



LUND  
UNIVERSITET

# A Deep Learning Pipeline for Genome-Wide Imaging Screen Uncovering Cell Death Regulators

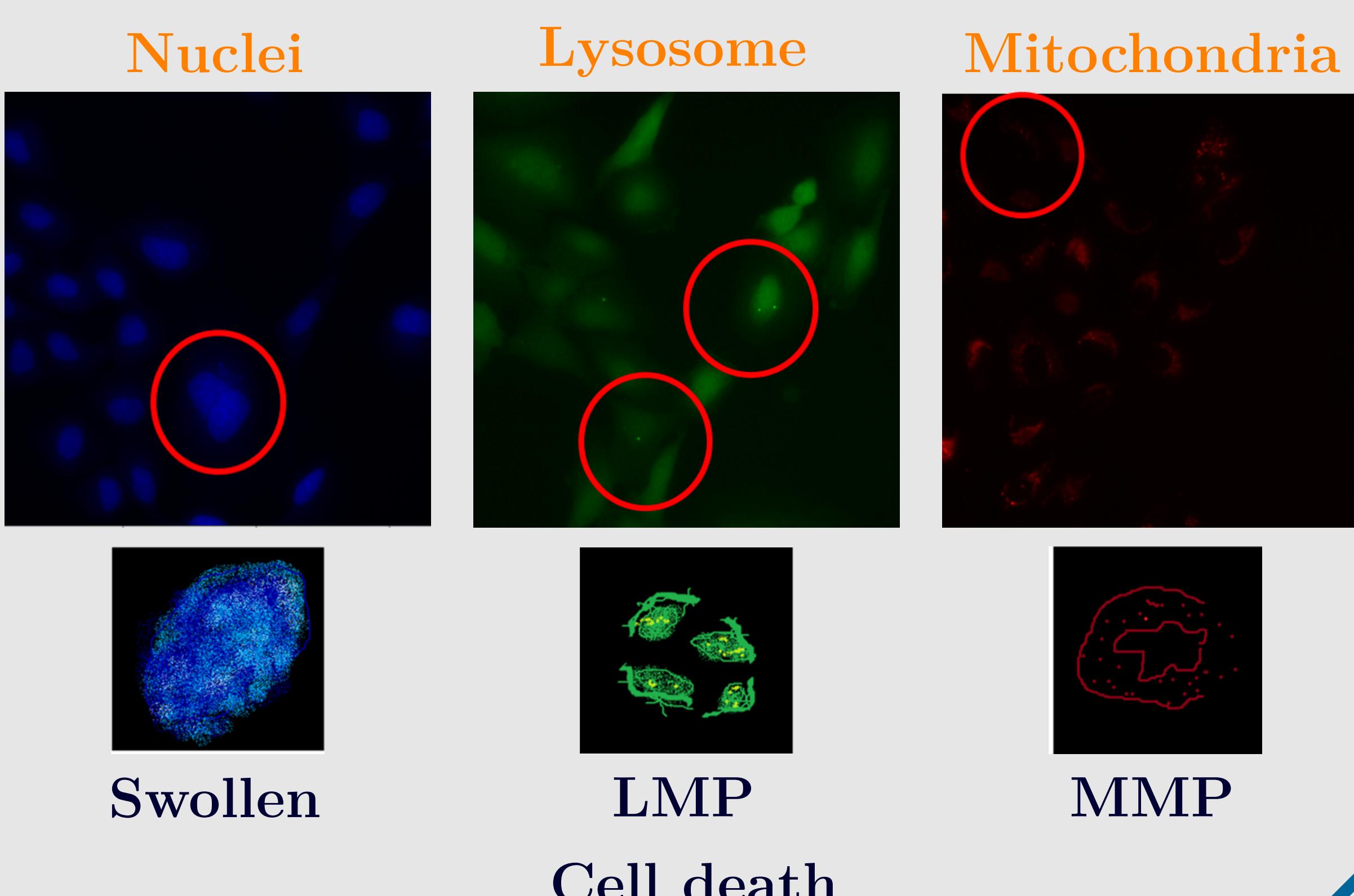
Salma Kazemi Rashed, Klara Esbo, Mariam Miari, Malou Arvidsson, Liu J, Cowley K, Simpson K, Johnstone R, Sonja Aits

Cell Death, Lysosomes and AI group, Lund University, Email: [Salma.kazemi\\_rashed@med.lu.se](mailto:Salma.kazemi_rashed@med.lu.se)

## Introduction

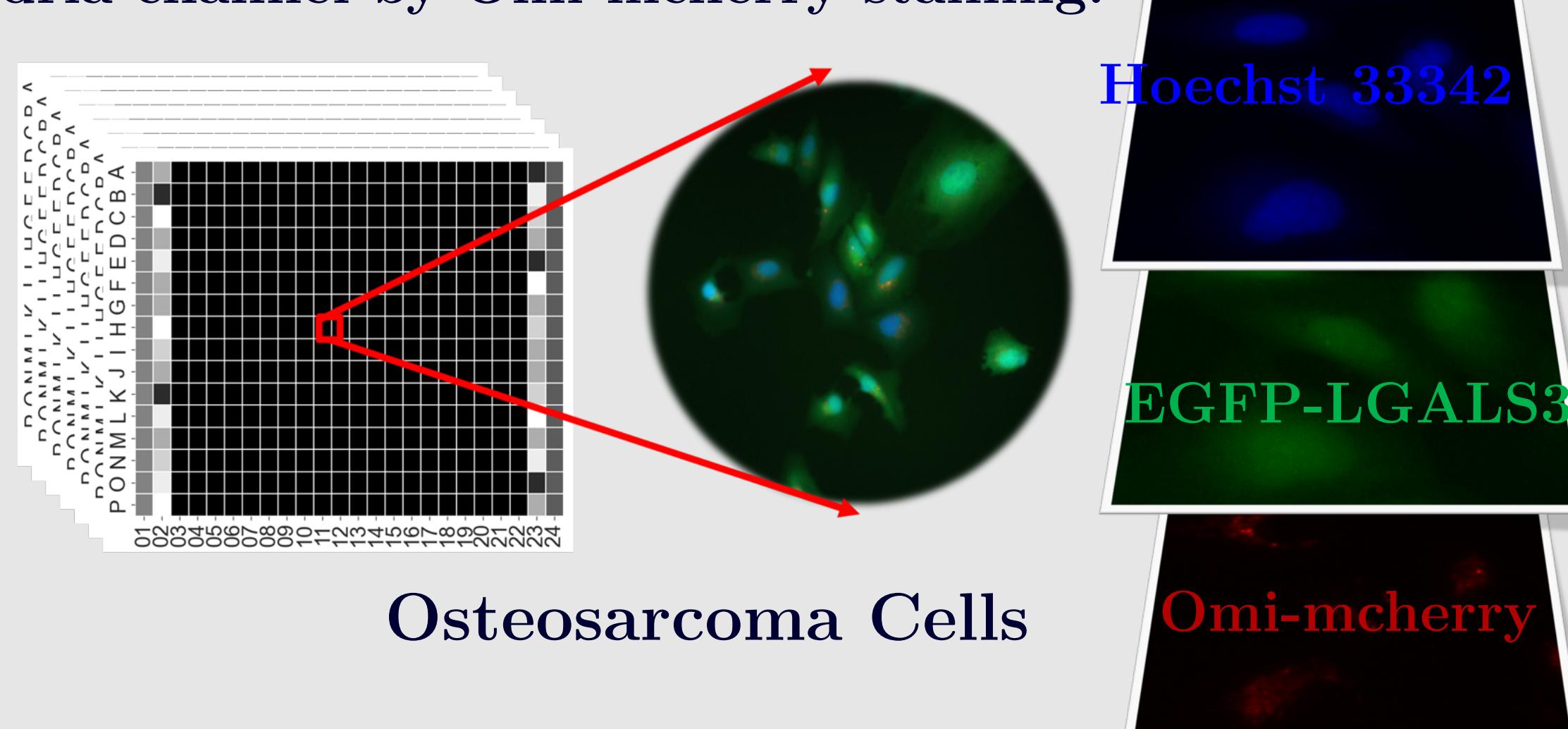
Cell death is a crucial biological process that creates space for new cells and prevents the spread of harmful mutated cells. It is highly regulated, but mutations in regulatory genes can lead to excessive or insufficient cell death, contributing to various diseases. There are different types of cell death, including a regulated process where the cell fragments and an uncontrolled one triggered by trauma, causing the cell to swell and burst.

Cell death process can alter the nucleus, cell body, and organelles like lysosomes, which contain harmful enzymes. Lysosomal membrane permeabilization (LMP) can trigger cell death and affect other organelles like mitochondria (MMP).



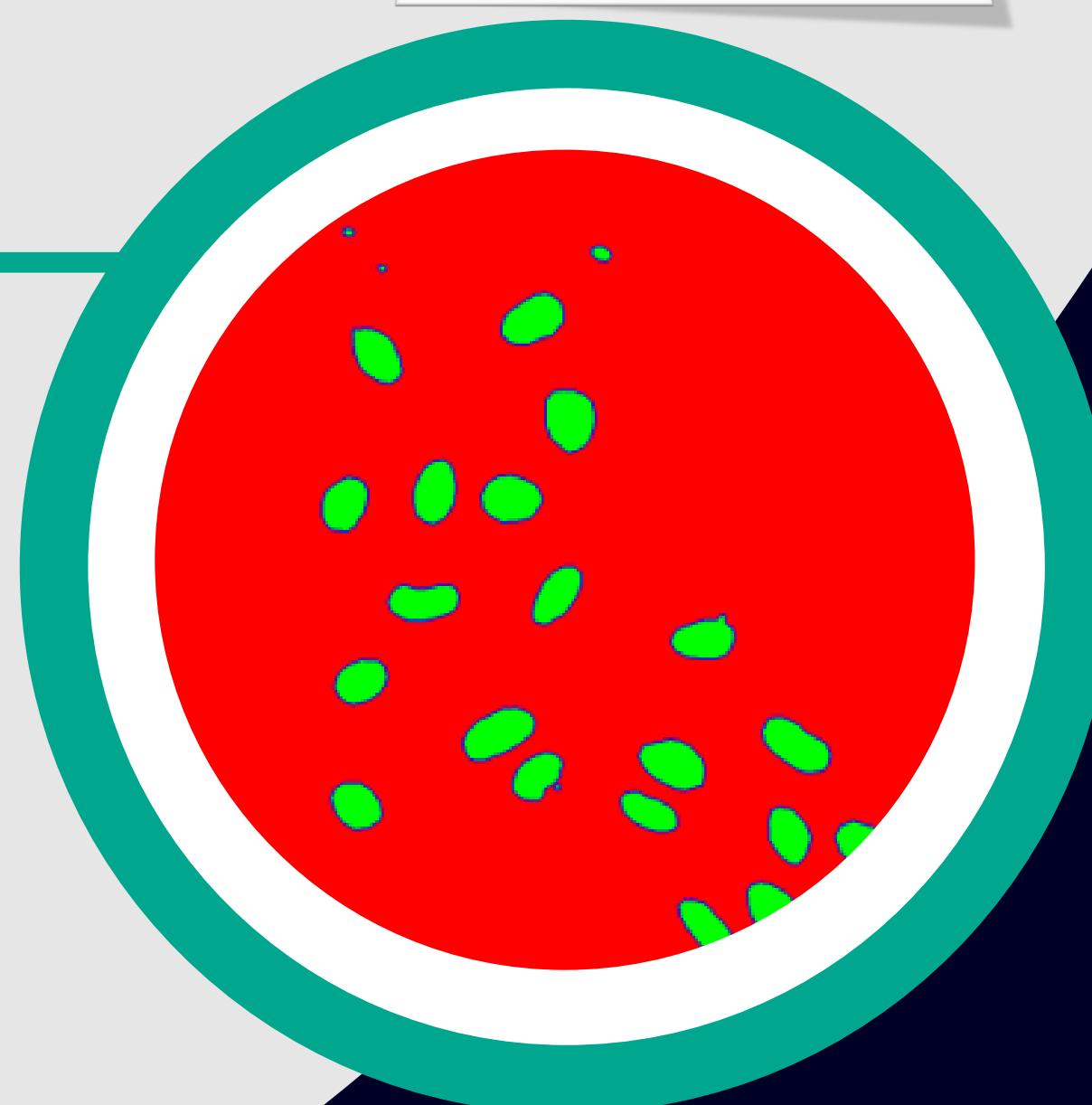
## Dataset

The dataset comprises 4,276,224 image frames ( $1104 \times 1104 \times 3$ ), collected from a gene knockdown study on U2OS osteosarcoma cells, conducted by Dr. Sonja Aits. Cells were transfected with an siRNA library targeting 18,170 protein-coding genes, and phenotypic changes were visualized across 213 plates using three fluorescent channels to label key cell components. The nucleus was labelled by Hoechst 33342, cells by EGFP-LGALS3, and mitochondria channel by Omi-mcherry staining.



## Annotation

We manually annotated a small subset of our dataset, consisting of 50 images: 30 for training, 10 for validation, and 10 for testing the model before applying it to the full dataset [1][2].

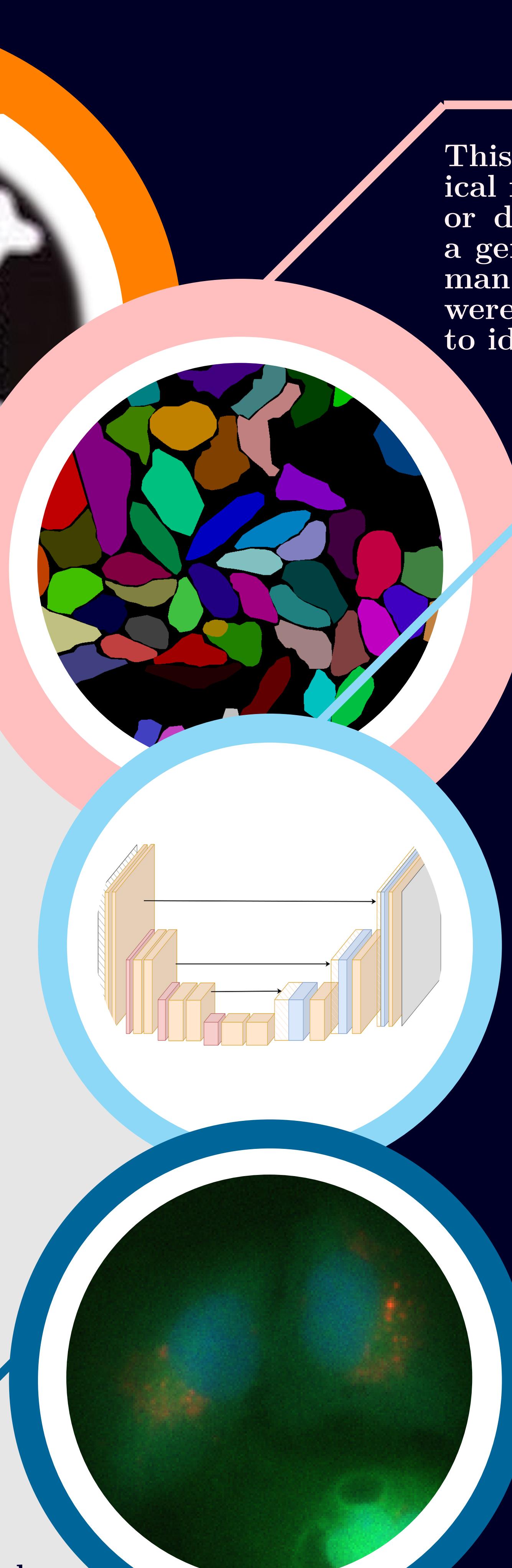


This study focuses on cell nuclei, extracting morphological features of all nuclear objects in order to find dying or dead cells, such as shape, area, and count. Using a genome-wide knockdown experiment, where each human gene was silenced and corresponding cell images were captured, these features are linked to the genome to identify cell death pathways.

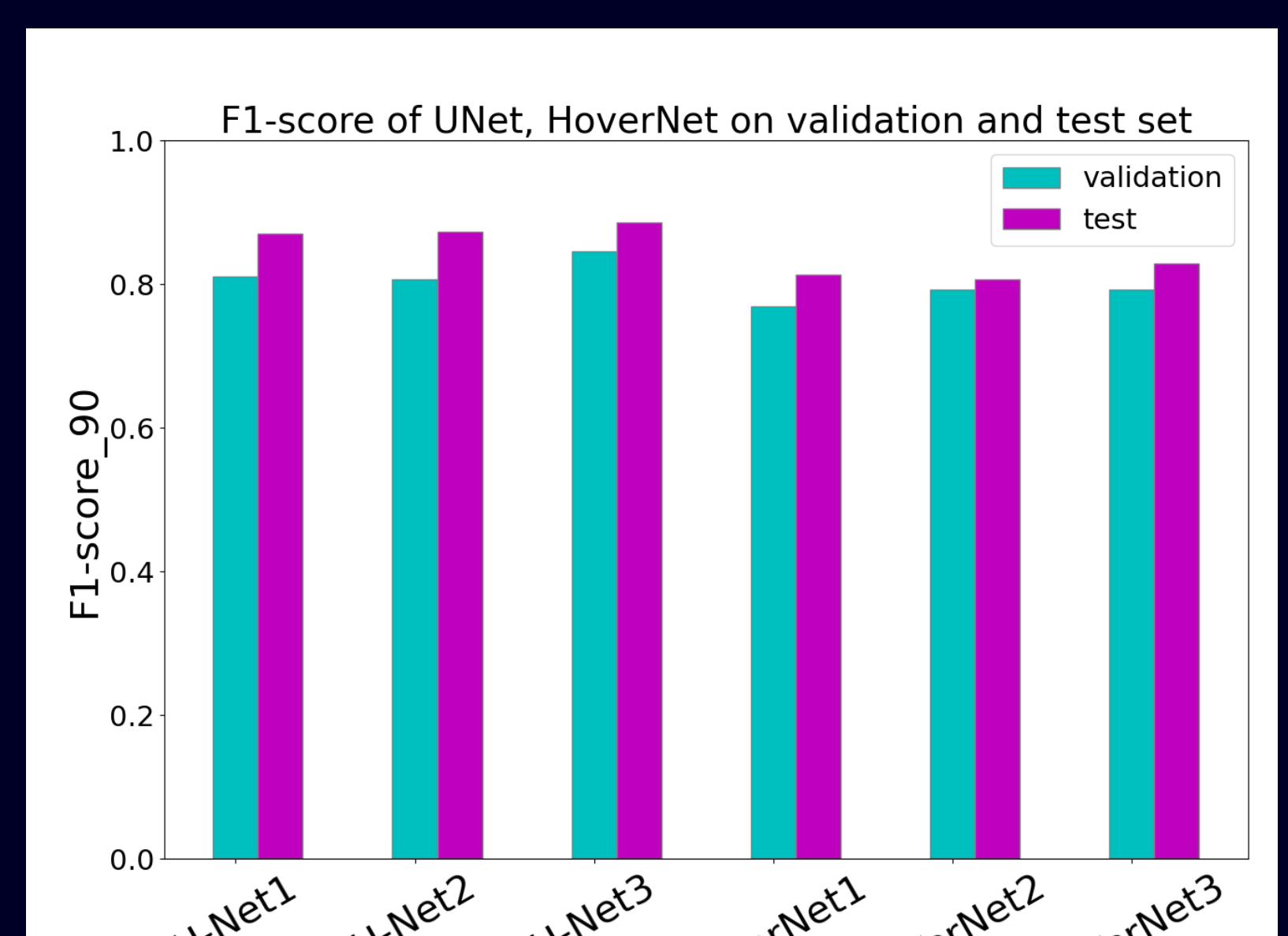
## Objective

The nearly 9 TB dataset of over 213 plates lacked annotations, prompting a stepwise solution. We manually annotated 50 images from nuclei[1], and cell[2] channels and utilized public datasets, such as the Broad Bioimage Benchmark Collection (BBBC), to train U-Net models from scratch, and experimented with advanced models like Hover-Net leveraging transfer learning. The best model achieved 89% F1-score and was applied to the entire dataset to segment nuclei and extract size and count features.

## Methodology

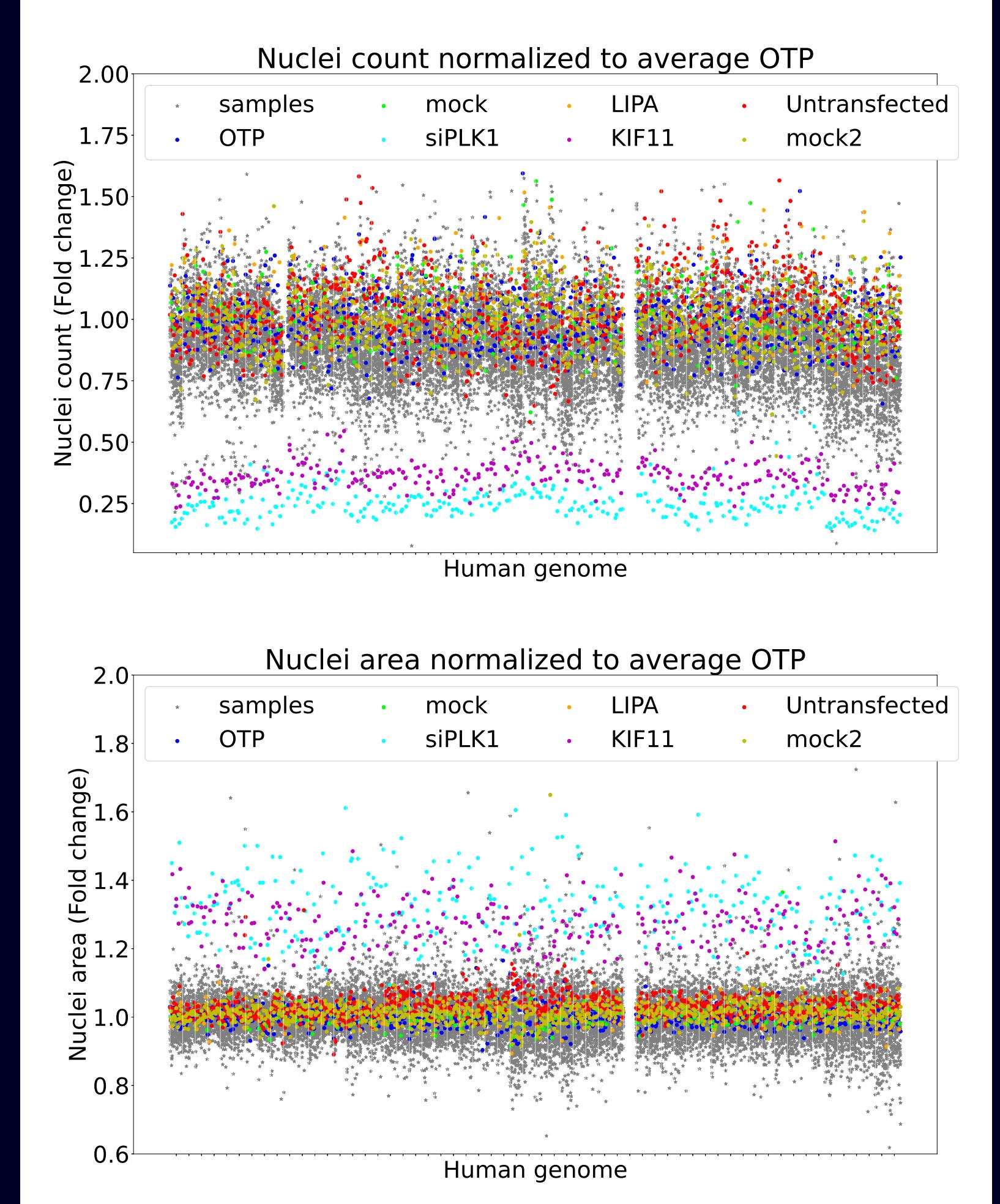


We evaluated using F1-score and Jaccard Index, with IoU ( $\text{IoU}(A, B) = \frac{|A \cap B|}{|A \cup B|}$ ) measuring overlap between predictions and annotations. F1-score, an object-based metric calculated as the geometric mean of recall and precision, was assessed at IoU thresholds ranging from 50% to 90%, with the final evaluation at 90%.



## Results

The best model was applied to the full dataset, and the nuclei count and area for each well were plotted.



## Conclusion

Identifying wells with significant nuclei changes allowed us to link them to genes, resulting in a list of candidates for further study in novel cell death pathways.

Gene	Count	Area
FBXO5	< 0.5	> 1.7
CCNA2	< 0.5	> 1.7
UBB	< 0.1	
POLR2A	< 0.1	

[1] Malou Arvidsson, Salma Kazemi Rashed, Sonja Aits, An annotated high-content fluorescence microscopy dataset with Hoechst 33342-stained nuclei and manually labelled outlines, Data in Brief, Volume 46, 2023, 108769, ISSN 2352-3409, <https://doi.org/10.1016/j.dib.2022.108769>.

[2] Salma Kazemi Rashed, Malou Arvidsson, Rafsan Ahmed, Sonja Aits, An annotated high-content fluorescence microscopy dataset with EGFP-Galectin-3-stained cells and manually labelled outlines, Data in Brief, 2024, 111148, ISSN 2352-3409, <https://doi.org/10.1016/j.dib.2024.111148>.