Using a novel multiplexed algal cytological imaging (MACI) assay and machine learning as a way to characterize complex phenotypes in plant-type organisms

Journal:	*Environmental Science & Technology*
Manuscript ID	es-2023-07733f.R2
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Ostovich, Eric; University of Wisconsin-Milwaukee, School of Freshwater Science Klaper, Rebecca; University of Wisconsin Milwaukee, School of Freshwater Sciences

Using a novel multiplexed algal cytological imaging (MACI) assay and machine learning as a way to characterize complex phenotypes in plant-type organisms

Eric Ostovich[1], Rebecca Klaper[1*]

[1]School of Freshwater Sciences, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin, 53204, USA

Corresponding author email: rklaper@uwm.edu

Keywords: algae, ecotoxicology, phenotypic profiling, deep learning, high-throughput, new approach methodologies

ABSTRAST

High-throughput phenotypic profiling assays, popular for their ability to characterize alternations in single-cell morphological feature data, have been useful in recent years at predicting cellular targets and mechanisms of action (MoAs) for different chemicals and novel drugs. However, this approach has not been extensively used in environmental toxicology due to the lack of studies and established methods for performing this kind of assay in environmentally relevant species. Here, we developed a multiplexed algal cytological imaging (MACI) assay, based on the subcellular structures of the unicellular microalgae, *Raphidocelis subcapitata*, a toxicology and ecological model species. Several different herbicides and antibiotics with unique MoAs were exposed to *R. subcapitata* cells and MACI was used to characterize cellular impacts by measuring subtle changes in their morphological features, including metrics of area, shape, quantity, fluorescence intensity, and granularity of individual subcellular components. This study

demonstrates that MACI offers a quick and effective framework for characterizing complex phenotypic responses to environmental chemicals, that can be used for determining their MoAs and identifying their cellular targets in plant-type organisms.

Synopsis

This work proposes novel high-throughput phenotypic profiling and fluorescence imaging techniques to predict/characterize the mechanisms of action of environmental chemicals.

INTRODUCTION

With increasing quantities and classes of contaminants introduced into commerce and therefore found in the environment, there is a call for more rapid techniques for evaluating their potential hazard in a quick and efficient manner. Therefore, there is a need for more nontargeted (i.e. quantifying hundreds of distinct properties to identify unknown responses), high-throughput profiling assays that can characterize biological activity, identify potency thresholds, and predict mechanisms of action (MoAs),[1] as compared to traditional targeted assays which only quantify singular, or few cellular functions or properties.[2] In recent years, morphological/phenotypic profiling has been shown to provide rich sources of data for interrogating biochemical perturbations as the morphology of a cell is extremely sensitive and strongly influenced by factors such as metabolism, genetic state, and environmental cues.[3] Additionally, it has been shown that specific biological perturbations deliver specific phenotypic profiles, and therefore any subset of morphological features that deviate from that of healthy cells can serve as a fingerprint, or unique identifier, to characterize biological activity.[4] For example Gustafsdottir et al.

2013 demonstrated the ability of morphological profiling to capture a wide range of cellular phenotypes after exposing U2OS cells exposed to 1600 different commercially available compounds with a range of different MoAs. Furthermore, when comparing the fingerprint of cells treated with novel compounds to that of cells treated with compounds with previously established MoAs, the probable MoA of these novel compounds can then be identified.[5]

Common high-throughput phenotypic profiling assays, like the Cell Painting Assay[6], involve the use of multiplexed fluorescence cytochemistry to visualize multiple subcellular structures within a cell and high-content imaging to take hundreds of snapshots of their morphology in an automated and consistent manner. These image data can then be converted into quantitative data by using bioimage analysis to extract hundreds of morphological features at the resolution of a single cell. These morphological features include metrics related to cell size, shape, fluorescence intensity, texture, granularity, and even spatial relationships between organelles which all represent subtle unbiased descriptors of the phenotypic state. Currently, high-throughput phenotypic profiling assays are used most often in the context of drug discovery and disease models. For example, Hughes et al. (2020) used Cell Painting to screen 19,555 compounds and profile the phenotypic response across several esophageal adenocarcinoma cell lines; subsequent bioimage data was analyzed using hierarchical clustering and machine learning methods across 733 individual morphological features per cell, including measurements of size, shape, texture, and intensity. In doing so, this study successfully identified novel drug targets, predicted the MoAs of test compounds through comparison to a library of reference compounds, and discovered pharmacological classes that

targeted that specific type of cancer. However, this kind of assay may also have applications in other fields like eco- and environmental toxicology.

Recently, the Unites States Environmental Protection Agency (USEPA) has begun to use high-throughput phenotypic profiling for the screening and hazard identification of environmental chemicals, however, only human-derived cell models are still largely being used for this purpose.[4,8] While human-derived cell models provide the advantage of proven characterization and predictive power, they may not accurately represent phenotypic responses in environmentally relevant species, like plants and algae, whose cells are biologically distinct from animal cells. For example, DCMU (3-(3,4-dichlorophenyl)-1,1-dimethylurea), or Diuron, has been reported to cause DNA damage in certain types of human cancer cell lines[9] but in plants and algae targets photosystem II (PSII) proteins.[10] Furthermore, human cell lines may not accurately represent environmentally safe exposure levels for certain compounds, like ZnO nanoparticles, that are relatively benign to humans,[11] but acutely toxic to algal species at low concentrations.[12] Thus, cell models which are more environmentally relevant should be considered. In particular, *Raphidocelis subcapitata*, a prevalent type of freshwater green algae and an environmentally relevant organism, is a good candidate model to use for high-throughput phenotypic profiling as it is a USEPA established model for environmental toxicology[13] and an important bioindicator species for assessing/monitoring water quality.[14] Additionally, its strictly unicellular nature makes it beneficial for image-based assays in terms of downstream bioimage segmentation, which may be more difficult to do for other common microalgae like *Chlamydomonas spp*.,

Chlorella spp., and *Scenedesmus spp*. that tend to form colonies or coenobia under stress.[15,16]

Here we describe a multiplexed algal cytological imaging (MACI) assay for the phenotypic profiling of environmental chemicals, based on three subcellular structures that are important for the architecture of *R. subcapitata* cells: the chloroplast, nuclei, and lipid droplets. Each of these subcellular structures represent a different aspect of algal physiology and can be used to characterize complex phenotypes and predict phytotoxic mechanisms of action. For example, the chloroplast is an important subcellular compartment for conducting photosynthesis, and features related to chloroplast fluorescence can be used to describe relative levels of chlorophyll content between treatments, and can even be used to calculate the quantum yield of PSII.[17] Nuclei play a crucial role in regulating gene expression and facilitating cellular division, and features related to the number on nuclei per cell, as well as the relative amount of DNA content per nucleus, can be used to describe instances of cell cycle disruption/arrest.[15,18] And lastly, lipid droplets, which are a collection of neutral lipids, often triacylglycerol (TAG), and serve as an alternative form of energy storage to starch, are often indicators of cell stress when accumulated in large quantities.[19]

As a proof of concept, this study aims to assess the ability of the MACI assay to characterize and differentiate between cells which were exposed to various compounds with unique MoAs. The performance of this assay was evaluated by testing a small set of herbicides and antibiotics with already established MoAs, and preforming a hierarchical clustering analysis of their phenotypic fingerprints. Additionally, a convolutional neural network (CNN) machine learning model was trained off of a small subset of cell image

data in order to predict compound-specific perturbances. We propose that the MACI assay is a quick and effective way to characterize complex phenotypes and predict interactions with environmentally relevant chemicals in plant-type species.

MATERIALS AND METHODS

Algal Cell Culture.

A stock culture of *R. subcapitata*, inoculated at 1×10^5 Cells·mL^{-1}, was grown in a 1 L Erlenmeyer flask and cultured in OECD 201 media [20]. Cells were illuminated continuously with a full spectrum T8 light bulb at a photon flux of 70 μE·m^{-2}·s^{-1}. The stock culture was mixed with an orbital shaker at a speed of 111 rpm.

Exposure Setup.

Eight different environmental chemicals with unique established MoAs (described in **Table 1**), were exposed to cells for 24 hours at either 0 (control), 0.1, 1, or 10 μM.

Table 1: Environmental Chemicals with Known MoAs

Chemical	Mechanism of Action	Abbreviation	References
Aclonifen	Carotenoid Biosynthesis Inhibition	CBI	[21]
Carfentrazone	Membrane Disruption	MD	[22]
DCMU	PSII Photochemistry Inhibition	PPI	[23]
Glufosinate	N_2 Metabolism Inhibition	NMI	[24]
H_2O_2	Oxidative Stress	OS	[25]
Metolachlor	Very-Long-Chain Fatty Acid Synthesis Inhibition	VLCFASI	[26]
MSMA	OP Uncoupler/e- Transport Inhibition	OPU/e-TI	[27]
Zeocin	DNA Damage	DD	[28]

Each chemical was solubilized in either OECD 201 medium or 100% EtOH depending on its solubility, and sonicated for 30 minutes to prepare a primary 1000 µM stock. A secondary 100 µM stock solution was then prepared for each chemical by preforming a serial dilution from their respective primary stock solution into OECD 201 media. While cells were growing exponentially, 900 µL aliquots of algal stock culture (~5×10^5 Cells·mL^{-1}) were seeded into individual 1.5 mL microcentrifuge tubes. For each treatment, done in quintuplicate, the respective secondary stock solution and/or OECD 201 medium was added to each 900 µL cell suspension at a final volume of 1 mL. Resulting EtOH content in final exposure samples (≤ 1%) had a negligible effect on cell morphology (**Supplementary Figure S1**). The samples were then placed under full spectrum illumination, with tube lids open, at a photon flux of 70 $\mu E \cdot m^{-2} \cdot s^{-1}$ for 24 hours. 24 hours was chosen for the exposure duration as this timepoint has been shown to better delineate initial phenotypic impacts,[8] however, longer timepoints can also be chosen depending on the purpose of the exposure.

Multiplexed Algal Cytological Imaging (MACI) Assay.

At the conclusion of the exposure, aliquots from each sample were transferred to sterile 1.5 mL microcentrifuge tubes. Commercially available fluorescent probes and glutaraldehyde were used to stain and fix multiple subcellular compartments of the algal cells, respectively.

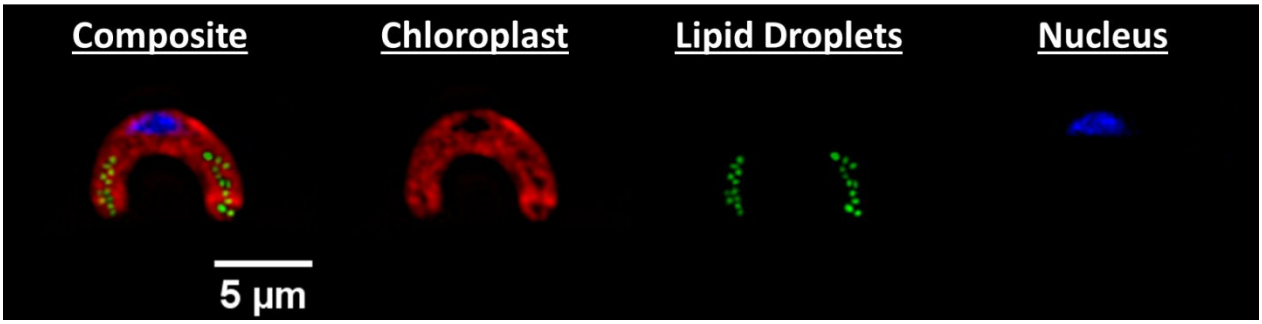


Figure 1: The MACI assay as seen in a *Raphidocelis subcapitata* cell. Representative fluorescence micrograph where each column represents a different fluorescently labeled subcellular compartment visualized by the MACI assay.

The chloroplast is auto fluorescent due to the presence of chlorophyll, and thus did not require a fluorescent probe, but NucBlue (Thermo Fisher, R37605) was used to label nuclei and BODIPY 505/515 (Thermo Fisher, D3921) was used to stain neutral lipid droplets, as seen in **Figure 1**. After the adding reagents to the sample aliquots, all reactions were incubated overnight at 4 °C to minimize enzymatic degradation and maintain the integrity of the subcellular structures. Cells can also be stored at 4 °C for as long as one week when fixed with higher concentrations of glutaraldehyde (~0.25%) for maximum recovery (Shapiro et al., 2001). Alternatively, for live cell imaging, it is recommended that all reagents, excluding glutaraldehyde, are added, and reactions are incubated in the dark at room temperature for 15-30 minutes. After incubating reactions, cells were centrifuged at 4000 x g for 5 min, washed 2x with 1X Phosphate Buffered Saline (PBS), and resuspended in PBS. Cells from each sample were loaded into a well of a glass bottom 384 well plate (Cellvis, P384W-1.5H-N) at a seeding density of ~2 $\times 10^3$ cells$\cdot$mm^{-2} for optimal distribution of cells across the well surface. After loading cells, the well plate was then spun gently at 600RPM for 1 minute to concentrate cells at the bottom of the well. Alternatively, loaded well plates can also be set aside for 30-60 minutes at room temperature to allow cells to settle before imaging. Images were acquired at 9 sites

per well with an ImageXpress Micro XLS High-Content Screening System with a 60X Plan Fluor 0.85 NA air immersion objective (Molecular Devices, 1-6300-0414), using the fluorescent channels described in **Table 2**.

Table 2: MACI Fluorescence Cytochemistry parameters

Organelle	Stain	Channel	Excitation (nm)	Emission (nm)
Chloroplast	Auto Fluorescent	Cy5	628/40	692/40
Lipid Droplets	BODIPY 505/515	GPF	472/30	520/35
Nuclei	NucBlue	DAPI	377/50	447/60

To enhance image contrast and resolution, the digital confocal feature was used during image acquisition. For representative cell images with higher resolution, some images were also acquired with a 100X CFI L PLAN EPI CC 0.85 NA air immersion objective (Molecular Devices, 1-6300-0419).

Bioimage Analysis.

After acquiring images, any image analysis software can be used to extract quantitative data from the images. In this study, CellProfiler,[30] an open-source modular bioimage analysis software, was used for image pre-processing, object segmentation, and morphological feature extraction at the resolution of individual cells. A pipeline for analyzing algal cells can be found in the supplementary information. The pipeline identifies the chloroplast from the Cy5 channel, which spans most of the cell area of *R. subcapitata*, to help aid a segmentation algorithm in identifying individual cells, or regions of interest (ROI). These ROI are then used as a mask to identify which subcellular structures belong to which cell. This pipeline extracts 450 unique morphological features per cell related to area, shape, intensity, and granularity of each subcellular structure,

which is then exported to a local SQLite database file. Data tables were extracted from the SQLite database file using the RSQLite package in R.[31]

NOTE: Image naming rules and module settings in the MACI pipeline may need to be optimized for other microscopes and experiments.

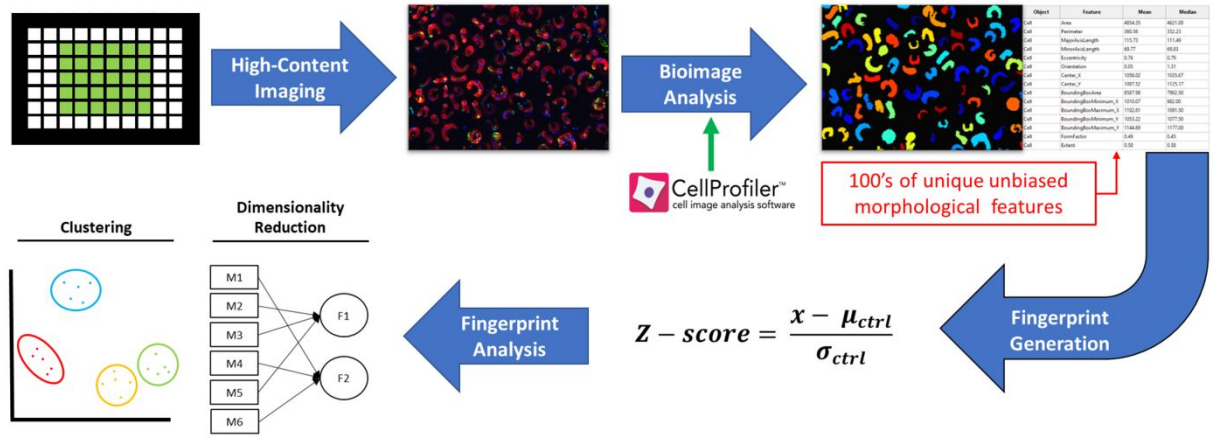


Figure 2: Phenotypic profiling workflow. General overview of the steps taken to conduct phenotypic profiling using image-based data. After perturbing and staining the algae cells, they are seeded into a 384-glass bottom well plate for high-content imaging. A CellProfiler pipeline (or a pipeline from any image analysis software) is used to convert the image data into quantitative data at the resolution of a single cell. Phenotypic fingerprints are then generated by calculating z-scores, and analyzed by reducing data dimensionality and/or preforming a clustering analysis.

Phenotypic Profiling - Fingerprint Analysis.

Phenotypic response data was analyzed using a general phenotypic profiling workflow (**Figure 2**). Data was firstly processed by aggregating single-cell morphological feature measurements to per-image and then per-well values, which was done by taking the cell and image means, respectively. Secondly, well data from each compound and dose were then normalized to the non-treated cell control by computing a Z-score:

$$Z - score = \frac{x - \mu_{ctrl}}{\sigma_{ctrl}} \quad (1)$$

where x is the feature value, μ_{ctrl} is the mean feature value of the control, and σ_{ctrl} is the standard deviation of the feature value of the control. In order to verify whether each compound elicited a change to the entire phenotypic profile of treated cells and to characterize compound-specific phenotypic changes, a partial least squares-discriminant analysis (PLS-DA) was performed in R using the mixOmics package.[32] Before feeding phenotypic response data into the PLS-DA models, an ANOVA was performed across all features for each refence chemical to remove any non-informative features with little variance (p-values > 0.05). Lastly, factor analysis was used to further reduce the dimensionality of phenotypic data vectors, and the fingerprints were subsequently compared to one another using hierarchical clustering based on Pearson correlation in R. Different data-analysis strategies are discussed in the supplementary information.

Phenotypic Profiling - Convolutional Neural Networks.

In addition to fingerprint analysis, a CNN was also trained on a small subset of cells (~10.5%) using the classifier module on CellProfiler Analyst (Ver 3.0).[33] Only a small percentage of the cells were chosen to build the CNN model as to avoid the possibility of overfitting (for example, the model may start to associate well location with the compounds instead of the actual cell features), however, it is worth noting that in the scope of this exposure, ~10.5% of cells is still a sufficiently large number of observations. A separate bin was created for each chemical-treated and the non-treated cell control in the classifier module, where around 1000 randomly fetched cells from each treatment were placed in each respective bin **(Figure 5)**. After training the CNN, it was used to score

the entire experiment by classifying individual cells into predicted phenotypic classes, and computing enrichment scores for each sample as the logit area under the receiver operating characteristic curve. An ANOVA and a Tukey post-hoc test was used to evaluate the significance of predicted phenotypic class enrichments for each treatment.

Statistical Analysis.

All statistical analyses were performed using R Studio [34,35]. A Shapiro-Wilk test was used to verify normal distribution and a One-Way ANOVA was used to compare variance among group means, while a Tukey post-hoc test was used for multiple comparisons. In each analysis, significant differences were determined with a 95% confidence interval.

RESULTS AND DISCUSSION.

Complex changes in phenotypes of cells upon chemical exposure can be defined using MACI.

After perturbing cells with respective chemicals and conducting MACI, a CellProfiler pipeline was used to convert the high-content image data into quantitative data. From these data 450 unique, unbiased, morphological features were extracted at the resolution of a single cell, which were used to generate phenotypic fingerprints of molecular interaction. In order to verify whether each chemical elicited a significant change to cell morphology, we characterized the cellular responses to each chemical, individually, by comparing changes in their complex phenotypic profiles with increasing concentration. This was done with a PLS-DA, which is a supervised machine learning algorithm that projects multidimensional datasets onto two-dimensional planes in order to predict

responses between groups. Based on the PLS-DA response plots (**Figure 3**), each chemical treatment displays a significant collective separation between response groups, thereby indicating that each chemical does elicit a significant, and measurable, change to cell morphology after 24 hours. Variable importance in projection (VIP) scores were also extracted from each PLS-DA response plot (**Supplementary Table S1**). VIP scores indicate the features, or predictors, which are most influential in driving the separation between response groups and can, therefore, help characterize groups of phenotypic markers that are unique to chemicals with specific MoAs. In this case, predictors with VIP scores above 1.0 were considered most important.

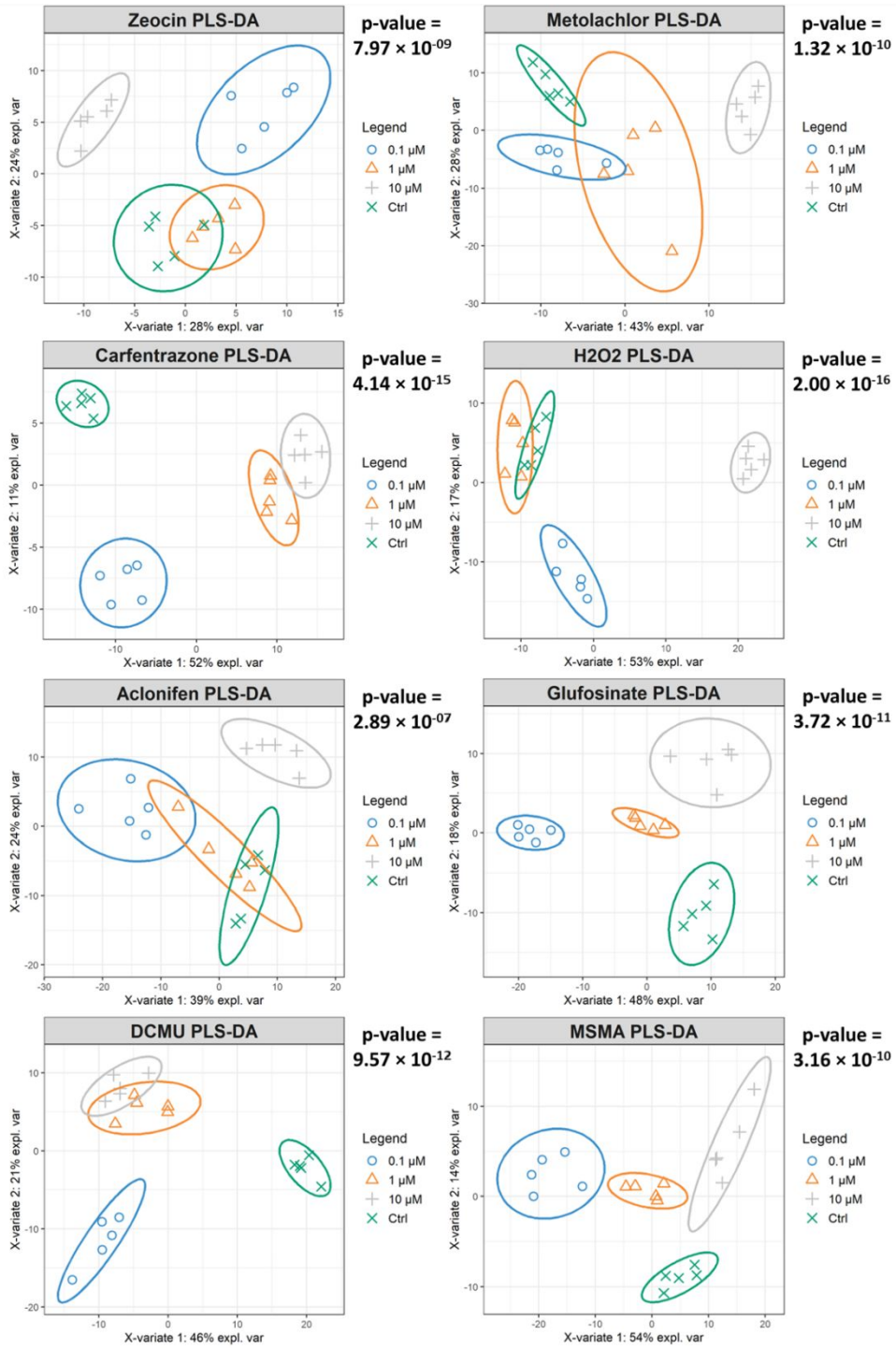


Figure 3: Phenotypic responses to environmental chemicals. A partial least squares-discriminant analysis (PLS-DA) response plot for each chemical graphically describes the change across complex morphological feature data with increasing concentration; ellipses represent 95% confidence intervals and p-values represent ANOVA statistics across the 1st latent variable between response groups.

Based on the top 10 VIP scores for each PLS-DA model, response groups of each chemical were delineated with a distinct combination of phenotypic markers, thus indicating that MACI can be used to characterize compound-specific interactions. For example, exposure to Aclonifen, a carotenoid biosynthesis inhibitor,[21] was most distinguishable by changes in nuclear shape features while exposure to Metolachlor, a very-long-chain fatty acid synthesis inhibitor,[26] was most distinguishable by changes in features related to lipid droplet granularity. There were also some phenotypic markers that overlapped for certain chemicals. For example, cells treated with H_2O_2 and Zeocin, a DNA damaging agent,[28] both garnered the same top three chloroplast normalized moment features. However, the VIP ranking order of these phenotypic markers, as well as the overall combination of markers, were still distinct between chemical profiles. Therefore, when using phenotypic profiling for predicting chemical MoAs, the entire profile, rather than the individual features, should be considered.

Impacts of chemicals can be separated using MACI through hierarchically clustering phenotypic fingerprints.

The ability of MACI to delineate subtle phenotypes of chemical-specific perturbance was evaluated, firstly, by comparing the phenotypic fingerprints of each chemical treatment to one another. For this purpose, the 10 μM data was used as this was the concentration that caused the largest change in morphology, compared to the control, for most of the chemicals after 24 hours. After constructing the phenotypic fingerprints, an ANOVA was used to identify individual features that carry little information, which were removed from the analysis given a p-value > 0.05. Additionally, factor analysis was used to further reduce the dimensionality of the phenotypic data vectors down to 7 eigen features/factors

in order to minimize redundant measurements adding noise while still preserving the variance within the dataset, as suggested by Young et al., 2008. The optimal number of factors was determined with a non-graphical Cattell's scree test.

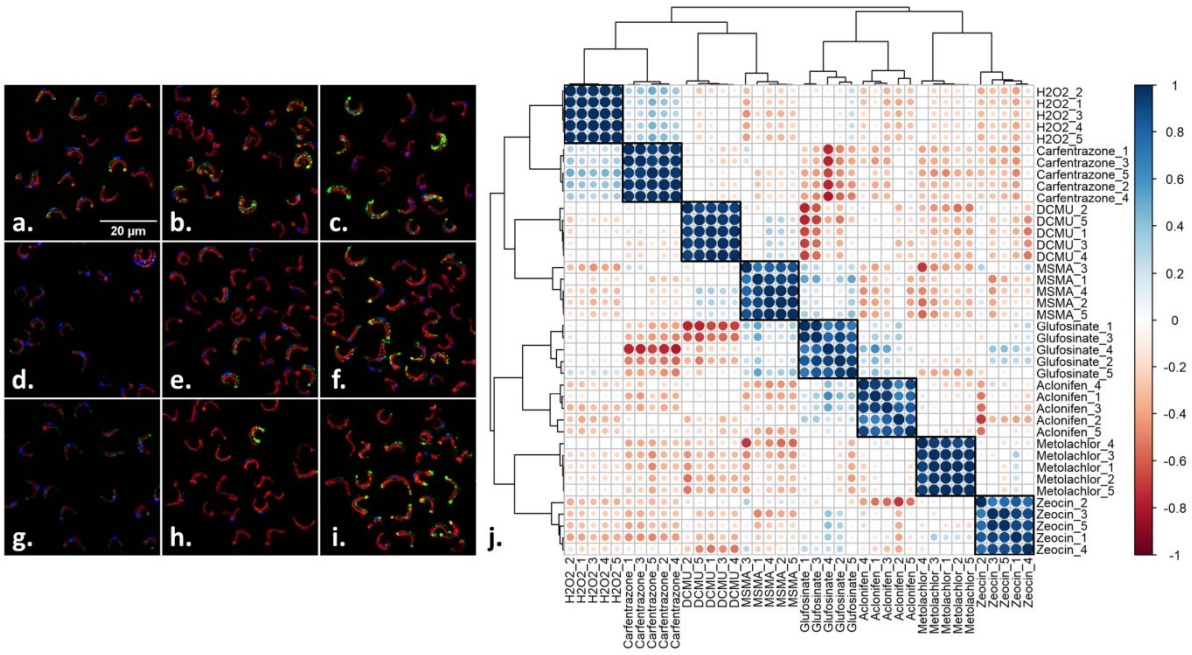


Figure 4: Phenotypic responses of environmental chemicals compared to one another. (a.-i.) MACI labeling patterns in nine different treatment groups; (a.) Healthy Cells, (b.) Zeocin, (c.) Metolachlor, (d.) Carfentrazone, (e.) MSMA, (f.) Glufosinate, (g.) H_2O_2, (h.) DCMU, and (i.) Aclonifen treated cells. (j) Pearson correlation matrix across unique phenotypic responses. Chemical-treated samples are hierarchically clustered based on their Pearson coefficient in relation to the other chemical-treated samples; dendrograms and boxes represent individual clusters.

The phenotypic fingerprints across all replicates for each chemical were hierarchically clustered based on their Pearson correlation coefficient in relation to one another (**Figure 4**). The hierarchical clustering analysis was able to identify 8 separate clusters (**Figure 4j.**). All of the clusters grouped individual replicates of the same chemical treatment together, thus indicating high correlation across replicates and reproducibility in cell-chemical interactions. However, some treatments were slightly less robust than others. For example, samples treated with Glufosinate and Aclonifen had less correlation

between replicates within their respective clusters in comparison to other chemical treatments, however, their overall correlations were still considerably high. Interestingly, some correlations between clusters could also been seen. For example, Carfentrazone and H_2O_2 clusters exhibited fairly high correlation to one another. This is not all that surprising, though, due to the similarity in the way each of these chemicals interact with algal cells and the stark visual similarities between their MACI labeling patterns (**Figure 4d.** and **4g.**, respectively). Additionally, DCMU and MSMA clusters, whose MoAs are both related to electron transport inhibition,[23,27] bore some slight correlation to one another. However, despite all of these intertreatment correlations, MACI was still sensitive to the subtle differences in their phenotypic responses as seen by the clear separation of treatment clusters (**Figure 4j**), thus suggesting that this assay can be used to successfully predict compound-specific perturbations and discriminate between chemicals with unique MoAs.

Chemical MoAs can be identified based on phenotypic response using convolutional neural networks.

In addition to hierarchical clustering analysis, we also took a deep learning approach to analyze complex phenotypes and delineate chemicals by their MoA, using convolutional neural networks. CNNs are a type of artificial neural network, which are most notable for the way they process image data similarly to the visual cortex of the human brain.[37]

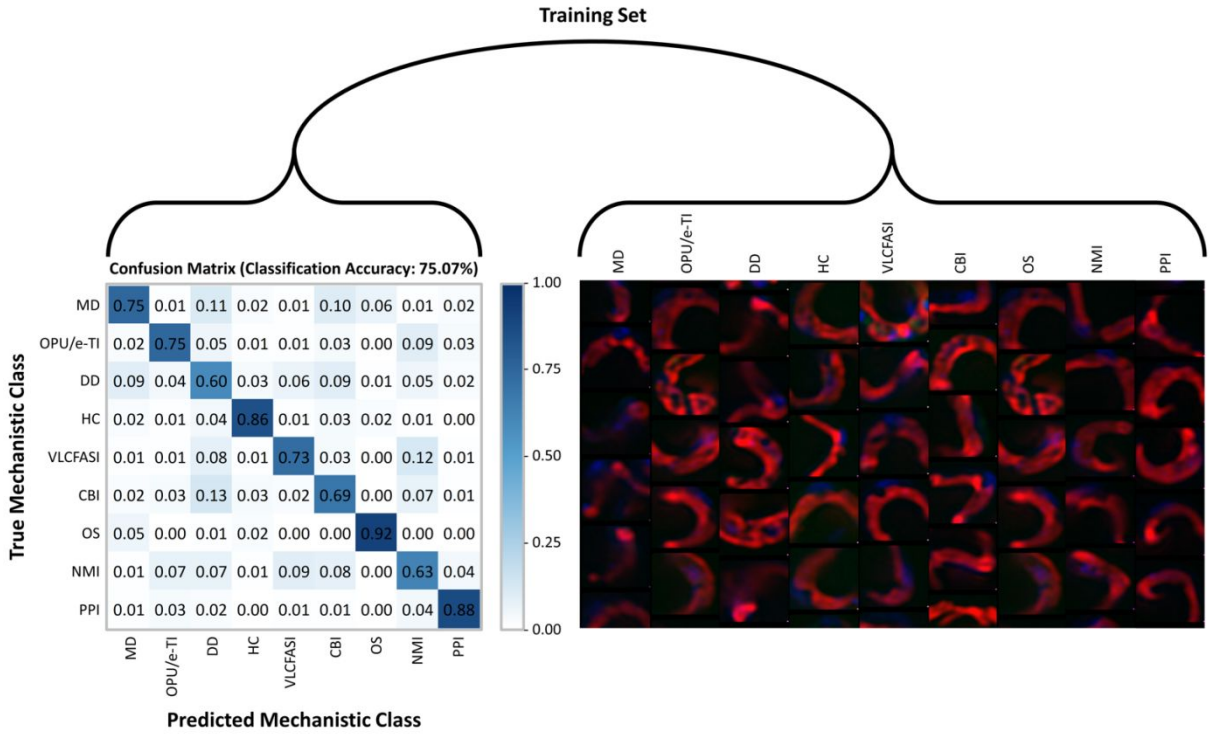


Figure 5: Convolutional neural network construction. A small subset of randomly fetched cells in each treatment are fed into a convolutional neural network model. Based on the training data, this CNN model yields a classification accuracy of 75.07% at correctly classifying cells by their true mechanistic class.

In CellProfiler Analyst, a CNN was trained on a small subset of randomly fetched cells from each treatment (~10.5% of cells from the entire experiment) using 50x50 neurons per layer. Based on the confusion matrix (**Figure 5**), the CNN model was less robust at distinguishing certain mechanistic classes from one another, such as cells with membrane disruption vs DNA damage or cells with inhibited N_2 metabolism vs inhibited very-long-chain fatty acid synthesis. However, the CNN model was still able to predict the correct mechanistic class across training cells with moderately good classification at an accuracy of 75.07% (**Figure 5**). Once trained, the CNN model was used to score each cell in the experiment, based on its individual phenotype, with a predicted mechanistic class, and then calculate enrichment scores for each sample.

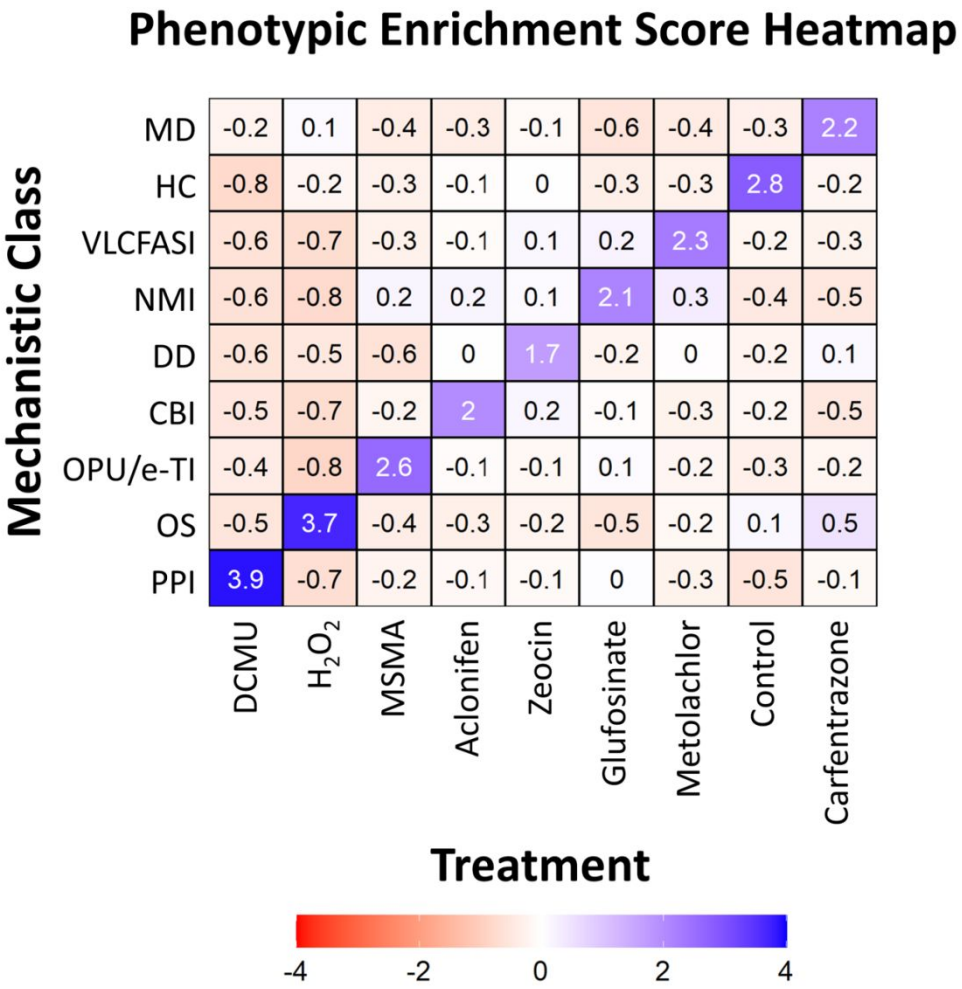


Figure 6: Phenotypic enrichment score heatmap. The CNN model classifies each cell across the entire experiment with a predicted mechanistic class based on its phenotype. Enrichment scores for each mechanistic class are calculated in each sample. Heatmap values represent average treatment enrichment scores; white colored numbers represent significantly enriched mechanistic classes for respective treatments.

Despite some confusion in the discrimination of mechanistic classes, the CNN model was still able to classify each treatment with the correct MoA. This is visualized in the heatmap of enrichment scores (**Figure 6**) where each treatment was significantly enriched in the appropriate mechanistic class. Based on these results, the deep learning approach reinforced the ability of MACI to separate chemicals by MoA. However, both deep learning and hierarchical clustering analyses proved to be robust and sensitive to subtle changes in complex phenotypes.

Complex phenotypic profiles are more efficient at predicting mechanisms of action rather than single interpretable features.

A majority of the morphological features used for phenotypic profiling are not interpretable on their own. Zernike moments, for example, measure specific aspects of an object's radial distribution,[38] and when multiple Zernike moments across multiple orders are combined together, they become powerful mathematical descriptors of that object's shape. Although they can be useful for reconstructing patterns and for detecting subtle changes in cell shape,[39] individual Zernike moments, by themselves, do not hold much intrinsic nor biological meaning. However, there are a select few of morphological features that do hold some biological relevance, such as those related to the intensity and quantity of fluorescence signals, which we can use to elucidate interesting biological phenomena. For example, measurements of integrated intensity, which is the sum of pixel intensity values over a ROI,[40] directly correlate to the number fluorophores in that ROI, and thus directly or indirectly measure relative levels of target biomolecular content. This kind of measurement has been used for analyzing endpoints related to changes in protein content[41] and for determining cell cycle stages based on the relative abundance of DNA content.[42,43] Another useful metric is quantifying the number of fluorescent objects within a single cell. For example, measuring the number of intracellular vesicles has been used to study endpoints related to the cellular uptake of micro/nano plastic particles[44] and for analyzing the intracellular trafficking of certain proteins.[45] In regard to *R. subcapitata*, three features that have relevance to algal physiology are the number of nuclei/cell, and the chloroplast and lipid droplet integrated intensities, which are related to chlorophyll and TAG content, respectively (**Supplementary Table S2**).

Often the phenotypic measurements that scientists are most interested in analyzing, such as the above-mentioned features, are not always the best predictors for characterizing the MoA of different compounds due to their lack of specificity. For example, 4 out of 8 treatments significantly increased chlorophyll content, while 4 out of 8 treatments also significantly increased the average number of nuclei/cell, in some cases to similar magnitudes while following similar trends to one another. TAG content was only significantly increased in Metolachlor treated cells, but most treatments did not elicit a significant change in TAG content. Furthermore, when using these three features in a hierarchical clustering analysis, based on Pearson correlation, the analysis was not sensitive enough to discriminate between chemical-specific perturbance (Supplementary **Figure S2**), as compared to the previous hierarchal clustering analysis using the entire profile (**Figure 4**). Ultimately, when conducting high-content phenotypic profiling, it is advised to evaluate changes in the entire profile, rather than changes in individual or select morphological features alone, as they do not hold enough information that can directly be linked to a specific MoA[5].

Applications of MACI, its advantages and limitations.

As the use of high-throughput phenotypic profiling assays become more popular, the MACI assay, in particular, could have several applications in environmental science. MACI could be used to screen the thousands of chemicals in the marketplace currently being reevaluated through efforts such as ToxCast for potential environmental toxicity and mechanism of action. Furthermore, it could be used to study the potential impacts of those emerging contaminants and environmental chemicals by characterizing their cellular targets and identifying their phytotoxic MoAs. This could be especially useful if

used alongside the Cell Painting assay with other environmentally relevant models like drosophila and rainbow trout gill cell lines to make cross-species toxicological comparisons that span multiple trophic levels. Additionally, in the context of environmental risk assessment, MACI could also be used to help prioritize which emerging contaminants and environmental chemicals require further evaluation. Once patterns are developed for a broad array of mechanisms of action it could also be used as a novel technology for the rapid detection and monitoring of chemicals and specifically emerging contaminants in the environment. We also hope to develop this technology further as a means to study more complex environmental samples with mixtures of chemicals that have different MoAs, which may be made possible with continuing advancements in machine learning techniques.

MACI could also be used as an *in vitro* model to drive developments in herbicide and agrochemical discovery. Green microalgae cells, like *R. subcapitata*, bear several similarities to the mesophyll cells of higher order terrestrial plants in terms of the cellular components, and their constituents, they contain, the environmental processes they carry out, and the evolutionarily conserved pathways and molecules they utilize.[46,47] Furthermore, since many herbicides and agrochemicals are delivered to plants via foliar application, MACI could be used for identifying cellular targets and determining MoAs of novel herbicides and agrochemicals. However, this technique may not be as useful for soil-based herbicides and agrochemicals.

Currently several high-throughput screening assays exist in addition to high-throughput phenotypic profiling, such as high-throughput transcriptomics and high-throughput proteomics. While both of these assays provide a rich molecular level understanding of

chemical interactions, they can be extremely time-consuming, costly, and computationally expensive. In comparison, MACI, as an image-based profiling assay, provides the advantage of low cost and high speed, while still retaining a capacity for in-depth characterization and classification.[48,49] Additionally, assays which measure the differential expression of transcripts and proteins rely heavily on well-established annotations for those transcripts and proteins. This is an issue for most environmentally relevant organisms, like *R. subcapitata*, which have not been annotated to the extent with which human disease models have.[50] However, another advantage of MACI, and other image-based profiling assays, is that they only require comparisons to a library of reference chemicals with established MoAs in order to derive meaning from the phenotypic response of novel or unstudied chemicals[5]. Lastly, MACI provides the advantage of greater experimental precision as each individual cell, of which there can be up to 100,000's-1,000,000's in any given experiment, serves as an independent, technical measurement, thereby, also limiting the impact of measurement error.[51] Some limitations of this approach also exist. As with all image-based high-throughput phenotypic profiling assays, this approach requires the use of a high-content automated fluorescence imaging system, which can be a significant initial investment. While, images taken with standard fluorescence microscopes can be analyzed in a similar manner to that of images taken on a high-content imaging system, the lack of automation can lead to an abundance of human error during image acquisition, thus decreasing the effectiveness of the assay. Secondly, small differences in image acquisition parameters, like Z-offsets, laser power, and acquisition times, could potentially greatly impact the downstream feature extraction and data analysis. As a result, it has not yet been proven that image data can be directly

compared between imaging platforms.[5] Since having a library reference set of image data is an important aspect of this kind of assay, this means that a separate library would have to be established for each individual imaging platform, rather than being able to share libraries across imaging platforms. However, with the development of machine learning and new statistical practices for phenotypic profiling, it may become possible to do so in the future.

Overall, based on the work described in this study, MACI provides a potentially quick and effective framework for characterizing complex phenotypes and compound-specific interactions which is suitable for predicting chemical MoAs in plant-type organisms. This work demonstrates the power and benefit of image-based phenotypic profiling in general, which is a technique that may continue to drive many advancements in the field of environmental science and technology.

Associated Content

Supporting Information: Description of different data-analysis strategies for image-based phenotypic profiling. Tables reporting the molecular weight & logKow (Table S1), and VIP scores (Table S2) for reference compounds. Table and description for interpretable features of biological relevance (Table S3). Figures displaying phenotypic responses to ethanol (Figure S1) and Pearson correlation matrix based on minimal feature data (Figure S2) (DOC).

Acknowledgements

This work was supported by the National Science Foundation under the Grant No. CHE-2001611, the NSF Center for Sustainable Nanotechnology. The CSN is part of the Centers for Chemical Innovation Program.

REFERENCES

1. Thomas, R. S., Bahadori, T., Buckley, T. J., Cowden, J., Deisenroth, C., Dionisio, K. L., Frithsen, J. B., Grulke, C. M., Gwinn, M. R., Harrill, J. A., Higuchi, M., Houck, K. A., Hughes, M. F., Hunter, E. S., Isaacs, K. K., Judson, R. S., Knudsen, T. B., Lambert, J. C., Linnenbrink, M., Martin, T. M., Newton, S. R., Padilla, S., Patlewicz, G., Paul-Friedman, K., Phillips, K. A., Richard, A. M., Sams, R., Shafer, T. J., Woodrow Setzer, R., Shah, I., Simmons, J. E., Simmons, S. O., Singh, A., Sobus, J. R., Strynar, M., Swank, A., Tornero-Valez, R., Ulrich, Elin M., Villeneuve, D. L., Wambaugh, J. F., Wetmore, B. A., & Williams, A. J. The next generation blueprint of computational toxicology at the U.S. Environmental protection agency. *Toxicol. Sci.* **169**, 317–332 (2019).

2. Caicedo, J. C., Singh, S. & Carpenter, A. E. Applications in image-based profiling of perturbations. *Curr. Opin. Biotechnol.* **39**, 134–142 (2016).

3. Gustafsdottir, S. M., Ljosa, V., Sokolnicki, K. L., Wilson, J. A., Walpita, D., Kemp, M. M., Seiler, K. P., Carrel, H. A., Golu, T. R., Schreiber, S. L., Clemons, P. A., Carpenter, A. E., & Shamji, A. F. Multiplex cytological profiling assay to measure diverse. *PLoS One* **8**, 1–7 (2013).

4. Willis, C., Nyffeler, J. & Harrill, J. Phenotypic Profiling of Reference Chemicals across Biologically Diverse Cell Types Using the Cell Painting Assay. *SLAS Discov.* **25**, 755–769 (2020).

5. Svenningsen, E. B. & Poulsen, T. B. Bioorganic & Medicinal Chemistry Establishing cell painting in a smaller chemical biology lab – A report from the frontier. *Bioorg. Med. Chem.* **27**, 2609–2615 (2019).

6. Bray, M., Singh, S., Han, H., Davis, C. T., Borgeson, B., Hartland, C., Kost-alimova, M., Gustafsdottir, S. M., Gibson, C. C., Carpenter, A. E. Cell Painting , a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. **11**, 1757–1774 (2016).

7. Hughes, R. E., Elliott, R. J.R., Munro, A. F., Makda, A., O'Neill, J. R., Hupp, T., & Carragher, Neil O. High-Content Phenotypic Profiling in Esophageal Adenocarcinoma Identifies Selectively Active Pharmacological Classes of Drugs for Repurposing and Chemical Starting Points for Novel Drug Discovery. *SLAS Discov.* **25**, 770–782 (2020).

8. Nyffeler, J. Willis, C. Lougee, R., Richard, A., Paul-friedman, K., & Harrill, J. A.Bioactivity screening of environmental chemicals using imaging-based high- throughput phenotypic profiling. *Toxicol. Appl. Pharmacol.* **389**, 114876 (2020).

9. Huovinen, M., Loikkanen, J., Naarala, J. & Vähäkangas, K. Toxicity of diuron in human cancer cells.

Toxicol. Vitr. **29**, 1577–1586 (2015).

10. Magnusson, M., Heimann, K. & Negri, A. P. Comparative effects of herbicides on photosynthesis and growth of tropical estuarine microalgae. *Mar. Pollut. Bull.* **56**, 1545–1552 (2008).

11. Nohynek, G. J., Dufour, E. K. & Roberts, M. S. Nanotechnology, cosmetics and the skin: Is there a health risk? *Skin Pharmacol. Physiol.* **21**, 136–149 (2008).

12. Aruoja, V., Dubourguier, H. C., Kasemets, K. & Kahru, A. Toxicity of nanoparticles of CuO, ZnO and TiO2 to microalgae Pseudokirchneriella subcapitata. *Sci. Total Environ.* **407**, 1461–1468 (2009).

13. U.S. Environmental Protection Agency. Ecological Effects Test Guidelines OCSPP 850.4500: Algal Toxicity. Office of Chemical Safety and Pollution Prevention; Washington, D.C. (EPA-712C-006. *United States Environ. Prot. Agency* 26 (2012).

14. Yamagishi, T., Yamaguchi, H., Suzuki, S., Horie, Y. & Tatarazako, N. Cell reproductive patterns in the green alga Pseudokirchneriella subcapitata (=Selenastrum capricornutum) and their variations under exposure to the typical toxicants potassium dichromate and 3,5-DCP. (2017). doi:10.1371/journal.pone.0171259

15. Zachleder, V. & Vítová, M. *The Physiology of Microalgae*. *The Physiology of Microalgae* (Springer Cham, 2016). doi:10.1007/978-3-319-24945-2

16. Roccuzzo, S., Couto, N., Karunakaran, E., Kapoore, R. V., Butler, Thomas O. Mukherjee, J. Hansson, E. M., Beckerman, A. P., & Pandhal, J. Metabolic Insights Into Infochemicals Induced Colony Formation and Flocculation in Scenedesmus subspicatus Unraveled by Quantitative Proteomics. *Front. Microbiol.* **11**, 1–17 (2020).

17. Jakob, T., Schreiber, U., Kirchesch, V., Langner, U. & Wilhelm, C. Estimation of chlorophyll content and daily primary production of the major algal groups by means of multiwavelength-excitation PAM chlorophyll fluorometry: Performance and methodological limits. *Photosynth. Res.* **83**, 343–361 (2005).

18. Hlavová, M., Vítová, M., Bišová, K. & Zachleder, V. M. DNA Damage during G2 Phase Does Not Affect Cell Cycle Progression of the Green Alga Scenedesmus quadricauda. (2011). doi:10.1371/journal.pone.0019626

19. Ischebeck, T., Krawczyk, H. E., Mullen, R. T., Dyer, J. M. & Chapman, K. D. Lipid droplets in plants and algae: Distribution, formation, turnover and function. *Semin. Cell Dev. Biol.* **108**, 82–93 (2020).

20. OECD. Test No. 201: Freshwater Alga and Cyanobacteria, Growth Inhibition Test. Organization for Economic Cooperation and Development. OECD Guidelines for Testing of Chemicals, Section 2. OECD Publishing Service, Paris, France. (2011). doi:https://doi. org/10.1787/9789264069923

21. Almeida, A. C., Gomes, T., Langford, K., Thomas, K. V. & Tollefsen, K. E. Oxidative stress in the algae Chlamydomonas reinhardtii exposed to biocides. *Aquat. Toxicol.* **189**, 50–59 (2017).

22. Li, X., Volrath, S. L., Nicholl, D. B. C., Chilcott, C. E., Johnson, M. A., Ward, E. R., & Law, M. D. Development of Protoporphyrinogen Oxidase as an Efficient Selection Marker for Agrobacterium tumefaciens-Mediated Transformation of Maize. *Plant Physiol.* **133**, 736–747 (2003).

23. Glauch, L. & Escher, B. I. The Combined Algae Test for the Evaluation of Mixture Toxicity in

Environmental Samples. **39**, 2496–2508 (2020).

24. Nagai, T. Sensitivity differences among seven algal species to 12 herbicides with various modes of action. *J. Pestic. Sci.* **44**, 225–232 (2019).

25. Geer, T. D., Kinley, C. M., Iwinski, K. J., Calomeni, A. J. & Rodgers, J. H. Comparative toxicity of sodium carbonate peroxyhydrate to freshwater organisms. *Ecotoxicol. Environ. Saf.* **132**, 202–211 (2016).

26. Machado, M. D. & Soares, E. V. Reproductive cycle progression arrest and modification of cell morphology (shape and biovolume) in the alga Pseudokirchneriella subcapitata exposed to metolachlor. *Aquat. Toxicol.* **222**, 105449 (2020).

27. Dayan, F. E. & Zaccaro, M. L. de M. Chlorophyll fluorescence as a marker for herbicide mechanisms of action. *Pestic. Biochem. Physiol.* **102**, 189–197 (2012).

28. Čížková, M., Slavková, M., Vítová, M., Zachleder, V. & Bišová, K. Response of the green alga Chlamydomonas reinhardtii to the DNA damaging agent zeocin. *Cells* **8**, 1–15 (2019).

29. Shapiro, H. M., Perlmutter, N. G. & Shapiro, H. M. A simple and highly efficient fixation method for Chrysochromulina polylepis (Prymnesiophytes) for analytical flow cytometry. *Cytometry* **44**, 126–132 (2001).

30. Stirling, D. R., Swain-Bowden, M. J., Lucas, A M., Carpenter, A. E., Cimini, B. A., & Goodman, A. CellProfiler 4: improvements in speed, utility and usability. *BMC Bioinformatics* **22**, 1–11 (2021).

31. Müller, K., Wickham, H., James, D. A. & Falcon, S. RSQLite: SQLite Interface for R. (2023). doi:https://rsqlite.r-dbi.org, https://github.com/r-dbi/RSQLite.

32. Rohart, F., Gautier, B., Singh, A. & Lê Cao, K.-A. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* **13**, 1–14 (2017).

33. Stirling, D. R., Carpenter, A. E. & Cimini, B. A. CellProfiler Analyst 3.0: accessible data exploration and machine learning for image analysis. *Bioinformatics* **37**, 3992–3994 (2021).

34. R Core Team. R: A Language and Environment for Statistical Computing. (2019).

35. RStudio Team. RStudio: Integrated Development Environment for R. (2020).

36. Young, D. W., Bender, A., Hoyt, J., McWhinnie, E., Chirn, G. W., Tao, Charles Y., Tallarico, J. A., Labow, M., Jenkins, J. L., Mitchison, T. J., & Feng, Y. Integrating high-content screening and ligand-target prediction to identify mechanism of action. *Nat. Chem. Biol.* **4**, 59–68 (2008).

37. Dürr, O. & Sick, B. Single-cell phenotype classification using deep convolutional neural networks. *J. Biomol. Screen.* **21**, 998–1003 (2016).

38. Allen, P., Calcagni, A., Robson, A. G. & Claridge, E. Investigating the potential of Zernike polynomials to characterise spatial distribution of macular pigment. *PLoS One* **14**, 1–19 (2019).

39. Boland, M. V., Markey, M. K. & Murphy, R. F. Automated recognition of patterns characteristic of subcellular structures in fluorescence microscopy images. *Cytometry* **33**, 366–375 (1998).

40. Subramanian, G. & Vijaya, A. Iterative Intensity Integration Technique (IIIT) for contouring reflective surfaces. *Opt. Lasers Eng.* **93**, 92–99 (2017).

41. Farid, K. M. N. & Derouiche, A. *Quantifying Compartment-Specific Protein Translocation in Astrocytes by Object-Oriented Image Analysis: Mitochondrial Translocation of PKCδ. In: Di Benedetto, B. (eds) Astrocytes. Methods in Molecular Biology* **1938**, (Humana Press; New York; NY., 2019).

42. Roukos, V., Pegoraro, G., Voss, T. C. & Misteli, T. Cell cycle staging of individual cells by fluorescence microscopy. *Nat. Protoc.* **10**, 334–348 (2015).

43. Gomes, C. J., Harman, M. W., Centuori, S. M., Wolgemuth, C. W. & Martinez, J. D. Measuring DNA content in live cells by fluorescence microscopy. *Cell Div.* **13**, 1–10 (2018).

44. Schmidt, A., Mühl, M., Brito, W. A. da S., Singer, D. & Bekeschus, S. Antioxidant Defense in Primary Murine Lung Cells following Short- and Long-Term Exposure to Plastic Particles. *Antioxidants* **12**, 1–24 (2023).

45. Cejas, R. B., Tamaño-Blanco, M. & Blanco, J. G. Analysis of the intracellular traffic of IgG in the context of Down syndrome (trisomy 21). *Sci. Rep.* **11**, 1–12 (2021).

46. Lu, Y. & Xu, J. Phytohormones in microalgae: A new opportunity for microalgal biotechnology? *Trends Plant Sci.* **20**, 273–282 (2015).

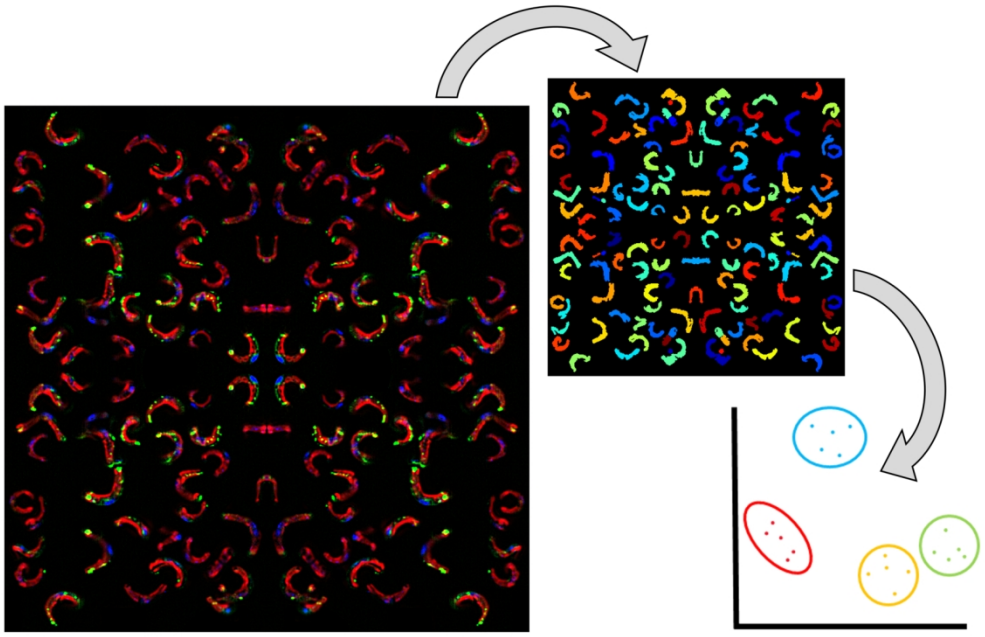
47. Riaz, A., Deng, F., Chen, G., Jiang, W., Zheng, Q., Riaz, B., Mak, M., Zeng, F., & Chen, Z. H. Molecular Regulation and Evolution of Redox Homeostasis in Photosynthetic Machinery. *Antioxidants* **11**, 1–23 (2022).

48. Ljosa, V., Caie, P. D., Horst, Rob., Sokolnicki, K. L., Jenkins, E. L., Daya, S., Roberts, M. E., Jones, T. R., Singh, S., Genovesio, A., Clemons, P. A., Carragher, N. O., & Carpenter, A. E., Comparison of Methods for Image- Based Profiling of Cellular Morphological Responses to Small-Molecule Treatment. *SLAS-DISCOVERY* **18**, 1321–1329 (2013).

49. Feng, Y., Mitchison, T. J., Bender, A., Young, D. W. & Tallarico, J. A. Multi-parameter phenotypic profiling: Using cellular effects to characterize small-molecule compounds. *Nat. Rev. Drug Discov.* **8**, 567–578 (2009).

50. Suzuki, S., Yamaguchi, H., Nakajima, N. & Kawachi, M. Raphidocelis subcapitata (=Pseudokirchneriella subcapitata) provides an insight into genome evolution and environmental adaptations in the Sphaeropleales. *Sci. Rep.* **8**, 1–13 (2018).

51. Blainey, P., Krzywinski, M. & Altman, N. Points of significance: Replication. *Nat. Methods* **11**, 879–880 (2014).



This work uses novel high-throughput phenotypic profiling and fluorescence imaging techniques to predict/characterize the mechanisms of action of environmental chemicals.

74x47mm (600 x 600 DPI)