lowest image similarity, while blue regions indicate regions of highest similarity.

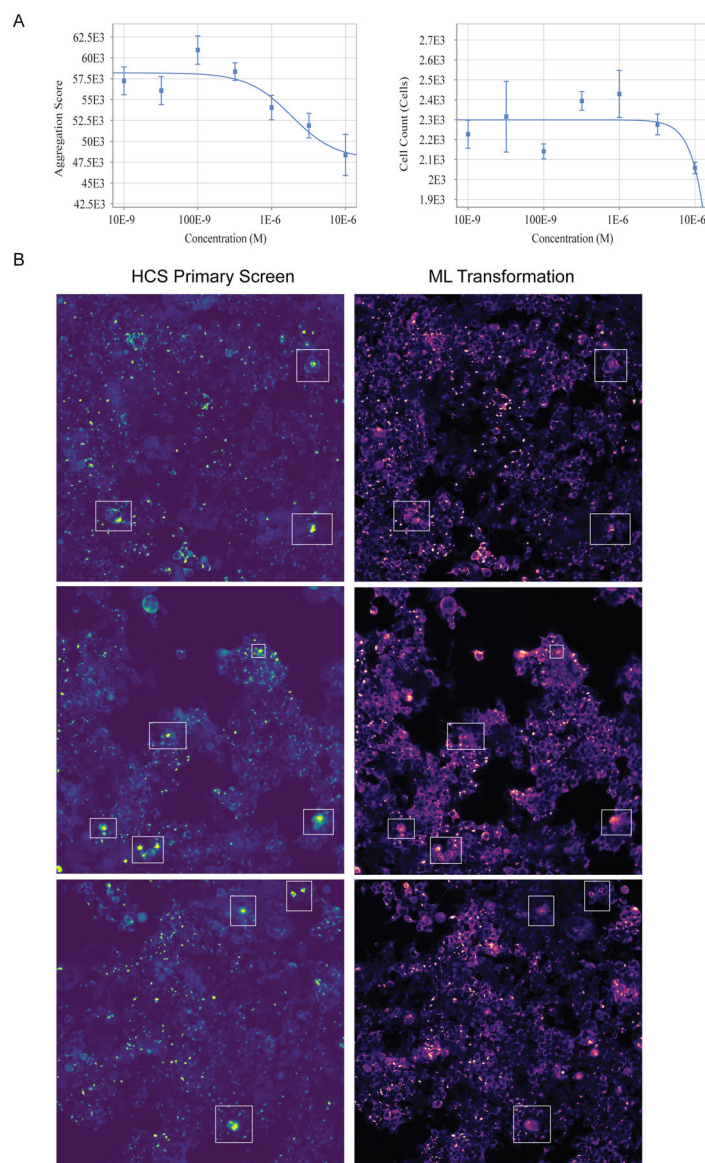


**Figure 3: The ML model outperformed all others.**

(a) The average PCC performance over the held-out test set for three models: the ML model and two null models derived from each of the input channels. Error bars show one standard deviation in each direction centered at the mean. Assessing the ML model's superior performance over each of the null models resulted in statistical significance of  $p < 0.00001$ .

(b) Equivalent MSE results after normalizing all images to zero-mean and unit-variance (Methods).

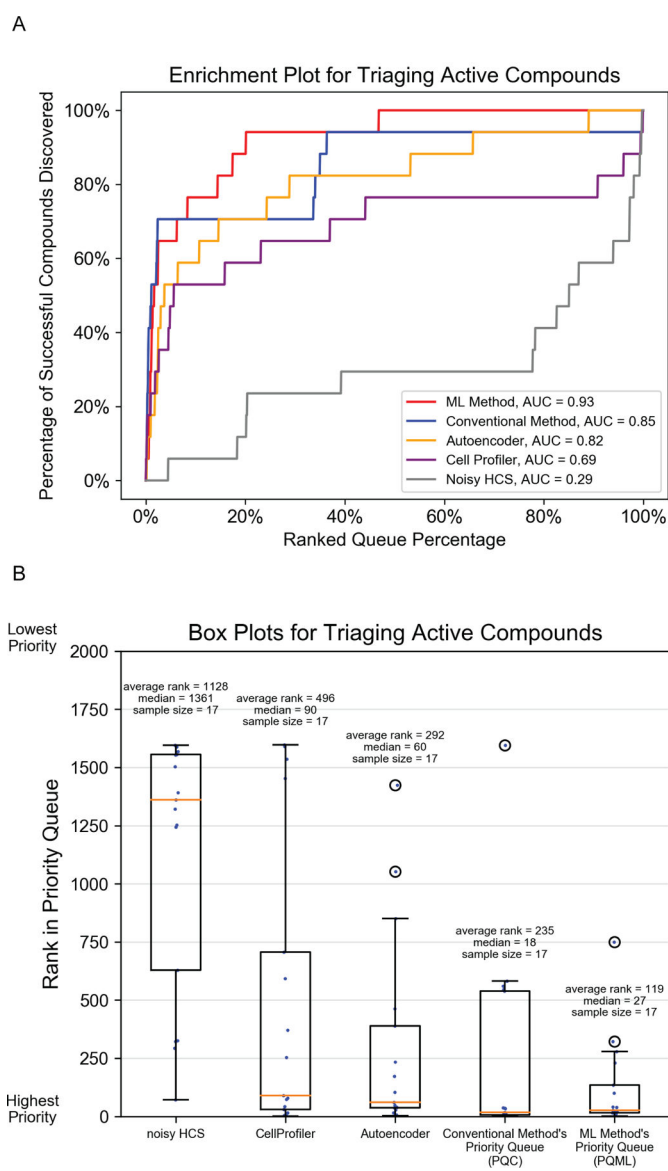
(c) A binary pixel threshold of 1.0 was used to binarize the label image, while the predicted images were assessed across a range of pixel thresholds (Methods). Left: the receiver operating characteristic (ROC) curves for the ML model (red) exceed the Null YFP Model (yellow) and the Null DAPI Model (blue). Right: Precision recall curves. The dashed horizontal line is the percentage of positive pixels (0.02) after thresholding the labels. Both ML and Null YFP models did similarly well at balancing true positives with false positives, but the ML model consistently maintained higher precision.



**Figure 4: Transforming the archival HCS data with ML-predicted AT8-pTau images revealed previously unknown active compounds.**

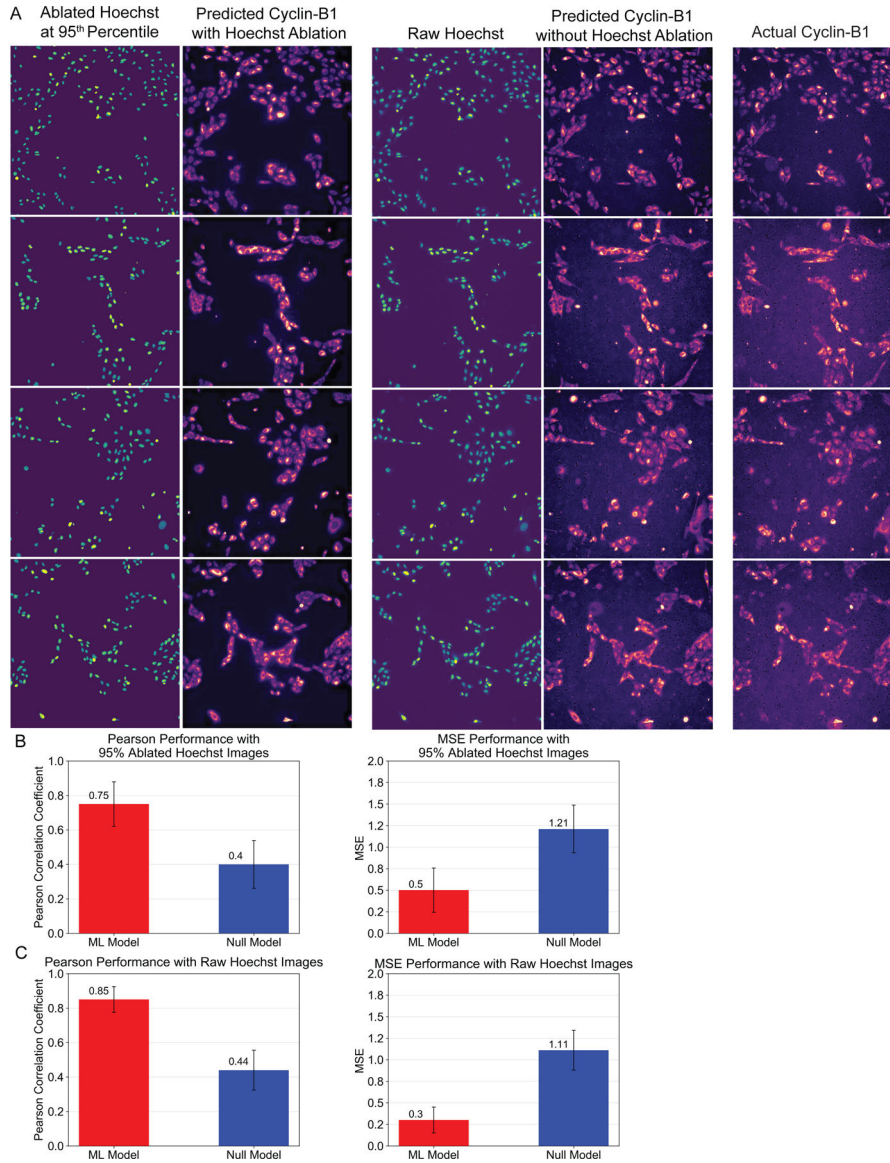
(a) We rescored a plate consisting of 1,600 compounds from the archival HCS, and prospectively collected dose-responses for compounds that resulted in the lowest tau aggregation according to either PQML or PQC. Example dose-response (left) and cell count assay (right) of an active compound. Although an indicator of cell viability, cell count is not a perfect metric for toxicity, since it is possible for the compounds to inhibit the innately prolific nature of HEK293T cells. These curves are drawn from the compound DRW545, which was ranked well by the ML method (22<sup>nd</sup>), but poorly (545<sup>th</sup>) by the conventional method. (b) Example images for active compounds that the ML-based rescoring rescued. Top row: Compound DRW1596 (ranked 1,596<sup>th</sup> in PQC versus 14<sup>th</sup> in PQML), Middle: compound DRW560 (ranked 560<sup>th</sup> in PQC versus 15<sup>th</sup> in PQML), Bottom: compound DRW545 (ranked 545<sup>th</sup> in PQC versus 22<sup>nd</sup> in PQML). YFP-tau images from the archival HCS dataset (left column) differed from the ML-predicted AT8-pTau images used for

rescoring the HCS (right column), with white boxes highlighting example regions. In these regions of interest, the ML predicted the aggregate signal as not being phosphorylated at the residue of interest. We auto-enhanced and colored images with ImageJ and Matplotlib solely for visualization.



**Figure 5: ML triaged active compounds better than other methods.**

(a) Enrichment plots comparing the conventional compound testing method and several alternatives with the prospective, trans-channel ML-guided dose-response assays. We considered the set of all active compounds ( $n=17$ ). The y-axis shows what fraction of active compounds were discovered in the top x% of the ranked priority queue (x-axis) (100% corresponds to rank 1,600). The ML method's area under the enrichment curve (AUC) indicates that its compound ranking achieved the highest success rate on prospective dose-response testing. (b) Box plots of rank accounting for all active compounds. The median line is in orange, whiskers are set at 1.5x interquartile range, and outliers are anything outside the 1.5x interquartile range (plotted as dots). PQML on average prioritizes active compounds about twice as well as PQC.



**Figure 6: The trans-fluorescence learning method generated accurate cyclin-B1 signal from Hoechst signal, on an independent and biologically unrelated dataset.** (a) Given ablated Hoechst images (leftmost column) a ML model trained to output cyclin-B1 images (rightmost column) achieved high *in silico* prediction of this channel (second column). Given unmodified Hoechst stain images (third column), a separate ML model constructed high-fidelity and detailed predictions (fourth column) of the actual cyclin-B1 channel. This panel contains images from the test set ( $n=108,330$  images) never seen during model training. For visualization, all images were auto-enhanced and colored as in prior Figures. (b) Assessing the Hoechst-ablated model’s performance. Left: average PCC coefficient when assessing pixel-wise similarity between actual and predicted Cyclin-B1 images in threefold cross-validation of a dataset of  $n=324,989$  total image pairs, across  $g=16,194$  unique functional genomic perturbations; Right: MSE from the same analysis. Assessing the ML Model’s superior average performance over the Null Model resulted in statistical significance of  $p < 0.00001$ . (c) Assessing the model trained using raw, unablated

Hoechst images, performing the same analysis as in (b). The ML Model's superior average performance over the Null Model likewise resulted in  $p < 0.00001$ . Error bars show one standard deviation in each direction centered at the mean.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript