Published in final edited form as:

Adv Exp Med Biol. 2016 ; 854: 597–603. doi:10.1007/978-3-319-17121-0_79.

Use of a machine learning-based high content analysis approach to identify photoreceptor neurite promoting molecules

John A. Fuller¹, Cynthia A. Berlinicke¹, James Inglese^{2,3}, and Donald J. Zack^{1,4,5}

¹Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, MD

²National Center for Advancing Translational Sciences, NIH, Rockville, MD

³National Human Genome Institute, NIH, Bethesda, MD

⁴Departments of Molecular Biology and Genetics, Neuroscience, and Institute of Genetic Medicine, The Johns Hopkins University School of Medicine, Baltimore, MD

⁵Institut de la Vision, Paris 75012, France

Abstract

Purpose—High content analysis (HCA) has become a leading methodology in phenotypic drug discovery efforts. Typical HCA workflows include imaging cells using an automated microscope and analyzing the data using algorithms designed to quantify one or more specific phenotypes of interest. Due to the richness of high content data, unappreciated phenotypic changes may be discovered in existing image sets using interactive machine-learning based software systems.

Methods—Primary postnatal day four retinal cells from the photoreceptor (PR) labeled QRX-EGFP reporter mice were isolated, seeded, treated with a set of 234 profiled kinase inhibitors and then cultured for one week. The cells were imaged with an Acumen plate-based laser cytometer to determine the number and intensity of GFP-expressing, i.e. PR, cells. Wells with values greater than twice the standard deviation of vehicle-treated wells for either number of objects expressing GFP or mean GFP total object intensities above an average level of 8877 RFU were re-imaged at a higher resolution with an INCell2000 automated microscope. The INCell images were analyzed with an open source HCA analysis tool, PhenoRipper (Rajaram et al., 2012) to identify the high GFP-inducing treatments that additionally resulted in the most diverse phenotypes compared to the vehicle control samples.

Results—The pyrimidinopyrimidone kinase inhibitor CHEMBL-1766490, a pan kinase inhibitor whose major known targets are p38 α and the Src family member Ick, was identified as an inducer of photoreceptor neuritogenesis by PhenoRipper, a cell-independent image analysis tool. This finding was corroborated using an alternate cell-based method of image analysis that measures quantitative differences in the mean neurite length in GFP expressing cells.

Conclusions—Interacting with data using machine learning algorithms may complement traditional HCA approaches by leading to the discovery of small molecule-induced cellular phenotypes in addition to those upon which the investigator is initially focusing.

XX.1 Introduction

A common challenge in HCA phenotypic screening is utilizing the image data to its fullest capacity. Generally, images are analyzed using investigator devised analysis algorithms that measure a certain set of defined parameters. Although this approach is powerful, it limits the assay to combinations of the defined parameters. A growing trend in the field is the use of supervised and unsupervised machine learning approaches that can both accelerate analysis of HCA data sets as well as facilitate discovery of novel induced phenotypes. Several open-source platforms containing machine-learning based methods have been developed and are available either as stand-alone applications such as PhenoRipper and CellCognition (Rajaram et al., 2012; Held et al., 2010) or as modules within popular HCA packages such as CellProfiler and CellProfiler Analyst (Sommer et al., 2011; Carpenter et al., 2006)

We have been using HCA phenotypic screening combined with primary and stem cell-derived retinal cell based assays to identify molecules that promote differentiation and survival of PRs and retinal ganglion cells (Fuller et al., 2014; Welsbie et al., 2013). In this study, we explored the added value of the machine learning approach to analyze images from a PR differentiation/survival screen. We used the publicly available open-source program PhenoRipper (Rajaram et al., 2012) to compare images of treated versus untreated cells. PhenoRipper uses a bag-of-features classification approach to characterize images (Csurka et al., 2004). The method consists of reducing image features to a quantized color state (q-color), characterizing blocks of pixels demonstrating different q-color states, and characterizing and classifying contiguous block (superblock) types across an image (Rajaram et al., 2012; Csurka et al., 2004). This approach does not require definition of cellular parameters. Using this software helped us identify a molecule that promotes PR neuritogenesis in vitro.

XX.2 Materials and Methods

XX.2.1 Primary Cell Dissociation

All animal procedures were performed in accordance with the ARVO statement on the “Use of Animals in Ophthalmic and Vision Research” and were approved by the Institutional Animal Care and Use Committees at the Johns Hopkins University School of Medicine. Retinal cells were isolated and prepared for cell culture and analysis as previously described (Fuller et al., 2014). Briefly, retinas from QRX-IRES-EGFP (114 strain) mice are isolated at postnatal day 4 (Wang et al., 2004). Animals are sacrificed by isoflurane overdose followed by decapitation. Eyes were enucleated, and retinas dissected and dissociated into single cell suspensions by incubation with activated papain in Hibernate-E without Ca^{2+} (BrainBits) for 15 min at 37°C.

XX.2.2 1536 well Cell Plating and Compound Library Treatment

1536 well plates were filled with 4 μL of neuronal culture medium consisting of Neurobasal-E, 2% B-27, 0.5 mM L-Glutamine, and 1X final penicillin/streptomycin (all Life Technologies) using a Multidrop Combi peristaltic bulk dispenser (ThermoFisher). 23 nL of a stock compound solution 10 mM-170 nM in DMSO was transferred to each well in the

culture plate from a library plate using a robotic pintool transfer tool (Wako) resulting in final concentrations ranging from 25 μ M-420 pM. Cells were resuspended at a concentration of 1.25×10^5 cells/mL, filtered, then dispensed into plates at a final well culture volume of 8 μ L (1000 cells/well). Plate were incubated for 7days in 95% humidity at 37°C.

XX.2.3 Fixation and Imaging

Fixation and imaging of the cells were performed as described previously (Fuller et al., 2014). Cells were treated with 2 μ L of 20% (4% final) fresh paraformaldehyde in PBS containing Hoechst 33342 (1:5000 final) for 15 min at room temperature. The wells were then aspirated and rinsed twice with PBS using an EL406 washer dispenser (BioTek Instruments) equipped with an aspirator and syringe dispenser for 1536 well plates, and imaged with an Acumen Explorer (TTP Labtech) plate cytometer equipped with 405, 488, and 633 nm lasers. PRs are defined as GFP positive objects with size and GFP fluorescence intensity above defined threshold values. “On the fly” analysis identified “hit” wells, defined as wells that display a fraction of GFP positive cells greater than 2*SD relative to the vehicle (DMSO) controls or wells with the mean total GFP intensity/object greater than 8877 RFU, a threshold found to be significant in previous screens. Brightfield, nuclei and GFP images of the hit wells were then acquired with a microscope-based INCell2000 HCA platform (GE) using a 20X objective lens.

XX.2.4 Analysis

The INCell images from the Hoechst (nuclei) and GFP (QRX promoter reporter) channels of the control and hit wells were loaded into PhenoRipper (Rajaram et al., 2012). All default parameters for threshold values as well as block size (15 pixels per block) for PhenoRipper were used for analysis. Representative vehicle (DMSO) control wells were used as a replicate image subsampling group. All points that are found on the periphery of the multidimensional scaling (MDS) plot (apparent outliers) were examined and excluded from analysis if deemed to contain artifacts. The putative positive hit and control images were also analyzed using a custom algorithm developed using the Neuronal Profiling 4.1 (ThermoFisher).

X.3 Results

XX.3.1 Photoreceptor neuritogenesis uncovered by machine learning

In order to screen for small molecules that promote photoreceptor differentiation, we developed a high-throughput assay utilizing dissociated retinal cells from a QRX-GFP transgenic mouse that we previously reported (Wang et al., 2004). Our intended readout of the assay was number of cells expressing GFP and expression level per cell. As expected, using this assay we detected a number of molecules that modulated GFP expression, and further characterization of these molecules is underway (Fuller et al., unpublished results).

To complement the GFP intensity measurement algorithm, we also analyzed the cell image data sets using a PhenoRipper-derived MDS plot (Fig 1a). This analysis revealed a superblock (common cell morphology) that appeared to be enriched in wells containing QRX positive neurites (Fig 1b arrow). Cells treated with 6-(2-chlorophenyl)-8-methyl-2-

(oxan-4-ylamino)pyrido[2,3-d]pyrimidin-7-one/CHEMBL-1766490/Pubchem CID 23551786, herein referred to by the ChEMBL identifier, were identified as having a significantly higher neurite counts as compared to vehicle treated cells (Fig 1d,e). Analysis of the same images using an algorithm developed with the Cellomics Neuronal Profiling package measured an increased neurite length over multiple compound concentrations compared to control (Fig 1c). It should be noted that the images analyzed were taken from a well-based preselection of images; therefore the neurites/clusters of images may not necessarily be reflective of the dataset of kinase inhibitors as a whole.

X.4 Discussion

High content analysis is typically undertaken by acquiring images of interest and performing feature-specific (e.g. fluorescent marker intensity) analysis. Although current HCA software algorithms are capable of discerning many patterns with high precision, the algorithms are generally selected and optimized to measure specific image features and phenotypes that are already of interest to the investigator. It is generally difficult to discern a ‘global’ treatment-specific phenotype (e.g. finding any and every different morphological parameter compared to control). Although it is possible to run an image set through every possible feature algorithm to maximize the discovery of cell differences relative to control, this can be labor and time intensive and is not typically done. Additionally, heterogeneous responses from a cell population can make subtle (though possibly important) changes difficult to identify, and is a common phenomenon that has been widely reported in HCA data. (Gough et al. 2014; Burrell et al. 2013; Huang 2009; Altschuler and Wu 2010).

Advanced informatics methods are also being used in conjunction with traditional HCA algorithms to identify novel morphological changes within an image set. Software such as PhenoRipper does not rely on cell segmentation and classification, and instead focuses on clustering similar and dissimilar morphological features within a particular experiment (Rajaram et al., 2012). It is then incumbent on the user to determine whether a discovered feature is of biological interest. We used PhenoRipper as an agnostic way to visualize dissimilar cell morphologies within an experiment, and discovered that a particular small molecule (CHEMBL-1766490) enhanced neurite outgrowth. This finding was then validated with a more “classical” algorithm specifically designed to quantify neurite outgrowth.

Although defining the mechanism of action for this small molecule is outside the scope of this discussion, this serves as a case in point for one of the main challenges of phenotypic screening, which is elucidating the molecular mechanism by which identified molecules exert their effect on the cell. CHEMBL-1766490 is reported to have p38 α MAPK inhibitory activity (Goldstein et al., 2011). It has been reported that p38 α inhibition induces neuronal differentiation, whereas other reports suggest that neurite outgrowth is inhibited following p38 α knockdown in PC12 and P19 committed neuronal cell lines (Morooka and Nishida, 1998; Iwasaki et al., 1999; Aouadi et al., 2006). Although inhibiting p38MAPK to enhance neuritogenesis may be a postnatal photoreceptor-specific phenomenon, it is possible that this is an “off-target” mechanism of the molecule, perhaps due to inhibition of one or more other kinases. Indeed, the kinase selectivity profile of the pyrimidinopyridone family from which this molecule was found demonstrates submicromolar inhibition of 10 out of the 300 kinases

tested (Goldstein et al., 2011). Future studies performed in a photoreceptor-specific systems pharmacology context may uncover currently unknown mechanisms of actions of similar small molecules on cellular phenotypes.

In this study, we started with an image data set that was derived from a screen that was carried-out with an assay that was originally designed to assess GFP reporter expression as a marker for photoreceptor differentiation. Then, using a different, machine learning-based analytical approach, we uncovered a small molecule that increased neurite outgrowth of developing photoreceptors. As many laboratories performing image-based analyses typically have extensive archives of data from previous studies, they could potentially benefit from a similar approach. By integrating machine-learning based approaches with established analysis algorithms, it should be possible to uncover previously unknown phenotypic features, and to identify molecules that modulate and regulate the pathways controlling these morphologies and phenotypes.

Bibliography

- Altschuler SJ, Wu LF. Cellular heterogeneity: Do differences make a difference? *Cell* (United States). 2010; 141:559–563.
- Aouadi M, Bost F, Caron L, et al. p38 mitogen-activated protein kinase activity commits embryonic stem cells to either neurogenesis or cardiomyogenesis. *Stem Cells*. 2006; 24:1399–1406. [PubMed: 16424397]
- Burrell RA, McGranahan N, Bartek J, et al. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* (England). 2013; 501:338–345.
- Carpenter AE, Jones TR, Lamprecht MR, et al. CellProfiler: Image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* (England). 2006; 7:R100.
- Csurka G, Dance C, Fan L, et al. Visual categorization with bags of keypoints. *Proceedings of Workshop on Statistical Learning in Computer Vision*. 2004; 1:22.
- Fuller JA, Shaw GC, Bonnet-Wersinger D, et al. A high content screening approach to identify molecules neuroprotective for photoreceptor cells. *Adv Exp Med Biol* (United States). 2014; 801:773–781.
- Goldstein DM, et al. Discovery of 6-(2,4-difluorophenoxy)-2-[3-hydroxy-1-(2-hydroxyethyl)propylamino]-8-methyl-8H-pyrido[2,3-d]pyrimidin-7-one (pamapimod) and 6-(2,4-difluorophenoxy)-8-methyl-2-(tetrahydro-2H-pyran-4-ylamino)pyrido[2,3-d]pyrimidin-7(8H)-one (R1487) as orally bioavailable and highly selective inhibitors of p38 α mitogen-activated protein kinase. *J Med Chem* (United States). 2011; 54:2255–2265.
- Gough AH, Chen N, Shun TY, et al. Identifying and quantifying heterogeneity in high content analysis: Application of heterogeneity indices to drug discovery. *PLoS One* (United States). 2014; 9:e102678.
- Held M, Schmitz MH, Fischer B, et al. CellCognition: Time-resolved phenotype annotation in high-throughput live cell imaging. *Nat Methods* (United States). 2010; 7:747–754.
- Huang S. Non-genetic heterogeneity of cells in development: More than just noise. *Development* (England). 2009; 136:3853–3862.
- Iwasaki S, Iguchi M, Watanabe K, et al. Specific activation of the p38 mitogen-activated protein kinase signaling pathway and induction of neurite outgrowth in PC12 cells by bone morphogenetic protein-2. *J Biol Chem* (UNITED STATES). 1999; 274:26503–26510.
- Morooka T, Nishida E. Requirement of p38 mitogen-activated protein kinase for neuronal differentiation in PC12 cells. *J Biol Chem* (UNITED STATES). 1998; 273:24285–24288.
- Rajaram S, Pavie B, Wu LF, et al. PhenoRipper: Software for rapidly profiling microscopy images. *Nat Methods* (United States). 2012; 9:635–637.

- Sommer, C., Strahle, C., Kothe, U., et al. Ilastik: Interactive learning and segmentation toolkit. 2011. p. 230-233.
- Wang QL, Chen S, Esumi N, et al. QRX, a novel homeobox gene, modulates photoreceptor gene expression. *Hum Mol Genet (England)*. 2004; 13:1025–1040.
- Welsbie DS, Yang Z, Ge Y, et al. Functional genomic screening identifies dual leucine zipper kinase as a key mediator of retinal ganglion cell death. *Proc Natl Acad Sci U S A (United States)*. 2013; 110:4045–4050.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. XX.1. Enhanced photoreceptor neurites following treatment with CHEMBL-1766490
a: 2D MDS plot of PhenoRipper derived image profiles for QRX-GFP retinal cells cultured 1 week. **b:** Histogram of representative superblocks that best distinguish vehicle control from CHEMBL-1766490 treated wells. Red arrow: superbblock containing neurites. **c:** Quantification of average neurite length using Cellomics Neuronal Profiling following treatment with DMSO vehicle or CHEMBL-1766490. **d, e:** Representative images from control (**d**) and CHEMBL-1766490 (**e**) treated wells.