Ben-Gurion University of the Negev
The Faculty of Engineering Sciences
The Department of Software and Information Systems Engineering

Anomaly detection for sensitive hits identification in high-content image-based phenotypic screens

Thesis submitted in partial fulfillment of the requirements
for the Master of Sciences degree

Naor Kolet

Under the supervision of **Dr. Assaf Zaritsky**

February 2023

Ben-Gurion University of the Negev
The Faculty of Engineering Sciences
The Department of Software and Information Systems Engineering

Anomaly detection for sensitive hits identification in high-content image-based phenotypic screens

Thesis submitted in partial fulfillment of the requirements
for the Master of Sciences degree

Naor Kolet

Under the supervision of **Dr. Assaf Zaritsky**

Author: _____ Date: _____

Supervisor: _____ Date: _____

Chairman of Graduate Studies Committee: _____ Date: _____

February 2023

Abstract

High-content image-based screening is a powerful tool for identifying phenotypic differences in cell populations. In my dissertation, I am proposing anomaly detection as a new measurement for the sensitive detection of phenotypic alterations in high-content image-based phenotypic screens. Anomaly detection, aimed at detecting abnormal observations that deviate from a baseline distribution, is especially suited for screening because many replicates of control experiments define the baseline distribution, and few replicates of many perturbations are potential anomalies from this baseline distribution. Specifically, I trained autoencoders to reconstruct the control phenotypic profiles and used the reconstruction error as a measurement for anomaly detection. I implemented this idea in three modes: feature-based organelle anomaly, feature-based inter-organelle anomaly, and image-based inter-organelle anomaly. My results suggest feature-based modes are more sensitive and complementary, but slightly less reproducible, compared to the common statistical approach. My preliminary results with image-based anomaly indicate the potential to uncover spatial dependencies between the organelles. Further investigations are needed to validate these results.

Acknowledgments

First and foremost, I would like to express my gratitude to my supervisor Dr. Assaf Zaritsky for his constructive criticism and patience throughout my dissertation. He filled me with inspiration even when difficulties arose during the research. I would also like to thank Dr. Shahar Golan, from the Jerusalem College of Technology, who accompanied me during my research from the very beginning and provided me with practical knowledge and tools to handle my first research. I would also like to thank Alon Shpigler, a PhD student and a fellow lab colleague, who joined my journey, helped, and added a new point of view to this research. Together, we pushed this research to a higher level. Lastly, I would like to thank my family and friends, especially Amit Shakarchy, a fellow lab colleague, who stood by my side, encouraged me throughout this journey, and believed in me even when I didn't.

Contents

List of Figures

Introduction

Cells are the fundamental unit of structure and function of all organisms. Organelle and subcellular organization form the underlying basis for cellular complexity. Visual cell phenotypes, characterized by the morphological features of a cell such as a cell shape and molecular composition, can serve as powerful readouts for cell state [1-5]. Alterations in visual cell phenotype, such as changes in cell shape and organelle organization, can provide insight into the cell's physiological state, as well as assist in the diagnosis and treatment of diseases [5-8].

High-content image-based screening combines automated microscopy and automated image analysis to identify phenotypic alterations in cell morphology, dynamics, and/or molecular composition [9-16]. To identify hits in a screen, the deviation of the cells' phenotype from each treatment is measured and compared to the phenotype of untreated/control cells. For example, one can define phenotypic variation as the fraction of features that dramatically deviated from the control in response to treatment [16]. Current measurements of magnitude, for which a treatment deviates from the control, hold the implicit assumption of independence between features. However, features are interdependent, even after feature selection, because an alteration in a feature can be derived from non-linear convoluted dependencies between other features [17]. For example, the variation and/or deviation of cells' intracellular organization may be largely explained by the cells' shape [18]. Explicitly measuring these complex dependencies for all features is not feasible due to the curse of dimensionality [17]. Thus, the biological function investigated may be incorrectly interpreted due to a simplified representation of the underlying data complexity.

I propose to employ anomaly detection methods to identify hits on screens. Anomaly detection aims at detecting abnormal observations that deviate from a predefined baseline pattern [19] and has vast applications in bioinformatics [20-21], healthcare [22], and cyber-security [23]. Anomaly detection relies on statistically characterizing the data distribution and defining observations that do not conform to this distribution as anomalous. Different approaches, such as neighbor-based [24-25] and isolation-based [26] were applied for anomaly detection, with deep learning emerging as especially powerful in recent years [27]. The advantage of deep neural networks stems from their ability to integrate massive amounts of complex data into a generalized model that captures the inherent representation of the data distribution [28-29]. High-content image-based screens obtain many replicates of control experiments and many perturbations, with a few replicates for each perturbation. These properties naturally align with the formulation of an anomaly detection problem: the many control experiments are used to statistically define reliable baseline patterns, and hits are defined and ranked according to their deviation from the baseline.

The Cell Painting assay is widely used for image-based high-content phenotypic screening. Cell Painting enables simultaneous analysis of multiple organelles, by staining the cells with six fluorescent dyes, imaged in five different channels (i.e., modalities) [30] The ability to capture the different modalities simultaneously makes it perfect to elucidate a detailed phenotypic profile.

There are several public datasets of high-content phenotypic screening using cell painting such as [10, 16, 31]. In my work, I used a public Cell Painting screening dataset with 30,000 small-molecule perturbations [10].

In my research, I explored three modes of hits identification using anomaly detection in Cell Painting high-content image-based screens. Two of the three modes are based on features derived from (1) organelles, (2) inter-organelle dependencies, and the third is image-based to harness spatial information. Common to all modes is the method used for anomaly detection. First, training deep neural network-based autoencoders to compress feature- or image-based representations of single cells in control experiments. Second, using the distribution of the control cells' reconstruction error as the baseline pattern. Finally, defining hits as treatments that have reconstruction error, which deviates from the baseline by some magnitude. The mode relying on organelle-derived features is based on autoencoders that map each modality (organelle) to itself, while the mode relying on inter-organelle dependencies maps several modalities to another modality. Applying these approaches, I demonstrate that anomaly detection can lead to more sensitive hits identification compared to direct assessment of feature deviation from the control. While inter-organelle anomaly detection proved to be more sensitive and complementary to the common statistical approach, it failed to outperform organelle anomaly detection. Further investigation of the organelle anomaly detection showed that it produces a slightly reduced level of reproducibility suggesting that the current implementation is not sufficiently powerful. Further investigations, especially related to the removal of batch effects, could harness the full potential of anomaly detection as an effective measure to capture complex feature dependencies. Promising preliminary results using image-based anomaly detection, that harness the spatial information between the organelles, highlight its potential for uncovering inter-organelle spatial dependencies.

Related Work

My thesis lies in between the computational and experimental disciplines, specifically covering anomaly detection, deep neural networks, high-content image-based phenotypic screening, and intra-cellular organization.

Anomaly Detection

Anomaly detection is the task of identifying observations that do not conform to the expected patterns in a dataset. Anomalies may indicate rare or unusual phenomena that warrant further investigation or may be indicative of problems that need to be addressed. Anomaly detection has a wide range of applications, including fraud detection, equipment failure prediction, and cybersecurity [19]. Among the computational approaches applied to detect anomalies are statistical methods, rule-based systems, and machine learning algorithms. Statistical methods involve analyzing the distribution of the data and identifying observations that fall outside of a specified range defined by the baseline distribution [32-35]. Rule-based systems rely on predefined rules or thresholds to identify anomalous data points [36-37]. Machine learning approaches involve training models on baseline data and using the trained models to identify anomalies that deviate from the distribution of the baseline data [38-39].

Traditional machine learning techniques for identifying outlier points typically rely on the examination of the local neighborhood of a candidate point, utilizing methods such as k-NN [40] or LOF [24]. These techniques assign a score to each data point based on its local neighborhood and use this score to determine which points are considered outliers. A significant limitation of these traditional techniques is the high computational complexity and resources required to process large datasets, making it infeasible to apply them in practice. To tackle this limitation, several techniques focus on learning the baseline distribution by subsampling the dataset. For example, decision trees partition the data into sections by the features' values. Each internal node represents a condition on one randomly chosen feature which separates the data points into two subgroups until every data point is isolated. These techniques assume that anomalous data points are more likely to be isolated from the baseline data, i.e., to be in an isolated partition [26, 41-43]. The limitation of such techniques is that such simple models can’t handle complex data with non-linear interdependencies between the features.

Deep learning, which involves the training of artificial neural networks on large datasets, has had a significant impact on the field of anomaly detection. The capacity of deep neural networks to learn and identify patterns and non-linear relationships in data renders them particularly well-suited for tasks such as anomaly detection [27]. For example, Han et al. trained generative adversarial networks (GANs) to reconstruct brain images and successfully identify abnormal brains that suffered from Alzheimer's disease [21]. The authors demonstrated the efficacy of their network through training on healthy brain images and evaluating its ability to identify Alzheimer's

brain images, as evidenced by a higher reconstruction error. Additionally, the authors reported a correlation between the reconstruction error and the stage of the disease.

In my research, the identification of treatments impacting cellular state is attempted through anomaly detection, where deviation from the baseline distribution of untreated control cells serves as a baseline for indication of treatment effects.

High-Content Phenotypic Screening

High-content phenotypic screening (HCS) is the integration of automated microscopy and image analysis to collect and analyze large volumes of cells under different perturbations. It combines the efficiency of high-throughput screening of small-molecule with the ability of cellular imaging to collect quantitative data from complex biological systems [44-45]. HCS serves multiple applications such as clustering compounds into functional pathways, identifying phenotype signatures of specific diseases, and identifying drug mechanisms of action [16, 46-48]. One application of HCS is screening large compound libraries by performing experiments with each compound and assessing those that caused the largest deviation of the cell phenotype [49]. Those compounds that caused the largest phenotypic variation are called “hits” and are followed up in more depth to validate their phenotypic effect and whether it is linked to a desired function/cell state [15]. HCS technology, and specifically the Cell Painting Assay, has now expanded throughout all the different stages of the drug development process and is considered a mainstream technology in the pharmaceutical industry [45, 49-50].

Bakal et al. analyzed cells under various treatment conditions using quantitative features [48]. To transform the extracted features into biologically meaningful morphological indicators, they trained a set of neural networks that take single-cell morphological features and give them a score for each treatment. Using those scores, they managed to discriminate seven reference treatment conditions with distinctive morphologies.

Gustafsdottir et al. used clustering to identify compounds with similar effects [47]. To create the morphological profile for each cell they used a software called CellProfiler [51]. This computer vision software creates high-quality morphological profiles for each cell in the microscopy images, by extracting features such as cell shape, intensity, and texture, i.e., for each cell. The morphological profiles can then be further analyzed for clustering similar profiles, classifying, visualizing, etc. [9]. After they have extracted the profiles, they calculated the baseline profile of a cell which is how an average cell looks without any special treatment. Then, they calculated the Euclidean distance between each compound and the mock profile and identified the clusters based on that distance.

Way et al. performed a systematic comparison between the Cell Painting assay and a gene-based assay called Gene Expression [16]. Differently from Cell Painting, in Gene Expression the levels of different RNAs that are being produced in the cell are measured. Both the gene expression and morphology change as cells respond to perturbations. In their work, they compared the sample

diversity, feature diversity, and other differences in characteristics between the assays. They also introduce a systematic way for measuring the ability to reproduce readouts from repetitions of the same compound, named 'Percent Replicating'. In my work, I will use this measurement to assess different hits identification approaches.

Organelle-Organelle Organization

Organelles are molecular machines that define cell architecture and function. While many studies focus on the role and structure of individual organelles, it is less known how different organelles coordinate their organization and function within the cell [5-7].

Friedman et al. observed the contact sites between two organelles, mitochondria and the endoplasmic reticulum (ER), and examined the role of ER in mitochondrial division showing that the ER-mitochondrial interface is indeed vital for function [4]. Wu et al. and Cohen et al. showed that the ER promotes a variety of key functions in the cell via its membrane contact sites with various other organelles, such as Ca^{2+} exchange and organelle biogenesis [5, 8]. Furthermore, dysfunctions in the inter-organelle organization were recently linked to several diseases, for example, defects in membrane contact sites between the ER and the mitochondria are thought to have implications in cancer, neurodegenerative disorders, and diabetes [5-7].

Monitoring the dynamics of multiple organelles simultaneously in the same cell and analyzing their interactions is key to elucidating the complex interaction network that plays a critical role in every living organism. Valm et al. applied spectral imaging to simultaneously image six endogenously-engineered labeled organelles, setting the current upper limit for simultaneous live imaging of multiple organelles in the same cell [52]. For fixed cells, Bray et al. describe an assay called 'Cell Painting' [30]. In 'Cell Painting' each image is composed of five fluorescent channels marking different cell organelles. In my work, I will use images that have been through such a process to be able to capture the dependencies between the organelles.

Deep learning in microscopy

Deep learning has revolutionized machine-learning-driven fields by generating a hierarchy of features directly from the data during the model optimization process. This is achieved by training the deep neural network to map its input to the corresponding ground truth annotation. This has led to new and improved solutions for a variety of tasks, including bioimage analysis and computational tasks.

One example of a computational task is the work of Ouyang et al., who used neural networks to reconstruct super-resolution images from sparse, rapidly acquired localization images and/or widefield images [53]. Similarly, Ounkomol et al. were able to construct fluorescence images of the cell's organelles from the bright-field image, suggesting that the bright-field image contains spatial features of its organelles [54]. Inspired by this work, I will use a neural network, specifically the 'U-Net' architecture, to construct a fluorescence image of one organelle by learning the spatial

dependence between the different organelles. The ‘U-Net’ architecture, invented by Ronneberger et al., is considered state-of-the-art in biomedical image processing [55] due to its localization ability, i.e., each pixel in the output image is influenced by the corresponding pixel in the input image.

Results

Anomaly detection for hit identification in Cell Painting data

For the application of screening, I used a cell painting dataset containing cells treated with over 30,000 different compounds [10] (Fig. 1A). This dataset comprises 406 plates, each plate representing a distinct experiment. Each plate is composed of 64 control wells and 320 treated wells, with each well containing hundreds of cells that were perturbed by the same compound. Each compound appears in multiple wells across different plates. The number of replicates varies from one component to another, and the number of replicates is between one and eight. Approximately 25,000 compounds have four replicates, making out the bulk of the experiment. CellProfiler [51] was used to extract, for each cell, five vectors, one per channel, for a total of ~600 features that describe the cell's morphology (e.g., size, shape), fluorescent intensity, and texture. Hits in this screen were determined according to their phenotypic deviation from the control experiments.

Formulating hit detection as an anomaly detection problem allowed me to harness the data asymmetry, represented by high volumes of control data (Fig. 1B), to train a baseline "autoencoder" deep neural network (DNN). The autoencoder compressed the information in control cells' feature vectors thus capturing the multi-dimensional interrelationships between the features in the control cells (Fig. 1C). The autoencoder architecture consists of an "encoder" DNN that transforms the input feature representation of a single cell to a "latent vector", a compressed representation of the information in control cells' feature vectors. This latent vector is given to a decoder DNN that reconstructs back the input. The autoencoder is trained to minimize the discrepancy between the input and reconstructed feature vectors. In inference, the baseline autoencoder is applied on feature vectors derived from perturbed cells, and the corresponding reconstruction error is calculated (Fig. 1D). Perturbations that did not impose a major shift in the cell are expected to result in a low reconstruction error compared to control cells, indicating the corresponding treatments have no effect or a subtle effect on the cell morphology. Treatments that result in a high reconstruction error, compared to the underlying error distribution of the control cells are considered as hits, i.e., candidates for further validation and exploration (Fig. 1E).

The autoencoders were optimized for each organelle independently, allowing differential analysis of the organelle-specific effect of the treatment. As a complementary analysis, I trained an autoencoder on the concatenation of all five organelles combined and achieved similar results (Comparison in Supplementary Fig. 1).

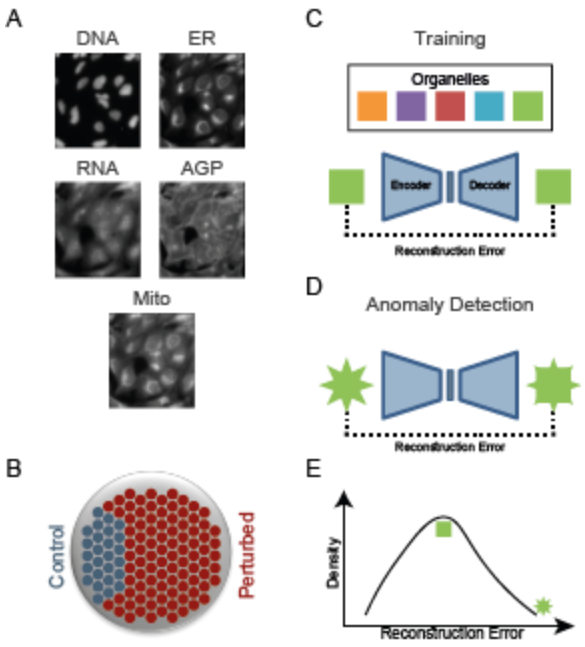


Figure 1. Detecting hits using Anomaly Detection. (A) The dataset consists of five different channels of organelles based on the 'Cell Painting' assay. (1) Nucleus (DNA), (2) Endoplasmic reticulum (ER), (3) Nucleoli, cytoplasmic RNA (RNA), (4) F-actin cytoskeleton, Golgi, plasma membrane (AGP), and (5) Mitochondria (Mito). **(B)** An illustration of a plate in the dataset. Each dot represents a well. Blue dots represent control wells, and each red dot represents a well with a unique treatment. **(C)** Training of an autoencoder, for each channel, to minimize the reconstruction error on the control cells. Each colored square represents an organelle's channel. **(D)** The trained autoencoder is used to predict the perturbed cells. A high magnitude of phenotypic deviation (star instead of square in the illustration) that is outside of the control variability, would cause a failure in reconstructing it by the autoencoder (anomalous square in the illustration), leading to high reconstruction errors. **(E)** The compounds are screened according to the reconstruction errors, and "hits" are identified as compounds that lead to high reconstruction errors with respect to the reconstruction error of the baseline control population.

Organelle anomaly detection as sensitive and complementary readout

I evaluated the performance of the suggested organelle anomaly detection and compared it to the current mainstream screening readout which is based on the statistical changes of the features from different organelles. To compare those readouts, I used a metric called 'Signature Strength' developed by [16]. Signature strength measures the fraction of features that passed a threshold for treatment. The signature strength depends on two parameters. One is the threshold, which I will refer to as the "signature threshold", which defines what is a "strong" feature. Second is the number of replicates as this metric provides higher value to compounds with more replicates, see Methods for full details. To systematically analyze the readouts, I present the distribution of signature strength using various signature threshold values while utilizing all compounds regardless of the number of replicates.

I systematically assessed the signature strength for different threshold cutoffs (Fig. 2A). Since the main focus of any screen is the "hits", those treatments that deviate significantly from the control and that are considered for further exploration, I assessed the full distribution, the top 5% and the top 1% of hits. When considering the full range of treatments, the statistical readout had higher signature strengths than the organelle anomaly readout. However, for subsets of treatments with a higher effect this trend flipped. Specifically, the top 1% of hits were characterized by higher signature strength for organelle anomaly concerning the statistical readout. Moreover, the sensitivity is gradually increasing when being more stringent about which features significantly deviate from the background, with increasing threshold, as measured systematically (Fig. 2B). One possible explanation for this observation is that, with lower signature thresholds, small variations are considered hits. The use of a method that does not account for the interdependencies of all features increases these small variations that otherwise would be described by the other features and thus increases the likelihood of false positives. By learning the interdependencies, my approach assigns a lower score for these treatments. Thus, the current statistical method appears more sensitive to these treatments than the proposed approach.

I next asked whether organelle anomaly finds more and different hits with respect to the statistical readout. I defined different thresholds for the signature strength, pooled all compounds that had a phenotypic alteration greater than a given threshold, and compared which hits were identified by each of these readouts. While there was a large agreement between the two readouts, as the criteria for deviation increased, the organelle anomaly method dominated with more hits, with respect to the statistical readout (Fig. 2C). These results suggest that organelle anomaly detection is a complementary readout that uncovers new compounds that could not be discovered otherwise.

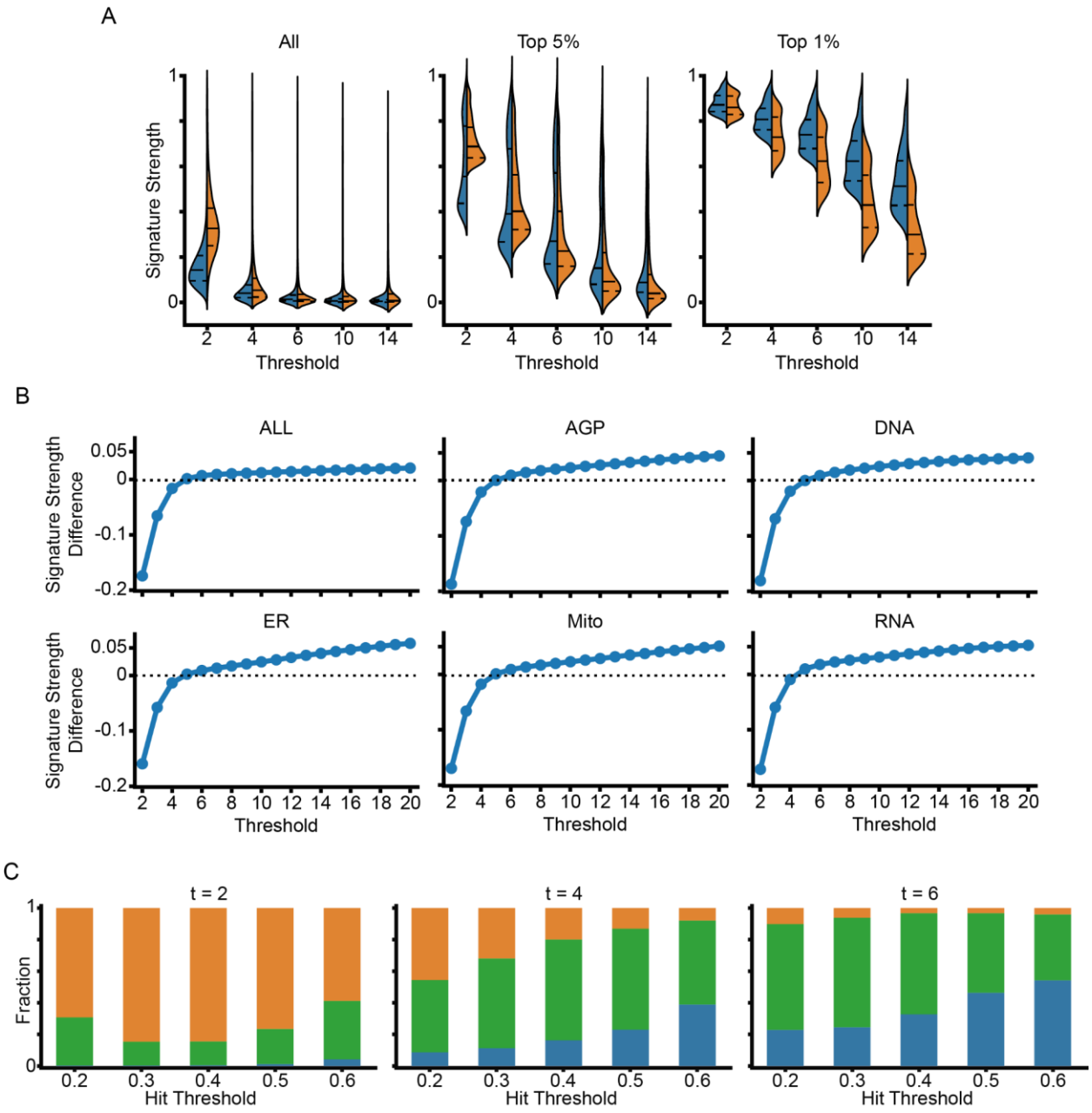


Figure 2. Organelle anomaly detection is a sensitive readout that finds complementary hits to the statistical readout. (A) The distribution of the signature strength with different signature thresholds (X-axis) for the organelle anomaly detection (blue) and statistical (orange) readouts. From left to right, each plot shows the distribution for all compounds, top 5%, and top 1% hits. **(B)** Each plot depicts the difference in signature strength between the organelle anomaly readout and the statistical readout (Y-axis) for each channel. The difference is computed across different thresholds that are used to compute the signature strength (X-axis). Each point is the mean difference across all the treatments. 'All' is the combined effect for all features across five organelles. **(C)** Each plot shows the fraction of 'hits' that are uniquely classified by each readout (Y-axis). In blue - organelle anomaly detection, In orange - statistical readout, and In green - both readouts identified those treatments. From left to right, each plot uses a different signature threshold for signature strengths, 2, 4, and 6. The "hits" are identified as treatments that cross a particular "hit threshold" (X-axis).

Reproducibility of organelle anomaly detection readout

Reproducibility is a critical aspect of the drug discovery process, as it helps to ensure that positive results are not due to chance or experimental error and increase confidence in the effectiveness of the treatment being screened. This is particularly important given the high failure rate of drug development, where only a small percentage of treatments that enter clinical trials ultimately receive regulatory approval [56]. Reproducibility is achieved through the independent replication of experiments by different research groups or by repeating the same experiment multiple times. This helps to ensure that the results are robust and that any potential biases or limitations of the study have been identified and addressed.

To analyze the reproducibility of the organelle anomaly readout compared to the statistical readout, I used a reproducibility measurement referred to as 'Percent Replicating' developed by [16]. This metric is based on the non-parametric test. For each treatment, the median correlation between its replications is compared to those of randomly generated compound combinations. Each treatment gets a p-value, based on this test, and the percent replicating is the proportion of the treatment population that gets a p-value below the significance level, see Methods for full details.

Here, I analyze treatments that have four repetitions in the dataset (Fig. 3) as they are the vast majority concerning their number of replicates. Analysis among all these treatments shows that organelle anomaly detection readout has slightly reduced percent replicating compared to the statistical readout with a significance level of 0.05 and the same percent replicating with a significance level of 0.01 (Fig. 3A). On the contrary, the proposed approach demonstrated a lower median correlation between replicates compared to the statistical readout, indicating that reproducibility was more challenging using this method. This discrepancy may be attributed to an over-amplification of the stringency of the treatments.

Reproducibility is especially critical when considering the 'hits' of treatments. Analysis of the percent replicating within the top ten percent of treatments based on signature strength revealed that the proposed approach exhibited lower levels of reproducibility in comparison to the current method. However, when the signature threshold for signature strength was increased, the gap between the approaches narrowed and even reversed (Fig 3B).

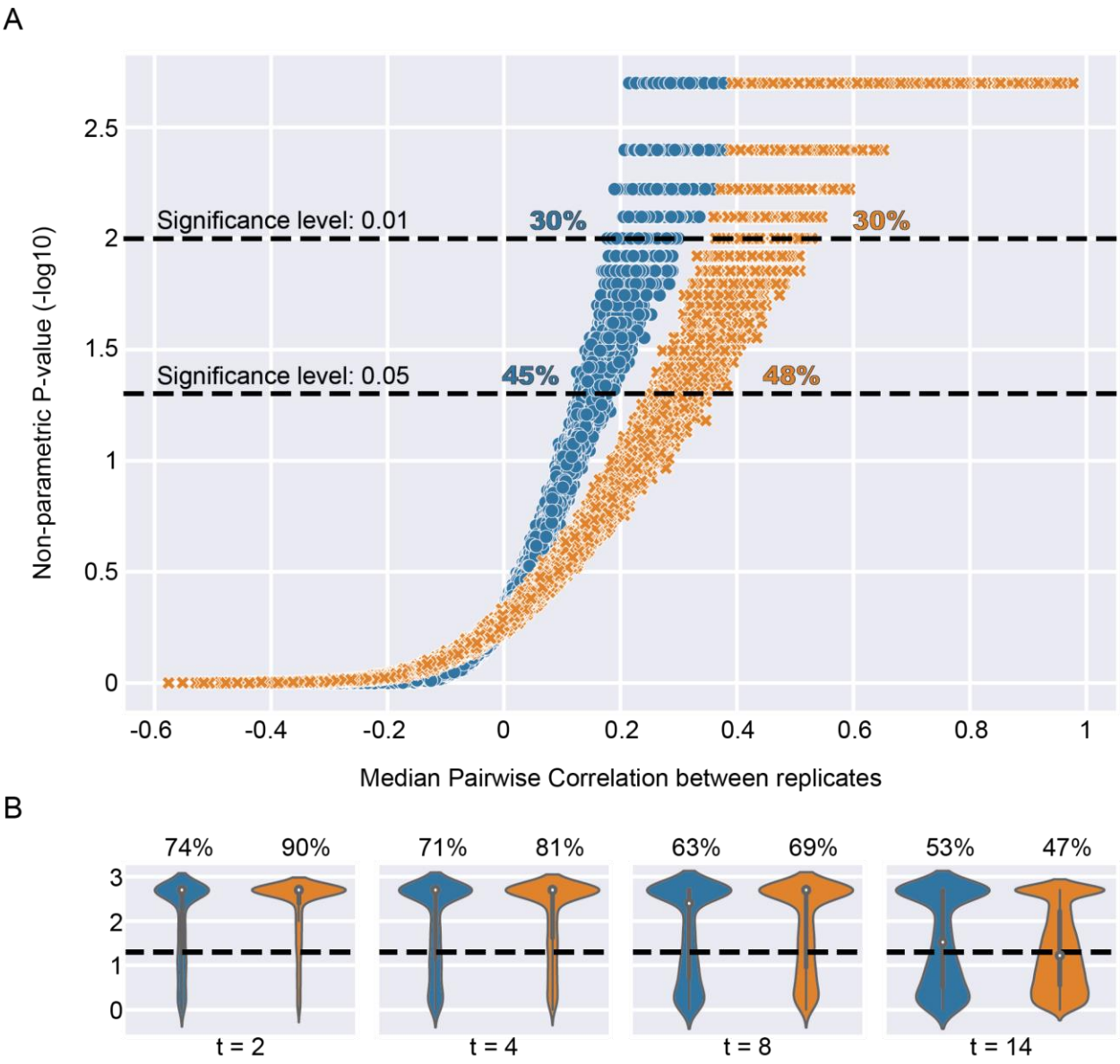


Figure 3. Organelle anomaly detection is slightly less reproducible. (A) The distribution of the reproducibility negative log non-parametric p-value (Y-axis) for the organelle anomaly detection (blue) and statistical (orange) readouts for each treatment corresponded to its median pairwise correlation (X-axis). From the bottom up, the black dashed lines represent significance levels of 0.05 and 0.01. The corresponding percent replicating is stated above each line, color encoded. **(B)** Each plot depicts the distribution of the negative log non-parametric p-value of the reproducibility test (Y-axis) for the organelle anomaly detection (blue) and statistical (orange) readouts for "hits". "Hits" are selected as the upper 10% of signature strength, using different signature thresholds, from left to right, 2, 4, 8, and 14. Above each distribution is the percent replicating using a significance level of 0.05 (shown as a dashed line).

Inter-organelle anomaly detection is a sensitive and complementary readout

Current measurements used for identifying hits in screens focus on the concatenation of features from individual organelles, thus uncovering compounds that lead to organelle perturbations (Fig. 4A, top-right). On the other hand, more and more recent studies indicate that the organelles' spatial relationship is the key to proper or improper functionality of the cell [5-6]. For instance, a recent study connected improper contact between the ER and the mitochondria thought to have implications in cancer, neurodegenerative disorders, and diabetes [5, 57]. Thus, the statistical approach may identify hits that seem to “fix” the organelle deviation but miss that these hits may perturb the inter-organelle organization or miss hits that “fix” inter-organelle deviations (Fig. 4A, bottom-left).

To identify compounds that perturb inter-organelle organization I designed an autoencoder that maps the concatenation of feature vectors from four organelles to the fifth (Fig. 4B). This is done for each 4-1 organelle mapping. During training, the autoencoders try to minimize the reconstruction error and thus learn the mappings between each organelle to the rest in control cells. When the inter-organelle dependencies are perturbed, the reconstruction errors significantly deviate from the reconstruction errors in the control cells, similar to what I described for a single organelle in Fig. 1.

Inter-organelle anomaly was more sensitive for the more prominent hits, as measured with those with high signature strengths (Fig. 4C) or ones that were selected based on a higher threshold on what defines a significant deviation of a feature (Fig. 4D). Inter-organelle anomaly mirrored the complementary capabilities of the organelle anomaly dominating the number of unique significant “hits” at higher hit identification criteria (Fig. 4E).

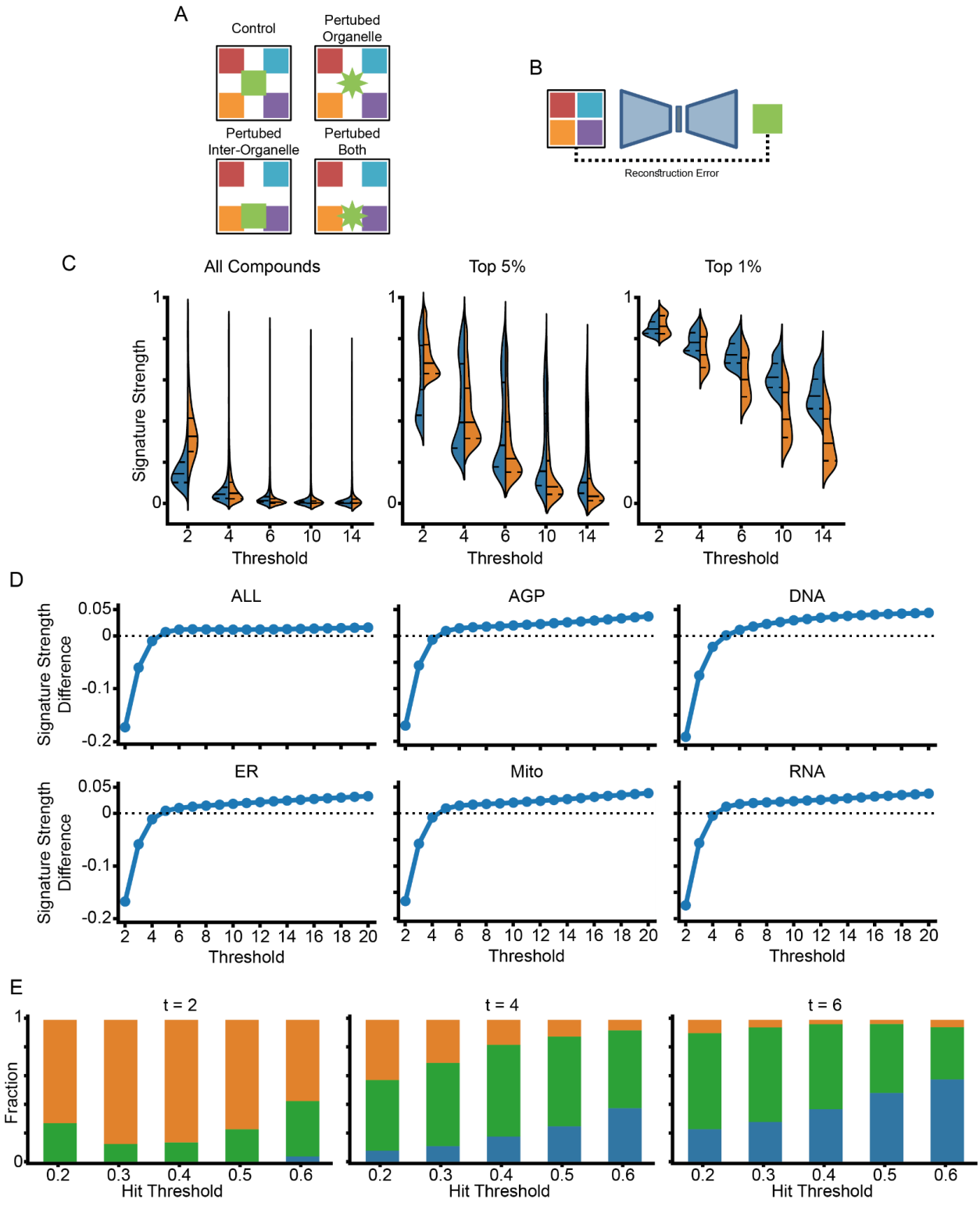


Figure 4. Inter-organelle anomaly detection framework provides a sensitive readout that finds complementary hits to statistical readout. (A) Schematic representation of the organization of five organelles (color code) in Control cells (upper left). Current readouts are designed for identifying alterations in single organelles (upper right: green star). My inter-organelle readout is designed to identify alterations in organelle-organelle dependencies (bottom left:

green square changing its position concerning the other squares). My readout is expected to capture alterations in inter-organelle that are not detected with the current readouts (bottom left), and to be more sensitive when both organelle and inter-organelle organization are altered (bottom-right). **(B)** Training of an autoencoder to minimize the 4-1 organelle mapping error on the control cells for each channel. **(C)** The distribution of the signature strength with different signature thresholds (X-axis) for the organelle anomaly detection (blue) and statistical (orange) readouts. From left to right, each plot shows the distribution for all compounds, top 5%, and top 1% hits. **(D)** Each plot depicts the difference in signature strength between the organelle anomaly readout and the statistical readout (Y-axis) for each channel. The difference is computed across different thresholds that are used to compute the signature strength (X-axis). Each point is the mean difference across all the treatments. 'All' is the combined effect for all features across five organelles. **(E)** Each plot shows the fraction of 'hits' that are uniquely classified by each readout (Y-axis). In blue - organelle anomaly detection, In orange - statistical readout, and In green - both readouts identified those treatments. From left to right, each plot uses a different signature threshold for signature strengths, 2, 4, and 6. The "hits" are identified as treatments that cross a particular "hit threshold" (X-axis).

Inter-organelle and organelle anomalies identify very similar hits

I compared the two anomaly-driven approaches, an organelle anomaly versus an inter-organelle anomaly. My analysis showed very similar distributions of signature strength (Fig. 5A), sensitivity analysis (Fig 5B), and complementariness analysis (Fig. 5C). The fact that both organelle anomaly and inter-organelle anomaly identified the same hits implies that they perform similar encoding the cells composition and intracellular organization. Thus, these results suggested that the inter-organelle approach relied mostly on the anomaly detection architecture itself as the power source rather than encoding the interrelationships between the organelles.

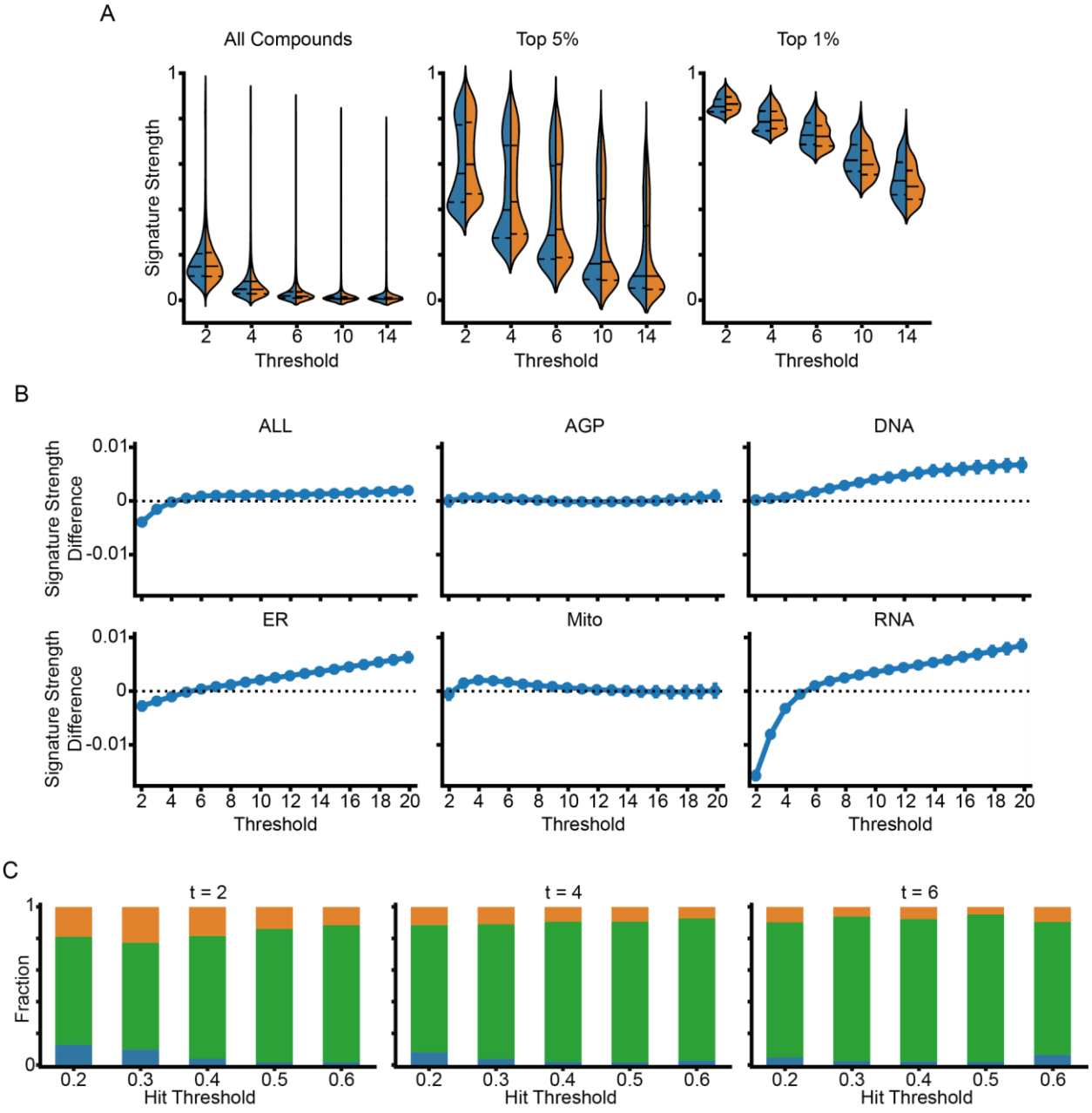


Figure 5. Inter-organelle anomaly detection is as sensitive and complementary as organelle anomaly detection. **(A)** The distribution of the signature strength with different signature thresholds (X-axis) for the inter-organelle anomaly detection (blue) and organelle anomaly detection (orange) readouts. From left to right, each plot shows the distribution for all compounds, top 5%, and top 1% hits. **(B)** Each plot depicts the difference in signature strength between the inter-organelle anomaly readout and the organelle anomaly detection readout (Y-axis) for each channel. The difference is computed across different thresholds that are used to compute the signature strength (X-axis). Each point is the mean difference across all the treatments. 'All' is the combined effect for all features across five organelles. **(C)** Each plot shows the fraction of 'hits' that are uniquely classified by each readout (Y-axis). In blue - inter-organelle anomaly detection, In orange - organelle anomaly detection readout, and In green - both readouts identified those treatments. From left to right, each plot uses a different signature threshold for signature strengths, 2, 4, and 6. The "hits" are identified as treatments that cross a particular "hit threshold" (X-axis).

Focus on uncorrelated organelles does not show the superiority of inter-organelle versus organelle anomalies

I hypothesized that the similarity in organelle and inter-organelle anomaly stems from the fact that the different organelles are highly correlated to one another. Organelle-organelle correlations were calculated as the average feature-wise Pearson correlation among the control cell population. These correlations showed high correlations between the DNA and ER channels and between the RNA and Mito channels (Fig. 6A). With this result, I focused on organelle combinations that were not strongly pairwise-correlated, specifically each autoencoder had two organelles as input and one as output. Details on the channel triplets used in Fig. 6B. Still, the results in terms of signature strength, sensitivity, and complementariness did not outperform the organelle anomaly measurement (Fig. 6). These results imply that capturing the relationship between the organelles by using the extracted features only cannot be attained as the features of the organelles are generally correlated.

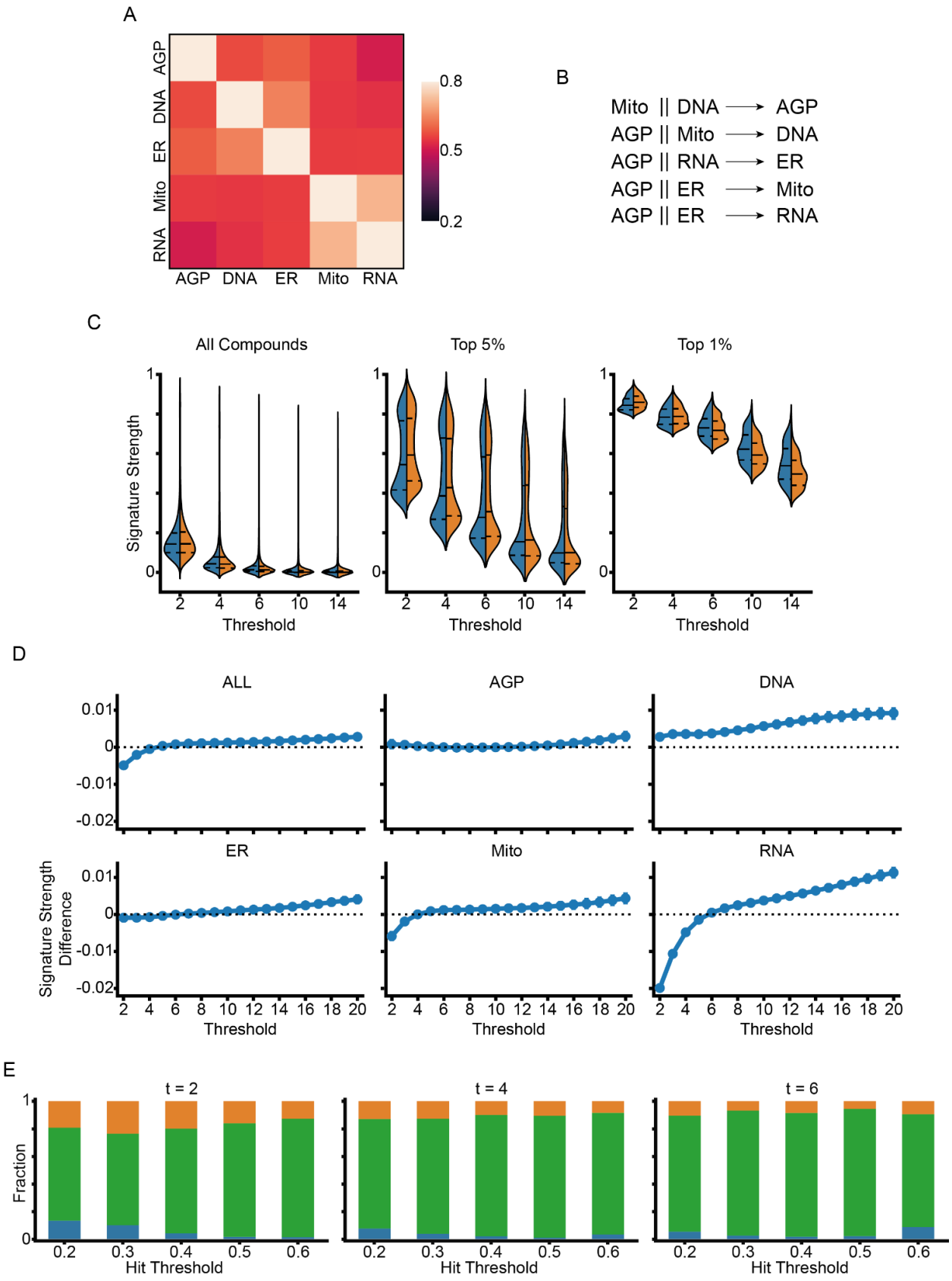

A
AGP
DNA
ER
Mito
RNA
0.8
0.5
0.2
B
Mito || DNA → AGP
AGP || Mito → DNA
AGP || RNA → ER
AGP || ER → Mito
AGP || ER → RNA
C
All Compounds
Top 5%
Top 1%
Signature Strength
Threshold
D
ALL
AGP
DNA
ER
Mito
RNA
Signature Strength Difference
Threshold
E
t = 2
t = 4
t = 6
Fraction
Hit Threshold

Figure 6. Removing correlated channels did not improve the performance of the inter-organelle approach. (A) The average paired single-cell level correlation between the channels. Each square depicts the correlation level between two different channels. **(B)** The organelles mapping in the inter-organelle anomaly detection after removing correlated channels. **(C)** The distribution of the signature strength with different signature thresholds (X-axis) for the inter-organelle anomaly detection (blue) and organelle anomaly detection (orange) readouts. From left to right, each plot shows the distribution for all compounds, top 5%, and top 1% hits. **(D)** Each plot depicts the difference in signature strength between the inter-organelle anomaly readout and the organelle anomaly detection readout (Y-axis) for each channel. The difference is computed across different thresholds that are used to compute the signature strength (X-axis). Each point is the mean difference across all the treatments. 'All' is the combined effect for all features across five organelles. **(E)** Each plot shows the fraction of 'hits' that are uniquely classified by each readout (Y-axis). In blue - inter-organelle anomaly detection, In orange - organelle anomaly detection readout, and In green - both readouts identified those treatments. From left to right, each plot uses a different signature threshold for signature strengths, 2, 4, and 6. The "hits" are identified as treatments that cross a particular "hit threshold" (X-axis).

Noisy reconstruction error leads to less reproducible anomaly detection

The lower reproducibility of anomaly detection methods was surprising, especially when focusing on the top hits. Anomaly detection was more sensitive than a direct assessment of feature deviation from the control, and this higher sensitivity was more pronounced as hits became more prominent. This describes an amplification of anomalous hits that intuitively should lead to enhanced reproducibility - deviating features are amplified above the noise, amplifying the correlation between replicates and leading to enhanced reproducibility. I looked closely at "hits" with high signature strengths that failed at reproducibility. I found that organelle anomaly detection features suffer from extremely high variation between the replicates (Fig. 7, blue), contrary to the low variations in the direct assessment of feature deviation from the control (Fig. 7, orange). This variability leads to lower correlations between replicates. The extreme z-scores values are probably caused due to the model's inability to generalize the control population distribution to the anomalous hits. A further examination of the generalization of the baseline model by adding a systematic technique to cancel out this noise, like adding *l2* regularization loss or using ensemble learning, is left for future work.

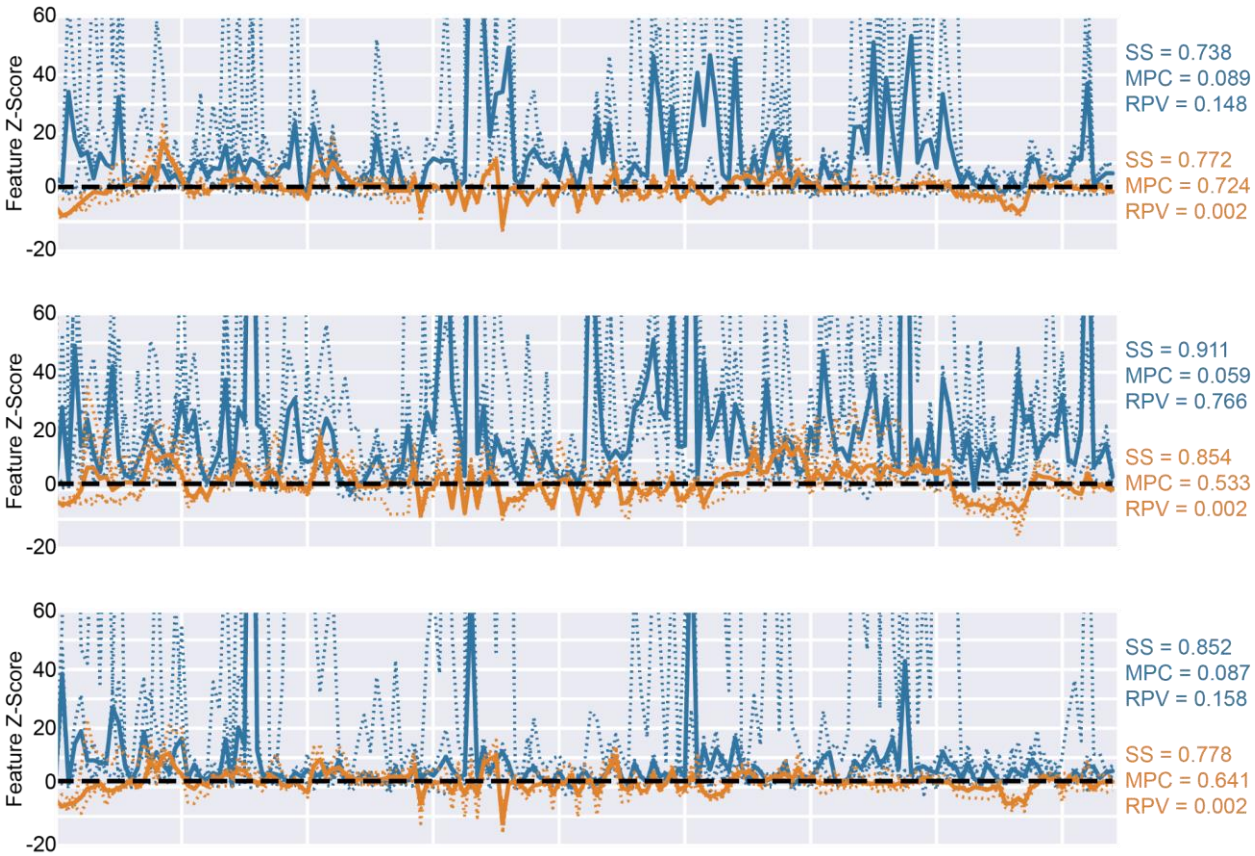


Figure 7. Anomaly detection readout contains noisy replicates. Each plot shows the deviation of each feature from the baseline population across the replicates of a single compound. The replicate values are shown as a dotted line and the median profile is in bold. Anomaly detection (blue), note that many replicate values are beyond the y-axis values

(not shown for visualization purposes), and direct assessment of feature deviation from the control (orange). On the right are the scores achieved by each compound. SS = signature strength, MPC = median pairwise correlation, RPV= reproducibility non-parametric p-value.

Image-based inter-organelle anomaly is more sensitive than organelle anomaly

As shown in Fig. 5-6, the inter-organelle approach suffers from the lack of spatial reference between the organelles that cannot be attained while using the extracted feature. Thus, my next attempt was to assess the idea of detecting anomalies directly from the raw image data. I used a U-Net network, introduced by [55], which is a convolutional autoencoder with unique localization ability. This ability is attained by using skip connections between the encoder and the decoder thus transferring localization information between the input and the output. This ability makes this network suited to the task of capturing the inter-organelle dependencies. For the inter-organelle anomalies, the target of the network was one organelle's channel, and the input was the other four organelles' channels. For the organelle anomaly, the network reconstructed the same organelle channel. The image data provides spatial context that is lost in the feature representations that encode organelle statistics without spatial context. My preliminary results were encouraging, showing that inter-organelle anomalies are more sensitive than organelle anomalies (Fig. 8). These results could imply that the spatial context in the images aid in the effective encoding of inter-organelle organization, but this has to be further validated in terms of reproducibility analysis, which is left for future work.

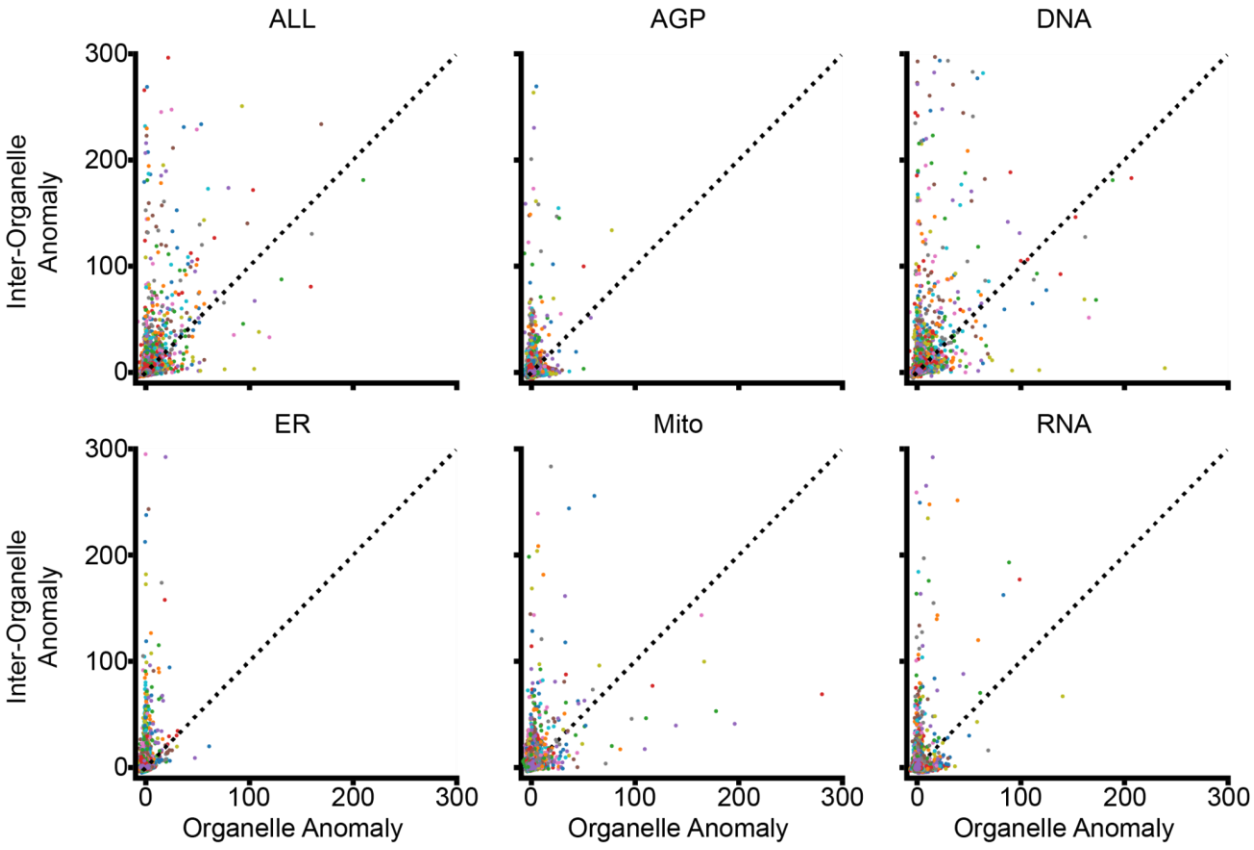


Figure 8. Inter-organelle visual anomaly detection framework provides a sensitive readout. Each plot depicts the normalized (concerning the controls) deviation in the corresponding organelle image anomaly detection (X-axis)

concerning the deviation in the inter-organelle image anomaly detection (Y-axis). Each data point is the normalized value for a given well. The color indicates the plate. “All” is the combined effect across the five channels (AGP, DNA, ER, MITO, RNA). The vast majority of “hits” are above the Y = X diagonal establishing inter-organelle visual anomaly detection as a sensitive readout for screening.

Discussion and outlook

High-content image-based phenotypic screening is a critical step in basic science, translational science, and drug discovery, as a first, cheap, and fast assay to identify treatments that alter the cells' morphological state. Current approaches to measuring how a perturbation altered the cells' morphology are based on statistically evaluating the deviation from the baseline control population. These approaches hold the implicit assumption of independence between the morphological features extracted to represent the cell's state. Here, I suggested anomaly detection as an approach that takes into account the interdependence between features. Anomaly detection is especially suitable for screening with many replicates of control experiments to statistically define reliable baseline patterns, and many perturbations, each with a few replicates that are assessed with respect to their deviation from the baseline. I assessed this new concept in three modes, focusing on tabular/vector representations of single cells (organelle, inter-organelle), and image-based inter-organelle.

Anomaly detection was found to be more sensitive amplified existing hits and identified new hits that were not identified by direct measurement of the deviation from the baseline distribution. However, hits identified by anomaly detection were less reproducible compared to the common statistical approach. I detected that, when focusing on the hits' replicates, they all show high deviation from the baseline control distribution, leading to high sensitivity. On the other hand, there is high variability between the replicates which leads to a low correlation and thus low reproducibility (Fig. 7). This variability is rooted in the fact that the autoencoder did not see the anomalous hits during training and thus contains some noise in inference time. This can be addressed by adding more generalization techniques to the training phase to make the model more stable, which is left for future work as the ability to reproduce the signals across replicates is vital for the readout reliability [56]. Applying regularization techniques to models can improve their performance and generalization ability. Adding *l1* or *l2* regularization can prevent overfitting on the control data by encouraging smaller weights, while adding dropout layers to the model can prevent overreliance on any one feature. Another way to reduce that noise would be by introducing perturbated data into the training phase, this will make the model more robust to anomalous hits during inference time.

High-content image-based phenotypic profiling has multiple applications beyond screening. One application is clustering compounds into functional pathways by identifying similarities between the compounds' phenotypic effects [46-47]. Anomaly detection may improve clustering as it is a more sensitive readout that, in turn, will lead to more strictly differentiated clusters. Another use case is identifying phenotype signatures of specific diseases or drug mechanisms of action [16, 46, 48], where anomaly detection can also improve the readout. In this case, anomaly detection can create a more accurate profile as it is more sensitive to the compounds' effect. In both cases, inter-organelle readouts can provide complementary information regarding the phenotypic effect that is currently not measured.

Indeed, the most promising application of anomaly detection is the attempt to measure the deviation in interrelationships between the organelles. The inter-organelle organization is vital to the cell's functionality and has implications in diseases like cancer, neurodegenerative disorders, and diabetes [4-8]. Therefore, a readout based on this organization may detect meaningful insights on perturbation effects that harm/correct those relations. Although the feature-based anomaly outperformed the common statistical approach, it failed to capture the relationship between the organelles from the single-cell extracted morphological features. The main reason for that is the high correlations between the different channels (Fig. 6A). These correlations led to the inter-organelle model being similar to the organelle model, which in turn led to similar performance (Fig. 6C-E). Another reason is that these organelle statistical features lack the spatial context that is critical for measuring spatial dependencies between organelles. My preliminary results on image-based inter-organelle anomalies (Fig. 8) hold great promise for future advances.

Methods

Dataset

I used the open dataset by Bray et al [10]. This dataset contains Cell Painting data of cells under treatment with 30,000 small-molecule compounds. The dataset is composed of 406 different plates. Each plate has approximately 64 control wells and 320 uniquely treated wells. Most of the treatments have several repetitions over different plates. The repetitions are used to measure the reproducibility of my approach. Each image contains five fluorescent channels representing different cell structures: (1) Nucleus (DNA), (2) Endoplasmic reticulum (ER), (3) Nucleoli, cytoplasmic RNA (RNA), (4) F-actin cytoskeleton, Golgi, plasma membrane (AGP) and (5) Mitochondria (Mito). Each image was of size 696x520 pixels. The dataset contains raw image data and single-cell morphological features extracted by the Cell Profiler. The morphological profile contains 86 features extracted from each fluorescent channel (except DNA which is composed of 70 features).

Data split approach

Each experimental plate has multiple control and treated wells. The control wells were used both for training the autoencoders and for accounting for inter-plate batch effects for hits identification during inference. For training, I used control wells to learn the baseline distribution of the control's data. For hits identification, I used control wells to account for the inter-plate batch effects. Thus, I split each plate into four subpopulations: (1) 12 control wells for training, (2) 4 control wells for validation, (3) 48 control wells for hits identification at inference, and (4) treated wells for evaluation at inference. Control wells for training and validation were sampled randomly from all well locations in the plate.

Morphological features analysis

The morphological features, extracted by the CellProfiler, are single-cell features that encompass a diverse range of statistics of cellular shape and adjacency, as well as statistics on intensity and texture that are calculated in each channel [10]. Based on these features, the analysis was conducted at the single-cell level.

Feature Selection and normalization

Pycytominer's [58] high correlation and low variance, both with the default parameters, were used for feature selection. High correlation: dropping features that are correlated with other features with a coefficient greater than 0.9. Low variance: dropping features that exhibit the same value for 95% of the cells. This feature selection was analyzed on the training subset and at the end, the number of features attributed for each fluorescent channel dropped from 86 (70 for DNA) to approximately 32 features. Feature-wise normalization was used using Z-score normalization

before introducing the autoencoders. The mean and standard deviation were estimated from the control wells used for training.

Autoencoder architecture

The autoencoders' architecture was derived from input and output sizes. The decoder and the encoder were symmetric. Each subsequent encoder/decoder layer was smaller/larger than the previous layer by a factor of 2 until reaching the latent encoding size of 8. The latent size was selected as the smallest size that succeeded to converge, validation loss < 1. For example, for input and output sizes of 32 and latent size of 8, the resulting architecture will contain five layers of size 32-16-8-16-32. Each layer contains a Relu activation layer except the final layer. The loss function of the network is the mean square error between the reconstructed output and the corresponding real target. Each autoencoder was trained until convergence or a maximum of 30 epochs. Convergence was defined as validation loss change < 0.002 for four epochs. The batch size was 2048 cells and the learning rate was 1.5e-3.

Feature-based treatments evaluation

I used the trained autoencoders and calculated the feature-wise prediction squared error for each cell. I used the prediction error for anomaly detection assessment compared to the direct assessment of feature deviation from the control. I pooled all cells in each well and used the average prediction error / features deviation as the well's phenotypic profile, and z-score normalized these profiles using the plate's control wells. These steps resulted in a vector of feature Z-scores for each cell.

Note that in anomaly detection, a negative Z-score implies lower reconstruction errors, i.e., closer to the baseline distribution. Therefore, to enable fair comparison with feature deviation (where negative z-scores imply a phenotype), I took the absolute Z-score values for each feature when measuring the phenotypic alteration's magnitude.

Measuring the magnitude of phenotypic alteration

The well's 'Signature Strength' score is the fraction of features above a predefined threshold, defining the magnitude of a phenotypic alteration that was induced by treatment, with respect to the control [59]. For treatment with multiple replicates, we first multiply the treatment's entries by the square root of the number of replicates and then calculate the average of the corresponding wells' signature strengths [59].

Measuring reproducibility

I used a metric called 'Percent Replicating' [16]. For each treatment, I calculate a bootstrapping statistical test rejecting the null hypothesis that the median pairwise correlation between replicates is within the distribution of pairwise correlations of different treatments. Statistical significance was calculated as follows. For each treatment x with n replicates, I created a null distribution of

1000 x *n* wells, where each of the *n* wells was a unique treatment (and not *x*). The median pairwise correlation between the *n* wells' feature vectors was calculated, and their distribution was evaluated concerning the median pairwise correlation of *x*'s *n* replicates. The p-value of treatment *x* was determined based on the fraction of the null distribution that had higher median pairwise correlations. The 'Percent Replicating' measurement was the proportion of treatments with a p-value < 0.05.

Images analysis

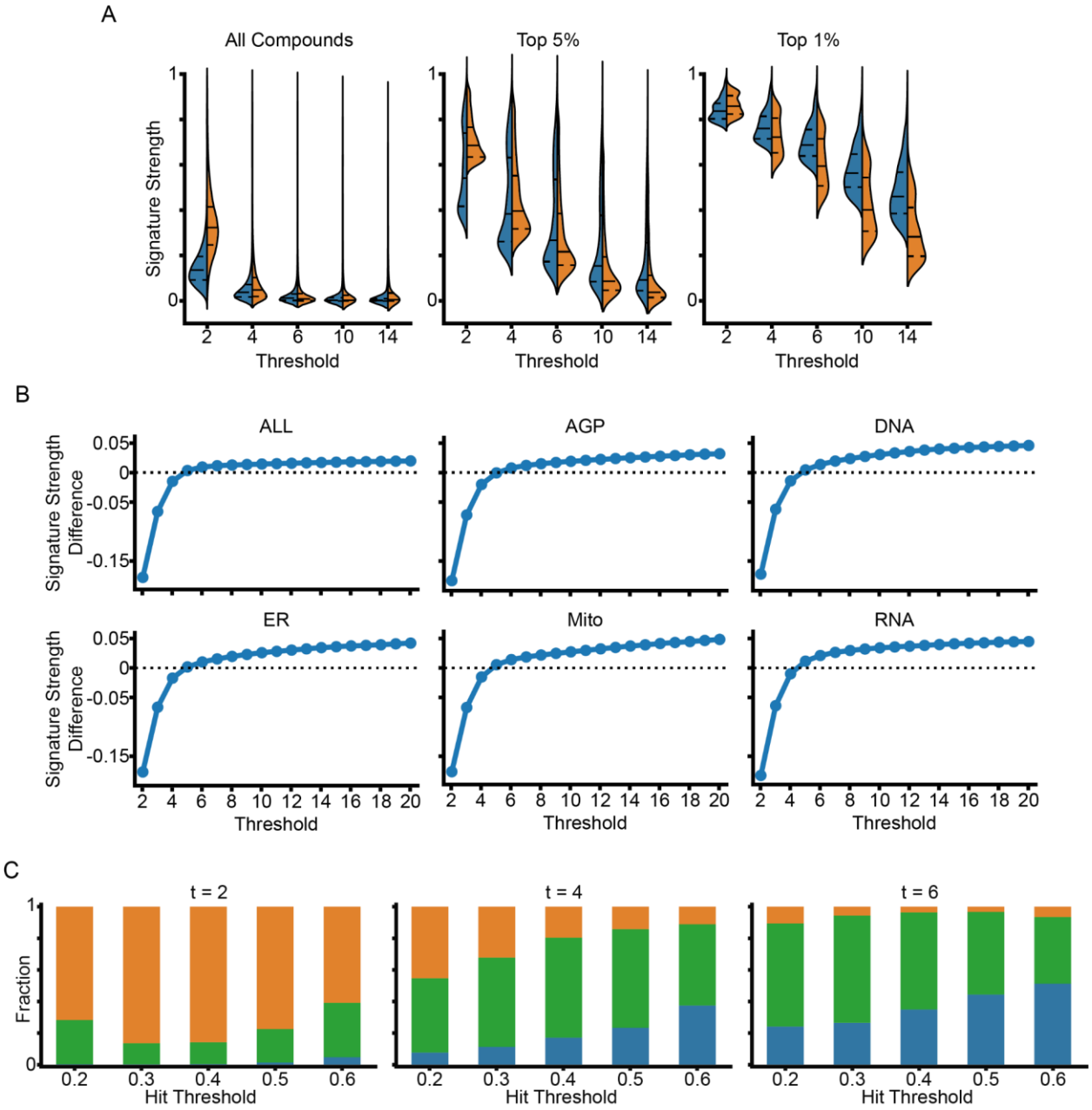
Training an image-based baseline model

As a baseline model, I used the U-Net architecture [55]. The model received as input an image patch of size 232x260 pixels, with a batch size of 16 image patches and a learning rate of 1.5e-4. The convergence criteria were the ones used in the tabular autoencoder described earlier. During each epoch of the model, a random crop of the image is generated. Each crop is normalized channel-wise using Z-score normalizations.

Image-base treatments evaluation

I used the trained models to predict the treated and control images that were reserved for hit identification. I calculated the pixel-wise mean squared error for each image's crop prediction, between the original image's region and the predicted one. For each image, I took enough crops to cover the entire image, then the average error is taken as the cumulative image reconstruction error. Each treated well is captured by several images, thus I took the average of their reconstruction errors. The resulting value was then utilized to determine the Z-score of the treated well in question, using the control wells of the same plate as the baseline for normalization. This results in a single Z-score for each treated well, indicating deviation from the baseline population.

Supplementary Figures



Supplementary Figure 1. Concatenate organelles anomaly detection is a sensitive readout that finds complementary hits to the statistical readout. **(A)** The distribution of the signature strength with different signature thresholds (X-axis) for the organelle anomaly detection (blue) and statistical (orange) readouts. From left to right, each plot shows the distribution for all compounds, top 5%, and top 1% hits. **(B)** Each plot depicts the difference in signature strength between the organelle anomaly readout and the statistical readout (Y-axis) for each channel. The difference is computed across different thresholds that are used to compute the signature strength (X-axis). Each point is the mean difference across all the treatments. 'All' is the combined effect for all features across five organelles. **(C)** Each plot shows the fraction of 'hits' that are uniquely classified by each readout (Y-axis). In blue - organelle anomaly detection, In orange - statistical readout, and In green - both readouts identified those treatments. From left to right,

each plot uses a different signature threshold for signature strengths, 2, 4, and 6. The “hits” are identified as treatments that cross a particular “hit threshold” (X-axis).

References

[1] Pincus, Z., and J. A. Theriot. 2007. "Comparison of Quantitative Methods for Cell-Shape Analysis." *Journal of Microscopy* 227 (2): 140–56. https://doi.org/10.1111/j.1365-2818.2007.01799.x.

[2] Keren, Kinneret, Zachary Pincus, Greg M. Allen, Erin L. Barnhart, Gerard Marriott, Alex Mogilner, and Julie A. Theriot. 2008. "Mechanism of Shape Determination in Motile Cells." *Nature* 453 (7194): 475–80. https://doi.org/10.1038/nature06952.

[3] Wu, Pei-Hsun, Daniele M. Gilkes, Jude M. Phillip, Akshay Narkar, Thomas Wen-Tao Cheng, Jorge Marchand, Meng-Horng Lee, Rong Li, and Denis Wirtz. 2020. "Single-Cell Morphology Encodes Metastatic Potential." *Science Advances* 6 (4). https://doi.org/10.1126/sciadv.aaw6938.

[4] Friedman, Jonathan R., Laura L. Lackner, Matthew West, Jared R. DiBenedetto, Jodi Nunnari, and Gia K. Voeltz. 2011. "ER Tubules Mark Sites of Mitochondrial Division." *Science* 334 (6054): 358–62. https://doi.org/10.1126/science.1207385.

[5] Wu, Haoxi, Pedro Carvalho, and Gia K. Voeltz. 2018. "Here, There, and Everywhere: The Importance of ER Membrane Contact Sites." *Science* 361 (6401). https://doi.org/10.1126/science.aan5835.

[6] Shai, Nadav, Eden Yifrach, Carlo W. T. van Roermund, Nir Cohen, Chen Bibi, Lodewijk IJlst, Laetitia Cavellini, et al. 2018. "Systematic Mapping of Contact Sites Reveals Tethers and a Function for the Peroxisome-Mitochondria Contact." *Nature Communications* 9 (1). https://doi.org/10.1038/s41467-018-03957-8.

[7] Scorrano, Luca, Maria Antonietta De Matteis, Scott Emr, Francesca Giordano, György Hajnóczky, Benoît Kornmann, Laura L Lackner, et al. 2019. "Coming Together to Define Membrane Contact Sites." *Nature Communications* 10 (1): 1–11.

[8] Cohen, Sarah, Alex M Valm, and Jennifer Lippincott-Schwartz. 2018. "Interacting Organelles." *Current Opinion in Cell Biology* 53 (August): 84–91. https://doi.org/10.1016/j.ceb.2018.06.003.

[9] Caicedo, Juan C, Sam Cooper, Florian Heigwer, Scott Warchal, Peng Qiu, Csaba Molnar, Aliaksei S Vasilevich, et al. 2017. "Data-Analysis Strategies for Image-Based Cell Profiling." *Nature Methods* 14 (9): 849–63. https://doi.org/10.1038/nmeth.4397.

[10] Bray, Mark-Anthony, Sigrun M Gustafsdottir, Mohammad H Rohban, Shantanu Singh, Vebjorn Ljosa, Katherine L Sokolnicki, Joshua A Bittker, et al. 2017. "A Dataset of Images and Morphological Profiles of 30 000 Small-Molecule Treatments Using the Cell Painting Assay." *GigaScience* 6 (12). https://doi.org/10.1093/gigascience/giw014.

[11] Lu, Alex X, Yolanda T Chong, Ian Shen Hsu, Bob Strome, Louis-Francois Handfield, Oren Kraus, Brenda J Andrews, and Alan M Moses. 2018. "Integrating Images from Multiple Microscopy Screens Reveals Diverse Patterns of Change in the Subcellular Localization of Proteins." *eLife* 7 (April). https://doi.org/10.7554/elife.31872.

[12] Lukonin, Ilya, Denise Serra, Ludivine Challet Meylan, Katrin Volkmann, Janine Baaten, Rui Zhao, Shelly Meeusen, et al. 2020. "Phenotypic Landscape of Intestinal Organoid Regeneration." *Nature* 586 (7828): 275–80. https://doi.org/10.1038/s41586-020-2776-9.

[13] Goglia, Alexander G., Maxwell Z. Wilson, Siddhartha G. Jena, Jillian Silbert, Lena P. Basta, Danelle Devenport, and Jared E. Toettcher. 2020. "A Live-Cell Screen for Altered Erk Dynamics Reveals Principles of Proliferative Control." *Cell Systems* 10 (3): 240–253.e6. https://doi.org/10.1016/j.cels.2020.02.005.

[14] Zaritsky, Assaf, Yun-Yu Tseng, M. Angeles Rabadán, Shefali Krishna, Michael Overholtzer, Gaudenz Danuser, and Alan Hall. 2017. "Diverse Roles of Guanine Nucleotide Exchange Factors in Regulating Collective Cell Migration." *Journal of Cell Biology* 216 (6): 1543–56. https://doi.org/10.1083/jcb.201609095.

[15] Chandrasekaran, Srinivas Niranj, Hugo Ceulemans, Justin D. Boyd, and Anne E. Carpenter. 2020. "Image-Based Profiling for Drug Discovery: Due for a Machine-Learning Upgrade?" *Nature Reviews Drug Discovery* 20 (2): 145–59. https://doi.org/10.1038/s41573-020-00117-w.

[16] Way, Gregory P., Ted Natoli, Adeniyi Adeboye, Lev Litichevskiy, Andrew Yang, Xiaodong Lu, Juan C. Caicedo, et al. 2022. "Morphology and Gene Expression Profiling Provide Complementary Information for Mapping Cell State." *Cell Systems* 13 (11): 911–923.e9. https://doi.org/10.1016/j.cels.2022.10.001.

[17] Thudumu, Srikanth, Philip Branch, Jiong Jin, and Jugdutt (Jack) Singh. 2020. "A Comprehensive Survey of Anomaly Detection Techniques for High Dimensional Big Data." *Journal of Big Data* 7 (1). https://doi.org/10.1186/s40537-020-00320-x.

[18] Viana, Matheus P., Jianxu Chen, Theo A. Knijnenburg, Ritvik Vasan, Calysta Yan, Joy E. Arakaki, Matte Bailey, et al. 2023. "Integrated Intracellular Organization and Its Variations in Human iPS Cells." *Nature* 613 (7943): 345–54. https://doi.org/10.1038/s41586-022-05563-7.

[19] Chandola, Varun, Arindam Banerjee, and Vipin Kumar. 2009. "Anomaly Detection." *ACM Computing Surveys* 41 (3): 1–58. https://doi.org/10.1145/1541880.1541882.

[20] Fernando, Tharindu, Harshala Gammulle, Simon Denman, Sridha Sridharan, and Clinton Fookes. 2021. "Deep Learning for Medical Anomaly Detection a Survey." *ACM Computing Surveys* 54 (7): 1–37. https://doi.org/10.1145/3464423.

[21] Han, Changhee, Leonardo Rundo, Kohei Murao, Tomoyuki Noguchi, Yuki Shimahara, Zoltán Ádám Milacski, Saori Koshino, Evis Sala, Hideki Nakayama, and Shin'ichi Satoh. 2021. "MADGAN: Unsupervised Medical Anomaly Detection GAN Using Multiple Adjacent Brain MRI Slice Reconstruction." *BMC Bioinformatics* 22 (S2). https://doi.org/10.1186/s12859-020-03936-1.

[22] Naidoo, Krishnan, and Vukosi Marivate. 2020. "Unsupervised Anomaly Detection of Healthcare Providers Using Generative Adversarial Networks." In *Lecture Notes in Computer Science*, 419–30. Springer International Publishing. https://doi.org/10.1007/978-3-030-44999-5_35.

[23] Meira, Jorge, Rui Andrade, Isabel Praça, João Carneiro, Verónica Bolón-Canedo, Amparo Alonso-Betanzos, and Goreti Marreiros. 2019. "Performance Evaluation of Unsupervised Techniques in Cyber-Attack Anomaly Detection." *Journal of Ambient Intelligence and Humanized Computing* 11 (11): 4477–89. https://doi.org/10.1007/s12652-019-01417-9.

[24] Breunig, Markus M, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. 2000. "LOF: Identifying Density-Based Local Outliers." In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, 93–104.

[25] Hautamaki, V., I. Karkkainen, and P. Franti. 2004. "Outlier Detection Using k-Nearest Neighbour Graph." In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.* IEEE. https://doi.org/10.1109/icpr.2004.1334558.

[26] Liu, Fei Tony, Kai Ming Ting, and Zhi-Hua Zhou. 2008. "Isolation Forest." In *2008 Eighth IEEE International Conference on Data Mining*. IEEE. https://doi.org/10.1109/icdm.2008.17.

[27] Pang, Guansong, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. 2021. "Deep Learning for Anomaly Detection." *ACM Computing Surveys* 54 (2): 1–38. https://doi.org/10.1145/3439950.

[28] Yang, Jingkang, Kaiyang Zhou, Yixuan Li, and Ziwei Liu. 2021. "Generalized Out-of-Distribution Detection: A Survey." arXiv. https://doi.org/10.48550/ARXIV.2110.11334.

[29] Ruff, Lukas, Jacob R. Kauffmann, Robert A. Vandermeulen, Gregoire Montavon, Wojciech Samek, Marius Kloft, Thomas G. Dietterich, and Klaus-Robert Muller. 2021. "A Unifying Review of Deep and Shallow Anomaly Detection." *Proceedings of the IEEE* 109 (5): 756–95. https://doi.org/10.1109/jproc.2021.3052449.

[30] Bray, Mark-Anthony, Shantanu Singh, Han Han, Chadwick T Davis, Blake Borgeson, Cathy Hartland, Maria Kost-Alimova, Sigrun M Gustafsdottir, Christopher C Gibson, and Anne E Carpenter. 2016. "Cell Painting, a High-Content Image-Based Assay for Morphological Profiling Using Multiplexed Fluorescent Dyes." *Nature Protocols* 11 (9): 1757–74. https://doi.org/10.1038/nprot.2016.105.

[31] Rohban, Mohammad Hossein, Shantanu Singh, Xiaoyun Wu, Julia B Berthet, Mark-Anthony Bray, Yashaswi Shrestha, Xaralabos Varelas, Jesse S Boehm, and Anne E Carpenter. 2017. "Systematic Morphological Profiling of Human Gene and Allele Function via Cell Painting." *eLife* 6 (March). https://doi.org/10.7554/elife.24060.

[32] Grubbs, Frank E. 1969. "Procedures for Detecting Outlying Observations in Samples." *Technometrics* 11 (1): 1–21. https://doi.org/10.1080/00401706.1969.10490657.

[33] Aggarwal, Charu C., and Philip S. Yu. 2008. "Outlier Detection with Uncertain Data." In *Proceedings of the 2008 SIAM International Conference on Data Mining*. Society for Industrial; Applied Mathematics. https://doi.org/10.1137/1.9781611972788.44.

[34] Horn, Paul S, Lan Feng, Yanmei Li, and Amadeo J Pesce. 2001. "Effect of Outliers and Nonhealthy Individuals on Reference Interval Estimation." *Clinical Chemistry* 47 (12): 2137–45. https://doi.org/10.1093/clinchem/47.12.2137.

[35] Solberg, Helge Erik, and Ari Lahti. 2005. "Detection of Outliers in Reference Distributions: Performance of Horn's Algorithm." *Clinical Chemistry* 51 (12): 2326–32. https://doi.org/10.1373/clinchem.2005.058339.

[36] Lim, Shu Yun, and Andy Jones. 2008. "Network Anomaly Detection System: The State of Art of Network Behaviour Analysis." In *2008 International Conference on Convergence and Hybrid Information Technology*. IEEE. https://doi.org/10.1109/ichit.2008.249.

[37] Roy, Jean. 2010. "Rule-Based Expert System for Maritime Anomaly Detection." In *SPIE Proceedings*, edited by Edward M. Carapezza. SPIE. https://doi.org/10.1117/12.849131.

[38] Xu, Xiaodan, Huawen Liu, and Minghai Yao. 2019. "Recent Progress of Anomaly Detection." *Complexity* 2019 (January): 1–11. https://doi.org/10.1155/2019/2686378.

[39] Samariya, Durgesh, and Amit Thakkar. 2021. "A Comprehensive Survey of Anomaly Detection Algorithms." *Annals of Data Science*, November. https://doi.org/10.1007/s40745-021-00362-9.

[40] Ramaswamy, Sridhar, Rajeev Rastogi, and Kyuseok Shim. 2000. "Efficient Algorithms for Mining Outliers from Large Data Sets." In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*. ACM. https://doi.org/10.1145/342009.335437.

[41] Siddiqui, Md Amran, Jack W. Stokes, Christian Seifert, Evan Argyle, Robert McCann, Joshua Neil, and Justin Carroll. 2019. "Detecting Cyber Attacks Using Anomaly Detection with Explanations and Expert Feedback." In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. https://doi.org/10.1109/icassp.2019.8683212.

[42] Aryal, Sunil, Kai Ming Ting, Jonathan R. Wells, and Takashi Washio. 2014. "Improving iForest with Relative Mass." In *Advances in Knowledge Discovery and Data Mining*, 510–21. Springer International Publishing. https://doi.org/10.1007/978-3-319-06605-9_42.

[43] Zhang, Xuyun, Wanchun Dou, Qiang He, Rui Zhou, Christopher Leckie, Ramamohanarao Kotagiri, and Zoran Salcic. 2017. "LSHiForest: A Generic Framework for Fast Tree Isolation Based Ensemble Anomaly Analysis." In *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*. IEEE. https://doi.org/10.1109/icde.2017.145.

[44] Carpenter, Anne E. 2007. "Image-Based Chemical Screening." *Nature Chemical Biology* 3 (8): 461–65. https://doi.org/10.1038/nchembio.2007.15.

[45] Zanella, Fabian, James B. Lorens, and Wolfgang Link. 2010. "High Content Screening: Seeing Is Believing." *Trends in Biotechnology* 28 (5): 237–45. https://doi.org/10.1016/j.tibtech.2010.02.005.

[46] Zock, Joseph. 2009. "Applications of High Content Screening in Life Science Research." *Combinatorial Chemistry &Amp$\mathsemicolon$ High Throughput Screening* 12 (9): 870–76. https://doi.org/10.2174/138620709789383277.

[47] Gustafsdottir, Sigrun M., Vebjorn Ljosa, Katherine L. Sokolnicki, J. Anthony Wilson, Deepika Walpita, Melissa M. Kemp, Kathleen Petri Seiler, et al. 2013. "Multiplex Cytological Profiling Assay to Measure Diverse Cellular States." Edited by Michael A Mancini. *PLoS ONE* 8 (12): e80999. https://doi.org/10.1371/journal.pone.0080999.

[48] Bakal, Chris, John Aach, George Church, and Norbert Perrimon. 2007. "Quantitative Morphological Signatures Define Local Signaling Networks Regulating Cell Morphology." *Science* 316 (5832): 1753–56. https://doi.org/10.1126/science.1140324.

[49] Haney, Steven A., Peter LaPan, Jing Pan, and Jing Zhang. 2006. "High-Content Screening Moves to the Front of the Line." *Drug Discovery Today* 11 (19-20): 889–94. https://doi.org/10.1016/j.drudis.2006.08.015.

[50] Bickle, Marc. 2010. "The Beautiful Cell: High-Content Screening in Drug Discovery." *Analytical and Bioanalytical Chemistry* 398 (1): 219–26. https://doi.org/10.1007/s00216-010-3788-3.

[51] Lamprecht, Michael R., David M. Sabatini, and Anne E. Carpenter. 2007. "CellProfiler: Free, Versatile Software for Automated Biological Image Analysis." *BioTechniques* 42 (1): 71–75. https://doi.org/10.2144/000112257.

[52] Valm, Alex M., Sarah Cohen, Wesley R. Legant, Justin Melunis, Uri Hershberg, Eric Wait, Andrew R. Cohen, Michael W. Davidson, Eric Betzig, and Jennifer Lippincott-Schwartz. 2017. "Applying Systems-Level Spectral Imaging and Analysis to Reveal the Organelle Interactome." *Nature* 546 (7656): 162–67. https://doi.org/10.1038/nature22369.

[53] Ouyang, Wei, Andrey Aristov, Mickaël Lelek, Xian Hao, and Christophe Zimmer. 2018. "Deep Learning Massively Accelerates Super-Resolution Localization Microscopy." *Nature Biotechnology* 36 (5): 460–68.

[54] Ounkomol, Chawin, Sharmishtaa Seshamani, Mary M. Maleckar, Forrest Collman, and Gregory R. Johnson. 2018. "Label-Free Prediction of Three-Dimensional Fluorescence Images from Transmitted-Light Microscopy." *Nature Methods* 15 (11): 917–20. https://doi.org/10.1038/s41592-018-0111-2.

[55] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation." In *Lecture Notes in Computer Science*, 234–41. Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28.

[56] Schaduangrat, Nalini, Samuel Lampa, Saw Simeon, Matthew Paul Gleeson, Ola Spjuth, and Chanin Nantasenamat. 2020. "Towards Reproducible Computational Drug Discovery." *Journal of Cheminformatics* 12 (1). https://doi.org/10.1186/s13321-020-0408-x.

[57] Prinz, William A, Alexandre Toulmay, and Tamas Balla. 2020. "The Functional Universe of Membrane Contact Sites." *Nature Reviews Molecular Cell Biology* 21 (1): 7–24.

[58] Way, GP, SN Chandrasekaran, M Bornholdt, SJ Fleming, H Tsang, A Adeboye, B Cimini, et al. 2021. "Pycytominer: Data Processing Functions for Profiling Perturbations."

[59] Subramanian, Aravind, Rajiv Narayan, Steven M. Corsello, David D. Peck, Ted E. Natoli, Xiaodong Lu, Joshua Gould, et al. 2017. "A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles." *Cell* 171 (6): 1437–1452.e17. https://doi.org/10.1016/j.cell.2017.10.049.

תקציר

סינון תרכובות מבוסס תמונה בתפוקה גבוהה הוא כלי רב עוצמה לזיהוי הבדלים פנוטיפיים באוכלוסיות התאים. בעבודת הגמר שלי, אני מציע להשתמש בגילוי אנומליות כמדד חדש לזיהוי רגיש של שינויים פנוטיפיים בעת סינון התרכובות. גילוי אנומליות, שמטרתו לזהות תצפיות חריגות החורגות מהתפלגות הבסיס, מתאים במיוחד לסינון תרכובות מכיוון שחזרות רבות של ניסויי בקרה מגדירות את התפלגות הבסיס, ומעט חזרות של טיפולים רבים הן חריגות פוטנציאליות מהתפלגות בסיס זו. באופן ספציפי, אימנתי מודלים, מסוג מקודד אוטומטי, לשחזר את הפרופילים הפנוטיפיים של תאי הבקרה והשתמשתי בשגיאת השחזור כמדידה לזיהוי חריגות. יישמתי את רעיון זה בשלושה מצבים: אנומליה מבוססת תכונה של אברונים, אנומליה בקשרי האברונים מבוססת תכונה ואנומליה בקשרי האברונים מבוססת תמונה. התוצאות שלי מצביעות על כך שאנומליות מבוססות תכונה הן רגישות ומשלימות יותר, אך מעט פחות ניתנות לשחזור, בהשוואה לגישה הסטטיסטית הנפוצה. התוצאות הראשוניות שלי עם אנומליה מבוססת תמונה מצביעות על הפוטנציאל לחשוף תלות מרחבית בין האברונים. יש צורך בבדיקות נוספות כדי לאמת תוצאות אלו.

אוניברסיטת בן-גוריון בנגב
הפקולטה למדעי ההנדסה
המחלקה להנדסת מערכות מידע ותוכנה

גילוי אנומליות לצורך זיהוי רגיש של תרכובות מטרה בתהליך סינון תרכובות מבוסס תמונה בתפוקה גבוהה

חיבור זה מהווה חלק מהדרישות לקבלת תואר מוסמך במדעים (M.Sc.)

נאור קולט

בהנחיית דר' **אסף זריצקי**

חתימת המחבר: _____ תאריך: _____

אישור מנחה: _____ תאריך: _____

אישור יו"ר ועדת תואר שני מחלקתית: _____ תאריך: _____

פברואר 2023

אוניברסיטת בן-גוריון בנגב
הפקולטה למדעי ההנדסה
המחלקה להנדסת מערכות מידע ותוכנה

גילוי אנומליות לצורך זיהוי רגיש של תרכובות מטרה בתהליך סינון תרכובות מבוסס תמונה בתפוקה גבוהה

חיבור זה מהווה חלק מהדרישות לקבלת תואר מוסמך במדעים (M.Sc.)

נאור קולט

בהנחיית דר' אסף זריצקי

פברואר 2023