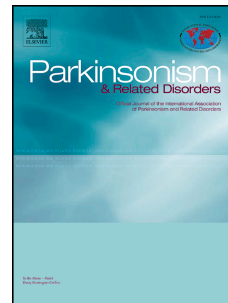


Journal Pre-proof



Machine learning-assisted neurotoxicity prediction in human midbrain organoids

Anna S. Monzel, Kathrin Hemmer, Tony Kaoma, Lisa M. Smits, Silvia Bolognin, Philippe Lucarelli, Isabel Rosety, Alise Zagare, Paul Antony, Sarah L. Nickels, Rejko Krueger, Francisco Azuaje, Jens C. Schwamborn

PII: S1353-8020(20)30120-6

DOI: <https://doi.org/10.1016/j.parkreldis.2020.05.011>

Reference: PRD 4339

To appear in: *Parkinsonism and Related Disorders*

Received Date: 3 February 2020

Revised Date: 7 May 2020

Accepted Date: 7 May 2020

Please cite this article as: Monzel AS, Hemmer K, Kaoma T, Smits LM, Bolognin S, Lucarelli P, Rosety I, Zagare A, Antony P, Nickels SL, Krueger R, Azuaje F, Schwamborn JC, Machine learning-assisted neurotoxicity prediction in human midbrain organoids, *Parkinsonism and Related Disorders* (2020), doi: <https://doi.org/10.1016/j.parkreldis.2020.05.011>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Ltd.

Machine learning-assisted neurotoxicity prediction in human midbrain organoids

Anna S. Monzel^a, Kathrin Hemmer^a, Tony Kaoma^b, Lisa M. Smits^a, Silvia Bolognin^a, Philippe Lucarelli^a, Isabel Rosety^a, Alise Zagare^a, Paul Antony^c, Sarah L. Nickels^a, Rejko Krueger^{c, d, e}, Francisco Azuaje^{b, f}, Jens C. Schwamborn^{a*}

^aDevelopmental and Cellular Biology, Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Esch-sur-Alzette, Luxembourg

^bComputational Biomedicine Research Group, Quantitative Biology Unit, Luxembourg Institute of Health (LIH), Luxembourg, Luxembourg

^cClinical and Experimental Neurobiology, Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Esch-sur-Alzette, Luxembourg

^dParkinson's Research Clinic, Centre Hospitalier de Luxembourg (CHL), Luxembourg, Luxembourg

^eLuxembourg Transversal Translational Medicine, Luxembourg Institute of Health (LIH), Luxembourg

^fData and Translational Sciences, UCB Celltech, Slough, United Kingdom.

*Corresponding author jens.schwamborn@uni.lu +3524666445536; 6, avenue du Swing, L-4367 Belvaux

Keywords: Midbrain organoids | Parkinson's disease | Neurotoxicity | Machine learning

Abstract

Introduction

Brain organoids are highly complex multi-cellular tissue proxies, which have recently risen as novel tools to study neurodegenerative diseases such as Parkinson's disease (PD). However, with increasing complexity of the system, usage of quantitative tools becomes challenging.

Objectives

The primary objective of this study was to develop a neurotoxin-induced PD organoid model and to assess the neurotoxic effect on dopaminergic neurons using microscopy-based phenotyping in a high-content fashion.

Methods

We describe a pipeline for a machine learning-based analytical method, allowing for detailed image-based cell profiling and toxicity prediction in brain organoids treated with the neurotoxic compound 6-hydroxydopamine (6-OHDA).

Results

We quantified features such as dopaminergic neuron count and neuronal complexity and built a machine learning classifier with the data to optimize data processing strategies and to discriminate between different treatment conditions. We validated the approach with high content imaging data from PD patient derived midbrain organoids.

Conclusions

The here described model is a valuable tool for advanced *in vitro* PD modeling and to test putative neurotoxic compounds.

Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized by the loss of midbrain dopaminergic neurons (DANs). The etiology of PD is multifactorial, with endogenous (genetic) and exogenous (environmental) contributors. Accumulating evidence suggests that in the majority of the cases, the combination and interaction of genetic risk variants, ageing, and environment leads to the development of PD [1,2]. This highlights the necessity to expand research to identify potential neurotoxic compounds and their harmful effects on the human brain. One of the major challenges is the development of disease models that can capture the complexity of the human brain. In recent years, stem cell-derived brain organoids have risen as promising disease models. These complex *in vitro* systems mimic the organ architecture and function, and have been shown to model neurological disorders (reviewed in [3]). With increasing complexity of the system, the availability of experimental tools is limited. So far, high-throughput techniques have limited relevance in the highly laborious organoid system. Due to the complex organization, comprehensive image analysis in organoids is challenging. Building on this, we developed methods to automatically acquire and process high-content imaging (HCI) data in organoids, which has been successfully demonstrated in brain organoids [4] and 3D microfluidic cultures [5]. In this study, we further refined this pipeline with optimized HCI data analysis in a neurotoxin-induced PD organoid model. Human midbrain organoids (hMOs) were exposed to 6-OHDA to specifically damage the dopaminergic system [6,7]. We used machine learning (ML) tools to analyze this *in vitro* toxicity assay. Using random forest (RF) classification, we assessed the neurotoxic effect of 6-OHDA on hMOs based on cellular features. This pipeline, from treatment to prediction, is valuable for the exploration of potential neurotoxic compounds in complex human brain organoids.

Methods

Organoids were generated as described in [8] from hiPSCs (Table S1). In order to identify, which 6-OHDA concentration leads to a significant reduction in the amount of dopaminergic neurons, organoids were treated for 48h with 50 μ M, 100 μ M, 175 μ M, 250 μ M and 500 μ M 6OHDA (Sigma) after five weeks of organoid culture, followed by six days of normal culture conditions (Fig S1). Organoids were fixed for immunofluorescence staining, snapfrozen for protein extraction, or dissociated into single cells for flow cytometry. Fixed organoids were sectioned and stained for neuronal markers. Images were acquired using an automated confocal microscope and further processed and cellular features were analyzed in MATLAB. In the R software environment (R version 3.5.1 -- "Feather Spray") we used

principal variance component analysis [9] to explore the contribution of experimental factors to the variance in the dataset and subsequently based data processing approaches on the output. Next, we built a ML model using random forest classification [10,11] to predict the treatment condition of the high content image analysis dataset. The dataset was split into five random subsets; four subsets (training datasets) were used to build the model and the remaining subset (test dataset) served to validate the model performance. This process was iterated ten times (i.e. 10x 5-fold cross validation) and the average classification accuracy was calculated. A detailed description of the experiments can be found in the supplementary information.

Ethical approval

Informed consent was obtained from all individuals donating samples to this study prior to the donation using a written form and protocol. The described work with human induced pluripotent stem cells has been approved by the Ethics Review Panel (ERP) of the University of Luxembourg and the national Luxembourgish research ethics committee (CNER, Comité National d'Ethique de Recherche). *In vitro* experiments were carried out with existing cell lines obtained from previous studies.

Data Availability

The data is openly available at <https://webdav-r3lab.uni.lu/public/data/machine-learning-assisted-neurotoxicity-prediction-in-human-midbrain-organoid/>.

Code Availability

The Matlab and R scripts for image and data analysis, as well as for RF classification are available on GitHub at https://github.com/LCSB-DVB/ML_Tox.

Results

6-OHDA induces concentration-dependent cell death in hMOs

We treated organoids derived from three independent human iPSC lines (Supplementary Table 1) with 6-OHDA concentrations ranging from 50 μ M to 500 μ M (Fig. S1). Cell quantification by flow cytometry revealed that exposure to 6-OHDA caused a concentration-dependent reduction in the amount of living DANs, identified by the rate-limiting enzyme of the dopamine synthesis, Tyrosine hydroxylase (TH) (Fig. S2). We fitted a non-linear regression curve for each line and determined a mean LD₅₀ at 147 μ M 6-OHDA (Fig. 1a). In further experiments, we used a concentration of 175 μ M, which led to a significant reduction of DANs in all three lines (Fig. 1b). Consistent with this, we observed a concentration-dependent reduction of TH in Western Blots, resulting in an average 2.3-fold decrease of the protein after 175 μ M 6-OHDA treatment (Fig. 1c, d). Furthermore, immunofluorescence staining reveals that 6-OHDA treatment leads to a decrease in the amount of TH⁺ cells (Fig. 1e) and to neurite fragmentation (Fig. 1f).

DANs within hMOs show typical signs of degeneration

We examined the effect of 6-OHDA on the neuronal network within hMOs using image-based cell profiling. Organoid sections were stained for neuron-specific-Class-III- β -tubulin (TUJ1), microtubule-associated-protein-2 (MAP2) and TH. We subdivided sections into center (5-6 80 μ m sections of the organoid core) and border sections (Fig. S3a, Fig. S4a) to correct for spatial asymmetry in hMOs. Images were acquired using an automated confocal microscope, processed in MATLAB and the amount of DANs was quantified. Due to their significant differences we analyzed border and center sections separately or corrected for the variation of the section by normalization (Fig. S4b-c). Upon 6-OHDA treatment, the amount of TUJ1⁺ neurons remained unaltered (Fig. 1g), while the amount of TH⁺ DANs decreased significantly (Fig. 1h). We computed a 3D mask for TH⁺ cells and generated a 3D skeleton of the DAN network to extract features such as nodes (dendritic and axonal points of branching) and links (total number of branches), as well as neurite fragmentation (Fig. S3c-d, Table S2). 6-OHDA treatment led to a decrease in the complexity of DANs and increase of fragmented neurites (Fig. 1i-j, Fig. S5-7).

Random Forest prediction of neurotoxicity

We used ML to build a classifier able to discriminate between CTRL and 6-OHDA-treated organoids; and consequently identify the measurements that describe the largest difference. We trained a RF algorithm with 10x 5-fold cross-validation in order to ensure an unbiased estimation of the model performance. We applied our strategy to the raw/unprocessed data. The generated model achieved a classification accuracy

of 75%. The prediction was influenced by dopaminergic features (Table S3, Figure S8a_i,-b_i). Since the prediction power of ML models highly depends on data quality, we attempted to remove highly variable and for the biological effect irrelevant experimental factors. We first assessed the contribution of those factors to the variability observed in the data. We used principal variance component analysis (PVCA) [9] and observed significant contribution of experimental factors (Fig. 2a_i). Building on this, we investigated whether we could improve classification accuracy by normalizing the data. We performed a z-score transformation across the entire dataset for each combination of experiment (four independent batches), cell line (hMO1, hMO2, hMO3) and section (border, center). Normalization strongly improved the classification accuracy of the RF model to 86%, while lowering the variance described by experimental conditions (Fig. 2a_{ii}, Table S3, Fig. S8a_{ii}-b_{ii}). Consistent with this, a clear separation between the treatments using hierarchical clustering (Fig. 2b), as well as principal component analysis (Fig. 2c) was achieved, suggesting that by ML-assisted optimization of data processing strategies, we can predict neurotoxicity using HCI data from human brain organoids.

Random Forest prediction of disease state in PD patient-derived midbrain organoids

To evaluate if RF classification of high content imaging data can be used to identify cellular manifestations in PD patient specific organoids we used a dataset based on a previous publication [4]. In this study, midbrain organoids, from PD patients with a LRRK2-G2019S mutation and controls were stained for the midbrain dopaminergic neuron marker TH/FOXA2. The amount of TH+, FOXA2+, TH+/FOXA2+ and TH-/FOXA2+ cells, as well as dopaminergic neuron complexity, were quantified at three different time points (10, 35, 70 days) using a similar custom image-analysis algorithm. For each time point, we trained a machine learning classifier model to predict the disease state and to identify which features are most important in determining the prediction. This model achieved a classification accuracy of 89% (day10), 92% (day 35), and 93% (day 70). Notably, in line with the previous results [4], each time point showed a different set of top features that determined the prediction outcome, with FOXA2+ cells as top feature at early time points, and amount/complexity of dopaminergic neurons at later time points (Table S4). This demonstrates the applicability of this workflow to monitor phenotypic traits in PD organoids.

Clustering the data obtained from PD patient specific organoids together with the data from the here described toxin model in a single computational model was not leading to meaningful results. This is the case because the two utilized midbrain organoid protocols differ in some important details (e.g. usage of matrix and identity of the stem cells used as starting population) and different markers were analyzed

(FOXA2 was an important marker in the patient specific organoids but was not analyzed in the toxin model) [4, 8].

Discussion

In this study, we have used the neurotoxin 6-OHDA to target specifically the dopaminergic system in organoids. To assess neuronal damage in the hMO system, among other methods we used microscopy-based phenotyping. However, organoids exhibit an architecturally complex heterotypic organization. Typically, multiple cell types including stem cells, glia cells, and neurons are arranged in close proximity in the 3D space, the latter one expanding long neurites in the surrogate matrix. This complexity makes it utterly difficult to measure and quantify neuronal features. The use of image processing algorithms is fundamental to extract features on the single cell level. Technological advances of high imaging throughput, precise analytical frameworks with high-performance computation opens new avenues for phenotypic profiling in brain organoids. In combination with ML approaches, we predicted neurotoxin-induced perturbations in hMOs. RF is by design a useful technique for reducing predictive variability, preventing overfitting and achieving high classification accuracy [12]. Importantly, RF gives estimates of which variables are most important in the classification [10]. Using PVCA, we identified the contribution of experimental factors to the total variance and designed optimized data normalization approaches to improve predictability. We further validated this approach and monitored phenotypic traits by a new analysis of previously published data in PD patient specific organoids with the LRRK2-G2019S mutation. This supports the concept of using image-based profiling studies in organoids to identify environmental or genetic factors that modulate phenotypes. However, the fact that we were not able to compare the data from the toxin model with the data from the LRRK2-G2019S model within the same computational model, highlights that for future meta-analysis of data coming from different studies, it is extremely important to use the same cell culture protocols and analysis strategies. Importantly, this RF model approach is highly sensitive to unbalanced data (control/treated ratio), which can be monitored by calculating sensitivity and specificity of the model (Table S3, S4). Hence, large and class-balanced sample sizes are beneficial. There might be situations where this is easier and less expensive to achieve with other cellular models like iPSC-derived dopaminergic neurons or LUHMES cells cultured in classical 2D conditions. Our model achieved an overall accuracy of 85%, which is acceptable. However, in order to further increase accuracy, it would be necessary to integrate more features beyond cell death and complexity/amount of DANs, which we could demonstrate with the validation dataset (Table S4).

Yet, despite the limitations, we suggest that organoids have the potential to be used as a platform from target identification to toxicity prediction using ML-assisted HCI-based cell profiling.

Acknowledgements

We thank Prof. Dr. Hans R. Schöler of the Max Planck Institute, Prof. Dr. Thomas Gasser of the Hertie Institute in Tübingen and the Coriell Institute for providing cell lines. Microscopy was supported by the LCSB bio-imaging platform. We thank the Disease Modeling Screening Platform from LCSB and LIH for their help with performing automated and high-throughput procedures. This work was supported by a Proof-of-Concept grant from the Fonds National de la Recherche (FNR) Luxembourg (FNR/PoC16/11559169). Further, this is an EU Joint Program - Neurodegenerative Disease Research (JPND) project (INTER/JPND/14/02; INTER/JPND/15/11092422) receiving funding from the FNR. A.S.M., L.M.S., and I.R. were supported by FNR Aides à la Formation-Recherche (AFR). A.S.M. received support from the Lush prize 2017. We also thank the private donors who support our work at the Luxembourg Centre for Systems Biomedicine.

Author Contribution

A.S.M. designed experiments, prepared the figures, performed image and data analysis and wrote the original draft. K.H. and I.R. designed experiments, and prepared the figures. T.K.M. performed RF classification, prepared figures and wrote the original draft. L.M.S., P.L., S.L.N., and A.Z. designed experiments and edited the manuscript. P.A. and S.B. designed image analysis, and edited the manuscript. R.K., F.A., J.C.S. conceived and supervised the project, designed the experiments, and edited the manuscript.

Financial disclosure/Conflict of interest:

JCS is co-founder and CEO of the biotech company OrganoTherapeutics which uses the here described model for drug development. RK is member of the advisory board of the same company. Apart from that the authors declare no conflict of interest.

Funding sources:

This work was supported by a Proof-of-Concept grant from the Fonds National de la Recherche (FNR) Luxembourg (FNR/PoC16/11559169). Further, this is an EU Joint Program - Neurodegenerative Disease Research (JPND) project (INTER/JPND/14/02; INTER/JPND/15/11092422) receiving funding from the FNR. A.S.M., L.M.S., and I.R. were supported by FNR Aides à la Formation-Recherche (AFR). A.S.M. received support from the Lush prize 2017.

Appendix A. Supplementary data

References

- [1] V. Bellou, L. Belbasis, I. Tzoulaki, E. Evangelou, J.P.A. Ioannidis, Environmental risk factors and Parkinson's disease: An umbrella review of meta-analyses, *Parkinsonism Relat. Disord.* 23 (2016) 1–9. <https://doi.org/10.1016/j.parkreldis.2015.12.008>.
- [2] J.C. Schwamborn, Is Parkinson's Disease a Neurodevelopmental Disorder and Will Brain Organoids Help Us to Understand It?, *Stem Cells Dev.* 27 (2018) 968–975. <https://doi.org/10.1089/scd.2017.0289>.
- [3] H. Wang, Modeling Neurological Diseases With Human Brain Organoids., *Front. Synaptic Neurosci.* 10 (2018) 15. <https://doi.org/10.3389/fnsyn.2018.00015>.
- [4] L.M. Smits, L. Reinhardt, P. Reinhardt, M. Glatza, A.S. Monzel, N. Stanslowsky, M.D. Rosato-Siri, A. Zanon, P.M. Antony, J. Bellmann, S.M. Nicklas, K. Hemmer, X. Qing, E. Berger, N. Kalmbach, M. Ehrlich, S. Bolognin, A.A. Hicks, F. Wegner, J.L. Sternecker, J.C. Schwamborn, Modeling Parkinson's disease in midbrain-like organoids, *Npj Park. Dis.* 5 (2019) 5. <https://doi.org/10.1038/s41531-019-0078-4>.
- [5] S. Bolognin, M. Fossépré, X. Qing, J. Jarazo, J. Ščančar, E.L. Moreno, S.L. Nickels, K. Wasner, N. Ouzren, J. Walter, A. Grünewald, E. Glaab, L. Salamanca, R.M.T. Fleming, P.M.A. Antony, J.C. Schwamborn, 3D Cultures of Parkinson's Disease-Specific Dopaminergic Neurons for High Content Phenotyping and Drug Testing, *Adv. Sci.* 6 (2019) 1800927. <https://doi.org/10.1002/advs.201800927>.
- [6] G. Jonsson, C. Sachs, Actions of 6-hydroxydopamine quinones on catecholamine neurons, *J Neurochem.* 25 (1975) 509–516. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1471-4159.1975.tb04357.x>.
- [7] U. Ungerstedt, 6-Hydroxy-dopamine induced degeneration of central monoamine neurons, *Eur J Pharmacol.* 5 (1968) 107–110.
- [8] A.S. Monzel, L.M. Smits, K. Hemmer, S. Hachi, E.L. Moreno, T. van Wuelen, J. Jarazo, J. Walter, I. Brüggemann, I. Boussaad, E. Berger, R.M.T.T.T. Fleming, S. Bolognin, J.C. Schwamborn, Derivation of Human Midbrain-Specific Organoids from Neuroepithelial Stem Cells., *Stem Cell Reports.* 8 (2017) 1144–1154. <https://doi.org/10.1016/j.stemcr.2017.03.010>.

- [9] P. Bushel, Principal Variance Component Analysis (PVCA), (2013) Version 1.22.0.
- [10] L. Breiman, Random Forests, *Mach. Learn.* 45 (2001) 5–32. <https://doi.org/10.1023/A:1010933404324>.
- [11] A. Liaw, M. Wiener, Classification and Regression by random Forest, *R News*. (2002) Vol. 2/3, pp 18–22.
- [12] A. Parmar, R. Katariya, V. Patel, A Review on Random Forest: An Ensemble Classifier, in: Springer, Cham, 2019: pp. 758–763. https://doi.org/10.1007/978-3-030-03146-6_86.

Figure Captions

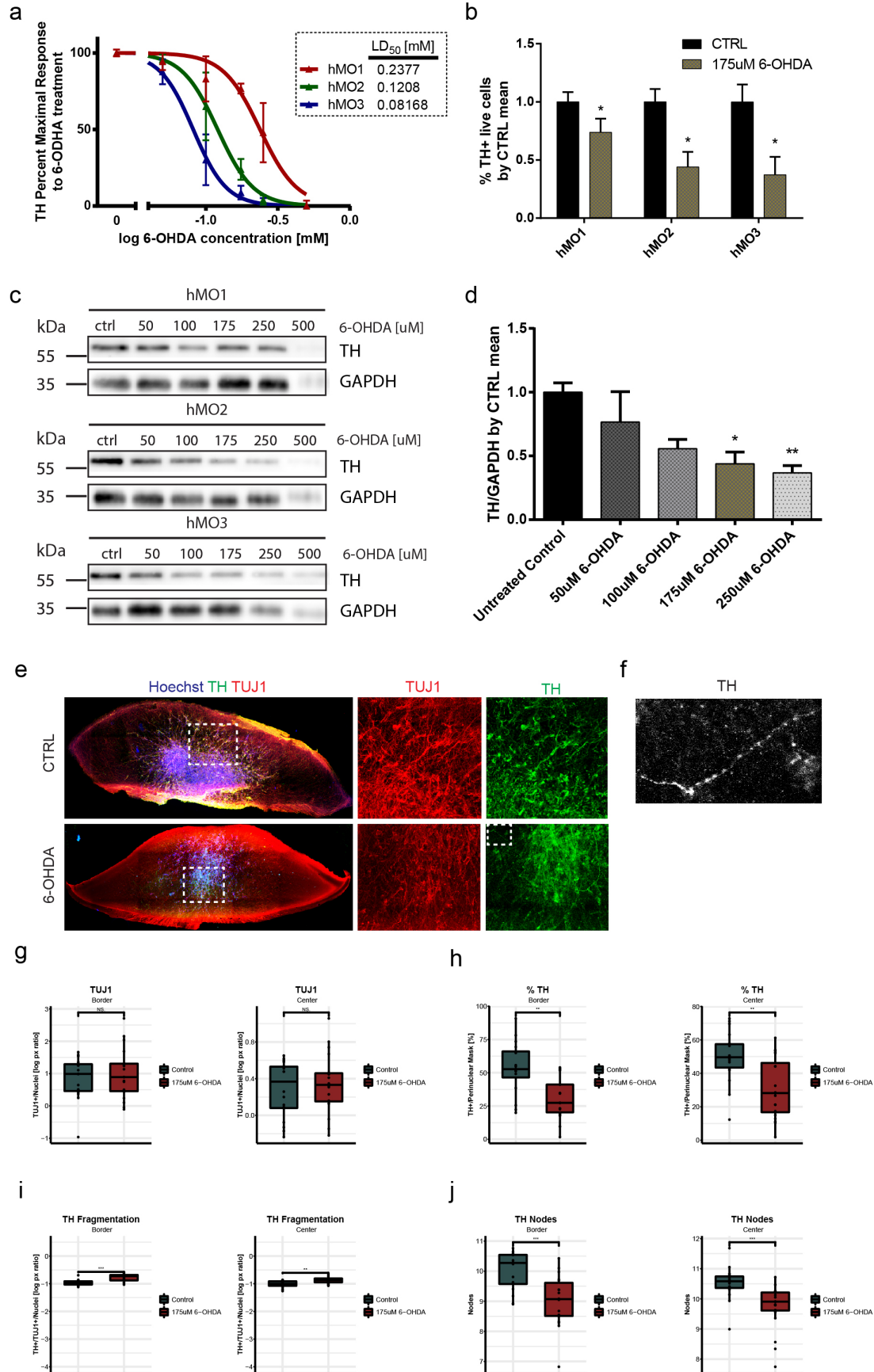
Fig. 1: 6-OHDA-induced concentration-dependent degeneration of DANs.

- a) 6-OHDA dose-response curves fitted to the flow cytometry data showing a 6-OHDA-concentration-dependent decrease in the amount of TH⁺ live cells. Data obtained for each cell line and condition from 6 pooled organoids of two independent organoid batches and 6-OHDA treatments.
- b) Barplot showing a robust decrease in the amount of TH⁺ live cells at a 6-OHDA concentration of 175 μ M. Data obtained for each cell line and condition from six pooled organoids of seven independent organoid batches and 6-OHDA treatments. Error bars represent mean + SEM. * $p < 0.05$
- c) Representative western blot revealing a concentration-dependent decrease of TH protein upon 6-OHDA treatment.
- d) Quantification of c, normalized to the mean of the untreated controls of nine organoids derived from three independent lines. Error bars represent mean + SEM. * $p < 0.05$, ** $p < 0.005$
- e) Immunofluorescence staining for dopaminergic neuronal marker TH and TUJ1 in untreated and 6-OHDA treated organoids reveals dopaminergic neurodegeneration upon treatment with 175 μ M 6-OHDA.
- f) Example of a fragmented TH⁺ neurite after 6-OHDA treatment.
- g) HCI analysis reveals that the overall amount of TUJ1⁺ neurons is unaffected from the treatment. Left: border sections, right: center sections
- h) HCI analysis of TH⁺ cells shows that DANs degenerate after 6-OHDA treatment. %TH: Total count of TH⁺ cells. Left: border sections, right: center sections
- i-j) 6-OHDA treatment leads to impaired neuronal complexity as indicated by increased neurite fragmentation and decreased number of nodes (branch origin and end-point). Left: border sections, right:

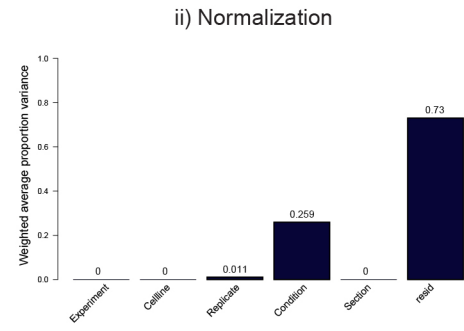
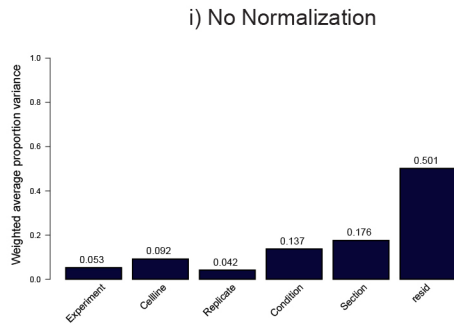
center sections. Data of g-k obtained from four independent organoid batches and 6-OHDA treatments from three cell lines. Wilcoxon rank sum test, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Fig. 2: ML-assisted optimization of HCI data analysis

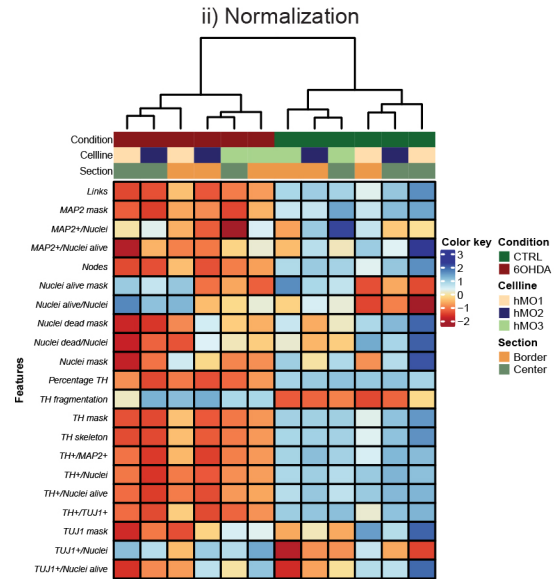
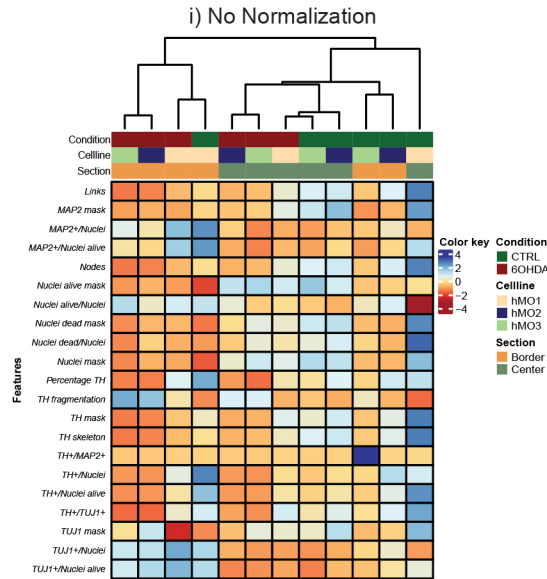
- a) Principal variance component analysis showing the relative contribution of each experimental factor to the total variance observed in the data i) before, and ii) after normalization. Resid. = residual weighted average proportion variance. Contribution of undefined residual effects to variation in the dataset.
- b) Hierarchical clustering using Ward's minimum variance method without (i) and with data normalization (ii).
- c) Principal component analysis of the scaled data before (i) and after data normalization (ii).



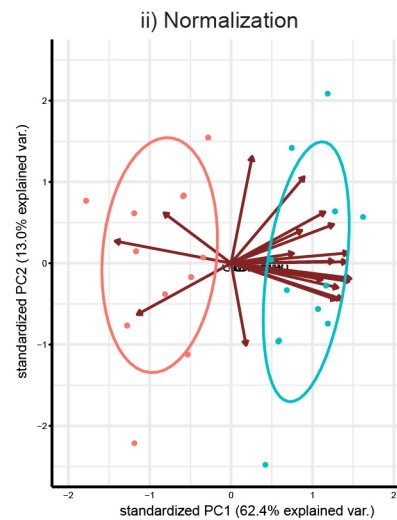
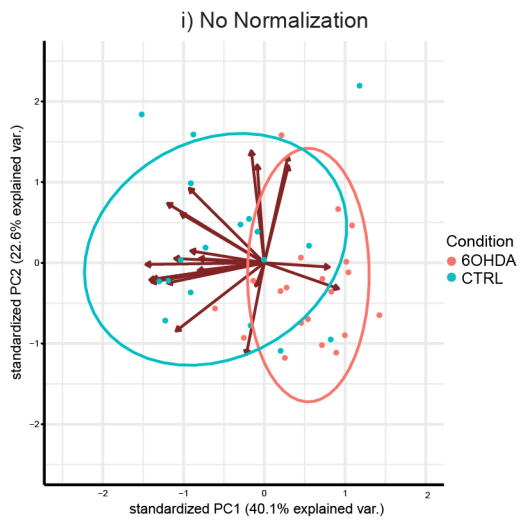
a



b



c



Highlights

- We describe an interdisciplinary approach to assess the effect of 6-OHDA on dopaminergic neurons in midbrain organoids
- 3D high-content screening of organoids was used to extract morphometric features of dopaminergic neurons in 6-OHDA-treated and untreated organoids
- A machine learning classifier was applied on the high-content data to predict neurotoxin-induced perturbations and to optimize data processing strategies