

H U M B O L D T - U N I V E R S I T Ä T Z U B E R L I N



LEBENSWISSENSCHAFTLICHE FAKULTÄT

INSTITUT FÜR BIOLOGIE

Bachelorarbeit

zum Erwerb des akademischen Grades Bachelor of Science

Image-Based Analysis of Knockdown-Dependent Shape Transitions of Migrating MDA-MB-231 Cancer Cells in a Wound Healing Assay

vorgelegt von

Name, Vorname: Sperling, Olyssa
E-Mail-Adresse: olyssa.sperling@hu-berlin.de
Erstprüferin: Prof. Dr. Dr. h.c. Edda Klipp
Zweitprüfer: Prof. Dr. Uwe Ohler
Abgabedatum: 04. April 2024

Abstract

Cancer metastasis describes the spread of cancer cells from the primary tumor to distant sites in the body. It is accomplished by various cellular and molecular mechanisms that are currently researched in order to understand them and the resulting migratory behaviour of cancer cells. A mechanism also present in migrating cancer cells is the epithelial to mesenchymal transition (EMT). It includes cell shape transitions and enables cells to invade tissue, resist stress and disseminate. In this thesis, the shape transitions of migrating MDA-MB-231 breast cancer cells are researched in a wound healing assay. The assay contains cells under 4 different control siRNA knockdowns, where migratory behaviour is known, as well as cells under 352 other siRNA knockdowns, where migratory behaviour is analysed. Therefore, a high-throughput experiment was conducted generating a time-series image data set with three experimental replicates. In order to gain information on the impact of various siRNA knockdowns on the migratory behavior of breast cancer cells, an image analysis workflow, utilizing several deep learning image analysis tools, is implemented. An algorithm in the workflow detects the leading edges of cells closing the wound, enabling the calculation of wound area closing over time, the leading edge velocities and positional features of single cells. Furthermore, single-cell trajectories are computed and analysed and single-cell shape features are extracted, used for directionality computations and a clustering assay.

Zusammenfassung

Die Metastasenbildung bei Krebs beschreibt die Ausbreitung von Krebszellen aus dem Primärtumor an entfernte Stellen im Körper. Sie erfolgt durch verschiedene zelluläre und molekulare Mechanismen, die derzeit erforscht werden, um sie und das daraus resultierende Migrationsverhalten von Krebszellen zu verstehen. Ein Mechanismus, der auch bei migrierenden Krebszellen zum Tragen kommt, ist der epitheliale zu mesenchymale Übergang (EMT). Er umfasst die Umwandlung der Zellform und ermöglicht es den Zellen, in das Gewebe einzudringen, Stress zu widerstehen und sich auszubreiten. In dieser Arbeit werden die Formveränderungen von migrierenden MDA-MB-231-Brustkrebszellen in einem Wundheilung-Assay untersucht. Der Assay enthält Zellen unter 4 verschiedenen Kontroll-siRNA-Knockdowns, bei denen das Migrationsverhalten bekannt ist, sowie Zellen unter 352 weiteren siRNA-Knockdowns, bei denen das Migrationsverhalten analysiert wird. Daher wurde ein Hochdurchsatz-Experiment durchgeführt, bei dem ein Zeitserien-Bilddatensatz mit drei experimentellen Wiederholungen generiert wurde. Um Informationen über die Auswirkungen verschiedener siRNA-Knockdowns auf das Migrationsverhalten von Brustkrebszellen zu erhalten, wurde ein Bildanalyse-Workflow implementiert, der mehrere Deep-Learning-Bildanalysetools verwendet. Ein Algorithmus im Workflow erkennt die Vorderkanten von Zellen, die die Wunde schließen, und ermöglicht die Berechnung der sich im Laufe der Zeit schließenden Wundfläche, der Vorderkantengeschwindigkeiten und der Positionsmerkmale einzelner Zellen. Darüber hinaus werden die Trajektorien einzelner Zellen berechnet und analysiert und Formmerkmale einzelner Zellen extrahiert, die für Richtungsbestimmungs-Berechnungen und für einen Clustering-Ansatz verwendet werden.

Abbreviations

cyto	cytoplasmic
D	nuclei displacement polarity angle
ECM	extracellular matrix
EMT	epithelial to mesenchymal transition
ERK	extracellular signal-regulated kinase
HCS	high content screening
IFN	interferon
MIST	Microscopy Image Stitching Tool
MSD	mean squared displacement
MTA1	Metastasis-associated protein 1
MYH9	myosin heavy chain 9
NF- κ B	nuclear factor κ -light-chain-enhancer of activated B
NTC1	non-targeting control 1
nuc	nuclear
P	persistency
PCA	principal Component Analysis
PLK1	polo-like kinase 1
POU5F1	POU class 5 homeobox 1
RNAi	RNA interference
siRNA	small interfering RNA
t	time
UMAP	uniform manifold approximation and projection
uPA	urokinase-type plasminogen activator
VAE	variational autoencoders

List of Figures

1	Wound healing assay of knockdown NTC1 of cell line MDA-MB-231. (a) At t=0 h a mechanical scratch was introduced into the monolayer. (b) By t=16 h the migrating cells almost closed the void.	10
2	Image stitching. Three sub-tiles (a) are combined into one image (b) using the stitching tool 'MIST'.	15
3	Composition of images in dataset. The scratch (magenta) has been done inclined in order to gain as much information about cells wandering into the scratch area as possible.	16
4	Technical functionality of the scratch detection algorithm. (a) Input image. (b) Blurred. (c) Segmentation of the foreground. (d) Labeling of disconnected components. (e) Selection of two biggest regions. (f) Extraction of scratch border coordinates.	17
5	Polygons created for extraction of positional features of cells. The cell centroids categorized by a polygon are colored accordingly.	18
6	Estimation of velocity of leading edge cells. The distance of points on linear functions fitted to the (a) leading edge was calculated for every hour, but for displaying reasons depicted here for the distance between leading edges of (b) t=0 h and (c) t=7 h.	19
7	Estimation of perpendicular bisector of scratch. The left (cyan) and right (orange) leading edges of the tissue were detected using the scratch detection algorithm. The vector of the estimated perpendicular bisector of the scratch $v_{scratch}$ is colored in magenta.	20
8	Estimation of curvature of cell border along its length. (a) represents the segmented cell with its predicted cell border generated by Cellpose 2.0. In (b) pink indicates convex regions while cyan indicates concave ones along the cell border length. (c) Plot of intensity of curvature along cell border length. Values greater zero infer convex regions and values smaller than zero correspond to concave regions. The red dots mark the local maxima of the curvature, also referred to as the feature 'protrusion' (feature nr. 22). (d) displays the calculation of the feature 'curvature' (feature nr. 21).	23
9	Cell segmentation after using Cellpose 2.0. In (b) and (c) the detected cell outlines colored in yellow can be inspected in more depth.	25
10	Scratch detection algorithm applied to NTC1 after (a) 0 h, (b) 8 h and (c) 16 h. Marked in magenta is the detected scratch contour.	26
11	Plate heatmap of scratch area ((a)-(c)) and its standard deviation ((d)-(f)) over time for t=0 h, t=8 h and t=16 h. Each square represents the mean of three replicates of a well in the multi-well plate. White squares represent wells where the scratch contours could not be detected, hence scratch area could not be determined.	27

12	Calculated mean scratch area of control knockdowns (n=8) and of all knockdowns (n=352) ((a)-(c)) and their standard deviations ((e)-(g)) over 16 hours. In (d) the mean and in (h) the standard deviation has been taken over all time-series area data over all experimental replicates for each knockdown (n=32).....	28
13	Estimated mean leading edge velocities of the control knockdowns and of all knockdowns for the three experimental replicates. (a) replicate 1. (b) replicate 2. (c) replicate 3.....	29
14	Cell centroids and nuclei centroids at t=16 h. Centroids of cells (cytoplasm) are marked in green and centroids of nuclei are marked in red. In (a) all detected cell centers (n=406) and their corresponding nuclei (n=125) are marked. The marked cells in (b) are labeled to be located the leading edge cells.....	30
15	Nuclei displacement at t=16 h. The centroids of the cells (cytoplasm) are marked with green 'x', while the centroids of the nuclei are marked with red 'x'. In (a) the white arrows symbolise the calculated displacement of the nuclei with respect to the cytoplasm. In (b) and (c) the detected nuclei displacements can be observed in more detail. For better visibility, in (b) and (c) the displacement arrows are colored in black.	31
16	Nuclei displacement polarity angle over time of leading edge cells compared to bulk cells. (a)-(c) Variability of the polarity angles for knockdown MYH9 for the different replicates. (d) Angle progress over time for all knockdowns in all experimental replicates. ((e)-(h)) Angle evolvement over 16 h for means of control knockdowns.	31
17	Tracked cells for a time difference of one hour. Cells colored in green (blue borders) are of t=0 h and cells colored in blue (green borders) are of t=1 h. The white arrows start at the centroid of the green cell at t=0 h and end with their tip at the centroid of the blue cell at t=1 h.	32
18	Whole cell tracks of one experiment plotted onto the cells at t=0 h. The tracking algorithm was able to follow 136 cells from t=0 h until t=16 h, while 762 cells were present at t=0 h, meaning that approximately 18 % of the cells were tracked for the whole time for this well. In (b) and (c) the cell trajectories can be observed in more depth.	33
19	Mean squared displacement of whole tracks of leading edge cells of control knockdowns. The graph is plotted with a confidence interval width of 95 %. ((a)-(c)) Plots for each replicate. (d) All replicates.	34
20	Persistency of whole tracks of leading edge cells of control knockdowns. All graphs are plotted with a confidence interval width of 95 %. ((a)-(c)) Plots for each replicate. (d) Legend. (e) All replicates. .	34
21	Cell shape feature distributions for t=16 h of all three replicates plotted on top of each other. Replicate 1 is colored in blue, replicate 2 in orange and replicate 3 in green. All of the distributions are overlapping very well. Displayed is a sample fraction of 0.01 % of the filtered cells for visibility. .	35

22	UMAPs with Leiden clustering of all experimental replicates at t=16 h in ((a)-(c)). In (d) using a sample fraction of 30 % of all concatenated replicate data for t=16 h an UMAP was calculated.	35
23	Script of Fiji macro for image stitching with MIST and drift-correction with NanoJ-Core.	48
24	Correcting cell segmentations generated by model 'cyto' from Cellpose 2.0 and thereby creating a data set for training on some random images from own data in order to create a segmentation model adapted to own data.	49
25	Case of failing of the automatic scratch detection due to not enough robustness of the algorithm when employed on a scratch with cells contained inside the scratch area at t=0 h.	50
26	Comparison of positional feature extraction methods. (a) Results from the in this thesis used scratch detection algorithm are showed. (b) Distance fields colored by distance to the scratch. The distance field was computed with the Python package skfmm and depicts another approach to categorize leading edge cells and bulk cells, especially for untidy scratches.	51
27	Nuclei displacement polarity angle of leading edge cells towards perpendicular bisector of the scratch, displayed for the control knockdowns MYH9, PLK1, POU5F1 and NTC1 and for all knockdowns over time. (a)-(c) depict the evolution of the nuclei displacement polarity angle for the three experimental replicates, (d) shows the legend for the different knockdowns and (e) displays the angle evolution for all experimental replicates combined. All graphs are plotted with a confidence interval width of 95 %.	52
28	Explanation of measures for calculation of cell shape features.	53
29	Cell shape feature distributions of all experimental replicates. Displayed is a sample fraction of 0.01 % of the filtered cells for visibility. All of the y-axes are normed throughout all replicates. . .	54
30	Cell shape feature distributions of replicate 1 of bulk cells (blue) compared to leading edge cells (orange). Displayed is a sample fraction of 0.01 % of the filtered cells for visibility.	55
31	Heatmaps of feature impact on cluster. Significant features for a cluster are marked by the value 1 (black), while insignificant clusters are marked by the value 0 (light blue).	55

List of Tables

1	Extracted cell shape features.	22
2	Filtering conditions of cell shape features.	24

Contents

1	Introduction	7
2	Methods	15
2.1	Image pre-processing	15
2.2	Cell segmentation	15
2.3	Cell migration analysis	16
2.3.1	The scratch detection algorithm	16
2.3.2	Calculation of the closing wound area	18
2.3.3	Estimation of velocity of the leading edge	18
2.3.4	Calculation of nuclei displacement polarity	19
2.3.5	Cell trajectory analysis	20
2.4	Single-cell shape feature analysis	21
2.4.1	Definitions of single-cell shape features	21
2.4.2	Filtering of cell shape features	23
2.4.3	Dimensionality reduction and leiden clustering of cell shape features	24
3	Results	25
3.1	Stitching and drift-correction	25
3.2	Cell segmentation	25
3.3	Cell migration analysis	26
3.3.1	Area of closing wound	26
3.3.2	Estimation of velocity of leading edges	29
3.3.3	Nuclei displacement polarity angle	29
3.3.4	Cell trajectory analysis	32
3.4	Single-cell feature analysis	35
3.4.1	Feature distributions	35
3.4.2	Clustering	35
4	Discussion	37
5	References	41
6	Acknowledgments	47
7	Appendices	48
7.1	Fiji macro for image stitching and drift-correction	48
7.2	Own cell segmentation trained onto model 'cyto' from Cellpose 2.0	49
7.3	Scratch detection algorithm fail case	50
7.4	Distance fields	51
7.5	Nuclei displacement polarity angle	52
7.6	Explanation of single-cell shape feature measures	53
7.7	Single-cell shape features	54

1 Introduction

Cancer metastasis involves the spread of cancer cells from the primary tumor to distant sites in the body [1]. Metastasis significantly contributes to cancer morbidity and mortality, with up to 90 % of cancer-related deaths attributed to this phenomenon [2]. Different types of cancer exhibit varying patterns of metastasis, with breast cancer commonly metastasizing to sites such as bone, lung, liver, and brain [3]. For instance, bone metastasis is a frequent complication in breast cancer, especially in patients with hormone receptor-positive cancers, where long-term survival is possible even with bone involvement [4]. Brain metastasis, on the other hand, is more common in certain cancers like malignant melanoma, lung, and breast cancer [5]. Despite being less common, stomach cancer can also metastasize to the brain, highlighting the diverse metastatic potential across cancer types [6].

Mechanisms of metastasis

Metastasis is a complex process involving various cellular and molecular mechanisms. Understanding the mechanisms underlying metastasis is crucial for developing targeted therapies and improving patient outcomes. For example, the expression of metastasis-associated proteins like Metastasis-associated protein 1 (MTA1) and E-cadherin can serve as indicators of cancer progression and metastasis in various cancer types [7]. Moreover, the identification of specific biomarkers associated with metastasis can aid in predicting and monitoring metastatic spread across different cancer types [8]. Research has also focused on illuminating the molecular pathways involved in metastasis, such as the extracellular signal-regulated kinase (ERK) signaling pathway, which has implications in proliferation, survival, differentiation, invasion and motility of cancer cells [9]. Cancer metastasis remains a significant challenge in cancer treatment, underscoring the importance of continued research to understand the complexities of this process. The process of cancer cell dissemination from primary tumors and their establishment in distant tissues involves a series of steps known as the invasion-metastasis cascade [10]. This cascade includes the local invasion of primary tumor cells into nearby tissues, their entry into the circulatory system (intravasation), survival during transport through the bloodstream, arrest and exit from blood vessels into distant tissues (extravasation), formation of small tumor colonies (micrometastases), and subsequent growth of these colonies into clinically detectable metastatic lesions, a process called colonization [11]. These processes in cancer cells are likely caused by genetic and epigenetic alterations occurring within both the tumor cell and its surrounding microenvironment [12]. While research has clarified the complicate mechanisms driving primary tumor formation, our understanding of the biological mechanisms underlying metastatic disease remains limited. The invasion-metastasis sequence may be a result of chromosomal instability arising from persistent errors in chromosome segregation during mitosis. Mistakes in chromosome segregation can lead to the rupture of micronuclei and the release of genomic DNA into the cytosol, triggering activation of cytosolic DNA-sensing

pathways such as cyclic GMP-AMP synthase–stimulator of interferon (IFN) genes, which in turn activate downstream nuclear factor κ -light-chain-enhancer of activated B (NF- κ B) signaling [13]. By investigating the unique patterns and mechanisms of metastasis in different cancer types, we can advance our comprehension of how cancer spreads and identify novel therapeutic targets to fight metastatic disease effectively.

Epithelial to mesenchymal transition

The epithelial to mesenchymal transition (EMT) is a mechanism present in wound healing. Also, in metastasis it is an important process by which transformed cells gain the ability to invade, resist stress and disseminate [14]. Epithelial cells are tightly bound in the extracellular matrix (ECM), while mesenchymal cells are plastic and have the characteristics for invasion and metastatic spread, such as enhanced motility, invasiveness, and capacity to break down ECM components [14]. Carcinoma cells, cancerous cells originated from epithelial cells, spread through two mechanisms: single-cell dissemination facilitated by EMT or collective dissemination of tumor clusters. Recent findings indicate that leader cells within tumor clusters may also undergo specific phenotypic alterations linked to EMT [12]. Cell polarity, marked by the uneven distribution of cell organelles, can be essential for supporting cell migration and research has focused on understanding its importance. To establish a polarity axis with a clear direction, referred to as 'front' and 'back', cells must undergo spatial reorganization of the plasma membrane, cytoskeleton, cell-cell junctions, or organelles. During these processes, cells must adapt and respond to numerous and sometimes conflicting signals from their environment [15]. Due to the mechanisms underlying metastasis, cells that are located at the leading edge of tumor tissue are exposed to flow [16]. The flow can be induced by several cues including growth factors or shear stress. The cells at the leading edge are actin-enriched and polarised towards flow and elongate, while cells under a static condition like epithelial cells have a random polarity. This polarity can be measured by an asymmetric localisation of cell organelles [17]. The nucleus, the most prominent organelle in eukaryotic cells, experiences continuous exposure to both internal and external mechanical forces, leading to dynamic alterations in its shape and position [15]. This aspect highlights the important role of cell polarity in coordinating cell migration, especially during events such as metastasis [15].

Cell shape transitions in metastasis

The ability of cancer cells to migrate and invade surrounding tissues is a main characteristic of metastasis [18]. On its way to the target site, the invading tumor cell interacts with its microenvironment and changes its motility and plasticity. [19] Interactions with the tumor microenvironment enable infiltrating cancer cells to overcome stromal obstacles, establish residence, and proliferate. For instance, breast cancer cells exhibit diverse and dynamic shapes that are closely linked to their functional properties, including migration, invasion and proliferation rates [20]. Aberrations and deregulation of morphological and migratory cell features occur frequently

with pathological conditions like cancer [18]. Every cell possesses a distinct shape, a specific arrangement of organelles and a certain orientation in space. Moreover, the cell also forms junctional contacts with neighboring cells, each exhibiting unique morphological characteristics. When cells experience the effect of toxic substances or are dividing they are characterized by a rounding up [21]. During the metastasis of cancer cells, there are intricate and highly coordinated changes in morphology of cells. Cancer cells show various migration modes when invading 3D tissue, all characterized by distinct morphological and morphodynamic signatures, including filopodial, lamellipodial, lobopodial, blebbing and actin-enriched leading edge mode [22]. The ability of cancer cells to transition between these migration modes is crucial tumor prognosis. Breast cancer cells can display a variety of cell shapes, such as round, elongated, spindle-like and irregular shapes, and these shapes can undergo dynamic changes throughout metastasis [23]. Hence cell shape and morphological transitions are linked to cell migration and signaling networks [19]. Due to cellular structure undergoing dynamic remodeling to promote processes such as proliferation, migration and metastasis, the cytoskeleton is often altered in cancer, and therefore also cell shape is transformed significantly in comparison to healthy cells [24].

Wound healing assay

Studying cell migration in controlled experimental settings provides valuable insights into the dynamic behaviors of cancer cells during the metastatic process as well as in understanding cancer cell phenotype [25]. Visualization of growth, individual cell movements, cell-cell interactions, and the overall progression of migration in response to various stimuli or treatments can be made visible through wound healing assays [16]. Image-based data of cells moving while reacting to a perturbation captures a snapshot of the cell's condition. To observe and quantify morphological and comparable features in time and space it is essential to understand the relation of signaling networks and resulting migratory behavior [19]. Wound healing, the movement of cells into a wound to accomplish gap closure, is a way to perturb a system to show its reaction on a change in its microenvironment. Other cell migration assays are chemotaxis, where cells migrate in response to a chemical environment, haptotaxis, meaning cell migration happens within a gradient of chemoattractants, and transmigration, which refers to cells moving through a vascular endothelium. When using the technique of a wound healing assay, a gap is created by mechanically scratching a confluent monolayer of cells with a needle in order to remove some of them [26]. After that, cells move inward to fill the void as visible in Fig. (1). Compatible with image based readout the scratch assay offers enhanced accuracy, sensitivity and robustness in learning about the underlying mechanisms governing cancer cell movement in comparison to other migration assays [16].

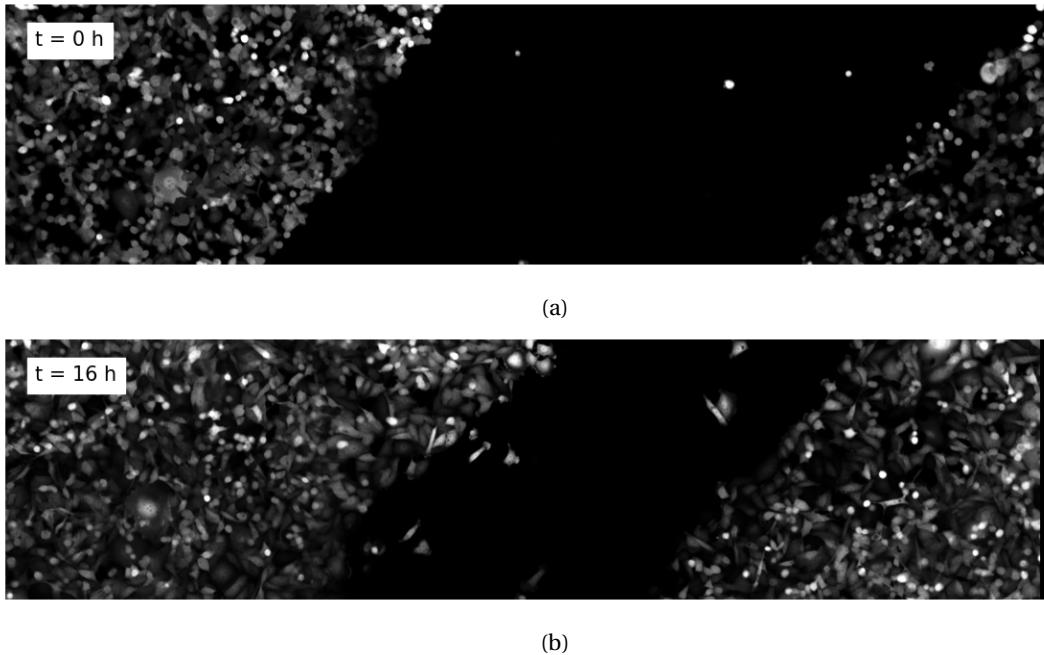


Figure 1: Wound healing assay of knockdown NTC1 of cell line MDA-MB-231. (a) At $t=0$ h a mechanical scratch was introduced into the monolayer. (b) By $t=16$ h the migrating cells almost closed the void.

RNA interference

Recent research has increasingly focused on identifying key regulatory molecules and signaling pathways that influence cancer cell migration, with particular attention given to small interfering RNA (siRNA) knockdown approaches [27]. By analysing the effects of gene knockdowns on cell morphology, migration dynamics, and collective behavior, critical insights into the roles of specific genes and proteins in signaling networks of cancer metastasis can be gained. By that, potential drug targets may be identified [28]. Gene knockdown experiments can be achieved through for example CRISPR/Cas or RNA interference (RNAi) gene editing [29], [30], which allows to selectively inhibit the expression of target genes involved in cell motility [28]. The effectiveness of siRNA in inhibiting the expression of key oncogenes and signaling pathways implicated has been shown in various cancer types. For instance, siRNA-mediated knockdown of genes like SPAG9 in hepatocellular carcinoma [31] and SATB1 in prostate cancer [32] has shown results in reducing cancer cell proliferation and migration. Additionally, siRNA targeting ribonucleotide reductase in melanoma cells has been shown to inhibit cell proliferation, either alone or in combination with chemotherapeutic agents [33]. Furthermore, siRNA knockdown of urokinase-type plasminogen activator (uPA) in breast cancer cells has been linked to suppressed tumor metastasis in preclinical models [34]. A wound closing intensely would suggest an aggressively growing siRNA knockdown. In comparison, cells closing a wound less fast than the wild type cells indicate the detection of a potential drug target.

Live-cell imaging and high content screening microscope

Studying the migratory behavior of cancer cells *in vitro* by live-cell imaging, where images of living cells are captured over a period of time using phase contrast and/or fluorescence microscopy, is a tool for investigating dynamic processes within living systems [35]. Therefore, image data that allows to distinguish cells and their forms are needed to be collected. Furthermore, automation of imaging is essential in order to collect timeseries data. Live-cell imaging provides a detailed understanding of the spatiotemporal dynamics of cancer cell migration, offering information on migration speed, directionality, and coordination among cells through observations inside the cells. For the purpose of making these processes inside cells visible, subcellular resolution of images is needed. In order to collect reliable and comparable data for a big set of treatment conditions, like many different gene knockdowns with multi-well plates, high content screening (HCS) microscopes are employed. A HCS is a specialised imaging system designed for automated and high-throughput analysis of biological samples. These systems feature automated imaging, meaning motorized stages, autofocus capabilities and multiple imaging channels, as well as the ability to handle multi-well plates, allowing simultaneous imaging of numerous condition in a single experiment. This high-throughput capability accelerates data generation and analysis compared to traditional manual microscopy methods [36]. By combining the scratch assay with live-cell imaging, researchers can gain comprehensive insights into the mechanisms underlying cancer cell migration [37].

Challenges of image analysis

Especially in biological image analysis, image pre-processing includes important tasks like stitching and drift-correction to improve the quality and precision of data analysis [38]. Stitching images is crucial for reconstructing high-resolution pictures, which are essential in fields like medical imaging [39]. Drift-correction is needed for accurate high-resolution mapping of time-series data and correcting discrepancies in shape, position, and orientation within acquired images [40]. To be able to further work with and reconstruct single cell movements from image data, cells have to be segmented, meaning an algorithm has to distinguish a cell from the background as well as a cell from other cells. A segmentation algorithm is trained to assign pixels of an image to different labels, producing masks of the input image, where each labeled region corresponds to one cell. Deep learning segmentation algorithms have rapidly advanced in recent years, allowing for the segmentation of individual cells and organelles with near-human accuracy [41], [42]. Recent advancements in imaging technologies and computational tools have made possible detailed investigations into how specific gene knockdowns influence cell morphology and behavior. The molecular signal of a wound nearby inside the cell tissue exists most primarily in the leading edge cells. The width of the leading edge cells is approximately three cell layers, a finding made in CyTOF experiments by the Linding Lab in collaboration with the Bodenmiller Lab [43]. Especially in the leading edge, different morphological features generated by various treatment conditions and affected by wound healing

can be observed. Therefore, measuring shape features of individual cells in the leading edge of the scratch and comparing them to the bulk cells and to other siRNA knockdown leading edge cells is of importance [44]. Aiming to explain patterns that may indicate similar phenotypes among grouped knockdowns, shape features with most discriminatory power can be selected for analysis [24], [45]–[47]. By utilizing advanced image analysis algorithms [41], [45] one is able to extract quantitative features related to cell shape, such as area, perimeter, aspect ratio, and curvature, across various siRNA knockdown conditions. Another key observation enabled by live-cell imaging experiments is tracking single-cell trajectories of cell movement in response to different stimuli. Cell trajectories can capture variations between individual cells and how these differences change over time [48]. Variances in migration dynamics due to siRNA knockdowns are sometimes only detectable by analysing single-cell trajectories [48]. To calculate single-cell trajectories, it needs to be determined which cell in an image of a time step corresponds to the same cell in an image of another time step, also called cell tracking. The latest progress in technologies has significantly improved the ability to track single cells as entities, allowing for continuous monitoring of individual cells over time [35]. DeepCell Tracking [35] for example is a method written in Python and built with TensorFlow 2, using deep convolutional neural networks, a supervised machine learning method, for creating cell lineages and trajectories. DeepCell Tracking was trained on a data set consisting of over 11,000 trajectories of cell nuclei and it performs robustly on a variety of cell types and imaging modalities [35]. Common measures from physics and biology to analyse trajectories are mean squared displacement (MSD) and persistency. MSD is a quantitative measure used in the analysis of particle or object movement in a given space over time [49]. It calculates the average of the squared distances traveled by the particle/object relative to its starting position at various time intervals. MSD is particularly valuable in studying Brownian motion, diffusive processes, and the dynamics of particles or molecules within biological systems. It provides insights into the nature of motion, such as random diffusion, directed movement, or confined motion, and can be utilized to characterize the properties of the system or particles under investigation. Persistency refers to the ability of a cell or particle to maintain its directionality over time [50]. It is a measure of how straight or persistent a trajectory is during movement. Cells or particles with high persistence tend to move in a relatively straight line for longer periods, while those with low persistence exhibit more random or erratic movement patterns with frequent changes in direction. For random walks, the persistence decreases to zero as the (sub-)track length increases.

High dimensional embedding and clustering

Using single cell shape features, each cancer cell from an image data set can be transformed to a data point in a high dimensional space. In that way, the dimensions represent the cell shape features that can be extracted out of the image data. Through dimensionality reduction of high-dimensional image data, shape-related characteristics that influence changes in cellular behavior can be revealed [51]. Methods such as Uniform Manifold Approx-

imation and Projection (UMAP), Principal Component Analysis (PCA), and Variational Autoencoders (VAE) allow to condense high-dimensional cell shape data into lower-dimensional representations while preserving essential information about cell morphology variations. UMAP is particularly known for its ability to capture complex, non-linear relationships within the data and visualize high-dimensional data sets in a lower-dimensional space [52]. PCA, a classical linear technique, identifies the directions of maximum variance in the data, aiding in dimensionality reduction and feature extraction [53]. VAE, a type of deep generative model, excels in learning the underlying structure of data. It consists of an encoder network, compressing input data into a lower-dimensional latent space and a decoder network, reconstructing the original data from the latent space representation [54]. Out of these three tools UMAP offers in combination the easiest way of implementation, highest level of reproducibility and most meaningful arrangement of clusters [51]. The dimensionality reduction of images of cancer cells affected by different conditions not only aids in visualization but also facilitates the identification of key shape-related features contributing to phenotypic changes in migration experiments induced by gene knock-downs [51]. The Leiden clustering algorithm [55], known for its effectiveness in identifying distinct communities within complex networks, can then be applied to the high-dimensional neighborhood graph. By grouping knock-downs based on similarities in cell shape features, Leiden clustering helps to identify potential clusters of knock-downs with underlaying phenotypic patterns. This clustering approach enables to uncover hidden relationships and patterns among knockdown conditions that may not be apparent through traditional analysis methods [56]. This approach has found applications in various fields such as single-cell gene expression analysis, metagenomic binning, and network graph clustering, demonstrating its versatility and robustness in handling diverse data sets [57]. Through an integrated approach of cell shape feature extraction, dimensionality reduction, and Leiden clustering, deeper insights into how specific gene knockdowns modulate cell morphology in breast cancer cells may be gained. Identifying knockdowns with similar phenotypic signatures not only enhances the understanding of underlying molecular mechanisms but also holds promise for targeted therapeutic interventions tailored to specific cancer subtypes.

Cell line, experiment and data

In this thesis, MDA-MB-231 cells were analysed. The MDA-MB-231 cell line is widely recognized as a valuable model in breast cancer studies. This cell line is categorized as triple-negative/basal-like, as it does not express hormone receptors (estrogen and progesterone) or HER2 [58]. MDA-MB-231 cells were derived from a metastatic breast adenocarcinoma and exhibit aggressive migratory and invasive properties, making it an ideal model for studying cancer cell progression and metastasis mechanisms [59]. MDA-MB-231 has been utilized to study the effects of various compounds on cell proliferation, migration, and apoptosis, underscoring its significance as a model for comprehending breast cancer biology and developing potential therapeutic interventions [60]. The im-

age data set with MDA-MB-231 cells that was analysed, was previously generated in the Linding Lab [43]. In order to examine the migration of breast cancer cells, the Linding Lab [43] performed an extensive literature research and chose to investigate the migratory response of 352 siRNA knockdowns and 4 known siRNA knockdowns as controls on the MDA-MB-231 cell line. The latter consist of the non-targeting control (NTC1), unperturbed cells, myosin heavy chain 9 (MYH9), which was expected to increase migration [61], POU class 5 homeobox 1 (POU5F1) [62], which was expected to decrease migration, and finally polo-like kinase 1 (PLK1) as a 'no migration' control due to causing apoptosis [63]. The experiment was conducted using a MORP_CSS21_1 384 PerkinElmer Cell-Carrier, a plate containing 384 wells. Each well contained cells subjected to one of the selected siRNA knockdowns. Each control siRNA knockdown was present in 8 separate wells for statistical robustness. The cancer cells underwent transfection with pooled Silencer Select siRNAs 48 hours before experimentation. After gene knockdown ongoing for 48 h, a mechanical scratch was introduced into the monolayer, and cells were monitored through imaging for 16 hours at 1-hour intervals using a HCS at 10x magnification (resolution of 1.3 μ m per pixel). The HCS used to acquire the data set analysed in this thesis was the 'Opera High Content Screening System' by PerkinElmer. This imaging aimed to detect red fluorescent protein-tagged histone (nuclear staining) and diffuse green fluorescent protein (cytoplasmic staining) and saved these information in two image channels [64]. The experiment was carried out three times the exact same way for reproducibility and data robustness. For each siRNA knockdown the data set covers one image file with a 17 time point z-stack containing two channels, which implies 19,584 (90 GB) images in total.

Research aims

This thesis centers on employing image-based analysis techniques to investigate knockdown-dependent shape transitions of migrating MDA-MB-231 cancer cells in a wound healing assay. It explores the image data set provided by the Linding Lab [43] and aids in revealing more of its potential capabilities by creating a workflow to analyse the detectable cell migration and single-cell shape transitions. To examine cell migration, methods to identify the leading edge of moving cells, to estimate scratch area over time and to assess the directionality of movement are developed. Further, tools for single-cell tracking and measures to analyse those trajectories are implemented. Additionally, to find out about morphological patterns due to siRNA knockdowns, an approach to extract and cluster single-cell shape features is created.

Currently there is still too limited systems-level comprehension regarding the organization and composition of signaling networks that govern cell morphology [65]. Thus, this thesis shows how to apply and construct tools for image analysis to help understand how specific gene knockdowns influence shape transitions and migratory behaviors of MDA-MB-231 cells.

2 Methods

All example images displayed in this thesis are from knockdown NTC1 (non-targeting control), derived from cell line MDA-MB-231. In further analysis, the word 'cell' in combination with an image analysis explanation refers to the objects contained in the cytoplasma channel of the images, since it characterizes the outline and area of the cell. An overview of the written Fiji and Python code is available on Github: <https://github.com/olxssa/bachelorproject>.

2.1 Image pre-processing

Due to the functionality of the microscope used, the images that were to be analysed consisted of three sub-tiles, each containing two channels, one for the cytoplasm and one for the nucleus staining. The sub-tiles were stitched horizontally using the Fiji [38] plugin 'Microscopy Image Stitching Tool' (MIST) [39] (Fig. 2). In order to properly operate inside the stitching tool 'MIST', a maximal horizontal overlap shift of 15 px and a maximal vertical overlap shift of 7 px was defined. Those values were approximated by manually stitching some images of the data set. After stitching, the time point stack was drift-corrected using the Fiji plugin 'NanoJ Core' [40]. In order to realise the plugins on all images in parallel, Fiji macros were developed. Per knockdown and per experimental replicate, one image stack, containing 17 time point images, was generated for further analysis. One pixel has the size of $(1.3 \mu\text{m})^2 = 2.6 \mu\text{m}^2$.

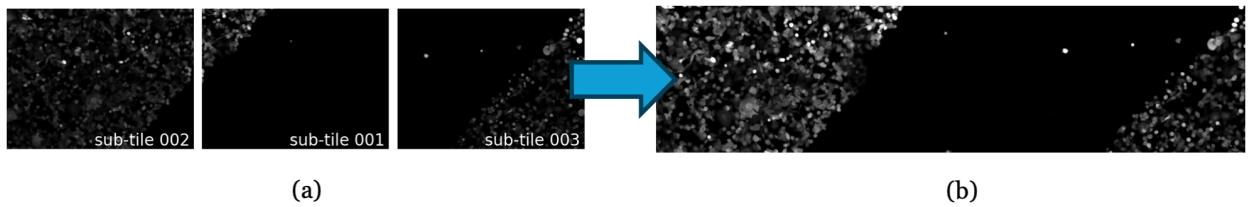


Figure 2: Image stitching. Three sub-tiles (a) are combined into one image (b) using the stitching tool 'MIST'.

2.2 Cell segmentation

For further analysis, the cells had to be segmented, meaning that each of the two-dimensional cells in the image had to be distinguished from the other cells as well as from the background. For every image and every color channel this task was resulting in an two-dimensional array containing masks of cells where pixel values had been set to the label of the corresponding cell. Using this array containing the cell masks, the cell centroids can be calculated, enabling the extraction of more information on the single cells. Segmenting single cells was achieved by utilizing Cellpose 2.0 [41]. The average cell diameter was calibrated to 16 px by Cellpose. The pre-trained model 'cyto' was improved by training it onto own annotated cytoplasm-channel data as can be seen in

Fig. (24). Additionally, the nuclei channel of the images was segmented with the pre-trained model 'nuc'.

2.3 Cell migration analysis

The general image composition of the image data set is displayed in Fig. (3). The scratch (colored in magenta) is located in the middle. On the left and on the right there is the monolayer of cell tissue. In the further analysis the left cell tissue will be characterized by the color cyan and the right cell tissue by the color orange, corresponding to Fig. (3).

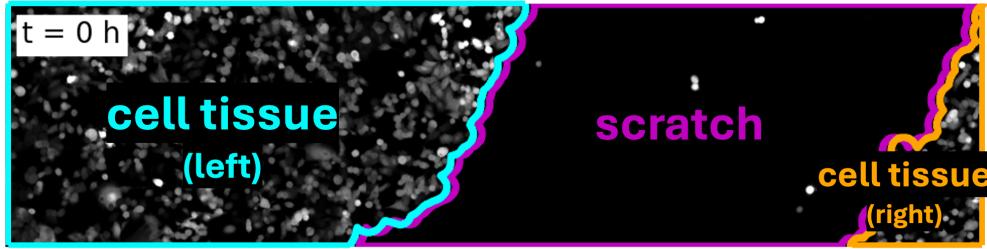


Figure 3: Composition of images in dataset. The scratch (magenta) has been done inclined in order to gain as much information about cells wandering into the scratch area as possible.

2.3.1 The scratch detection algorithm

In order to gain information on the scratch, its form, position and cells in connection to it, leading edge cells, a scratch detection algorithm was developed. Its technical functionality is inspired by Bowhead [64] and depicted in Fig. (4). At first, the raw image (Fig. 4a) is convolved with a Gaussian kernel at the standard deviation $\sigma = 5$ (Fig. 4b), and a dilation disk at disk size $d = 5$ for optimal morphological operations and then turned binary with the scipy function 'ndimage.binary_fill_holes()' [66] (Fig. 4c). The created mask of the original image is segmented by the scikit-image 'measure.regionprops()' label function [45] (Fig. 4d). Afterwards, the labeled regions in the mask image are sorted by size, keeping only the largest two regions, which are masks of the left and right tissue (Fig. 4e). Finally, the contours of the resulting two tissue masks are selected with the scikit-image function 'measure.find_contours()' and used as coordinates of the leading edge cells (Fig. 4f).

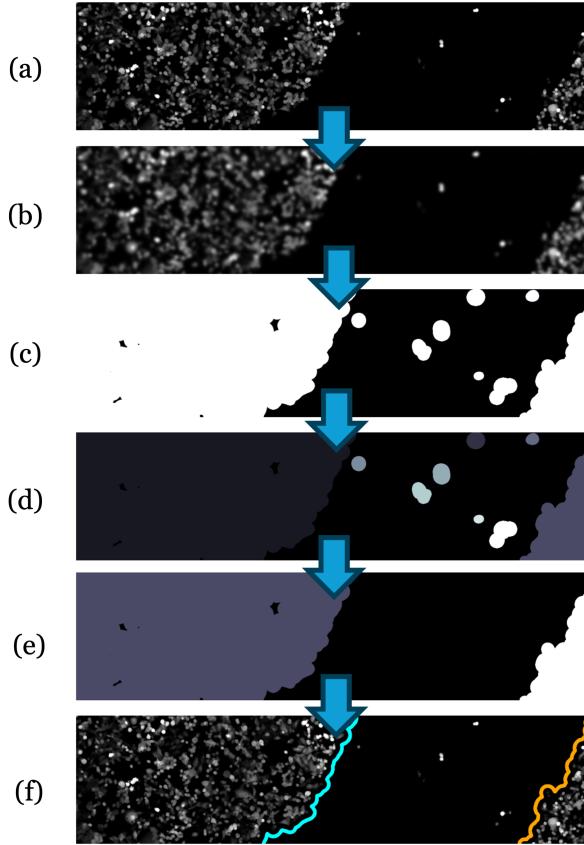


Figure 4: Technical functionality of the scratch detection algorithm. (a) Input image. (b) Blurred. (c) Segmentation of the foreground. (d) Labeling of disconnected components. (e) Selection of two biggest regions. (f) Extraction of scratch border coordinates.

2.3.1.1 Extraction of positional features

Using the segmented cell masks, the positional feature 'centroid' can be calculated by utilizing the scikit-image 'measure.regionprops()' function. The signal of a nearby wound inside the tissue is mostly present in up to three cell-layers perpendicular to the cells contacting the scratch [43]. Therefore, besides monitoring migration of tissue, the scratch detection method has also been adopted in order to gain information on border phenotype. That was realised by categorising single-cell position with respect to the scratch over time into the three categories 'inside_scratch', 'at_border' and 'bulk', as depicted in Fig. (5). A cell claims the feature 'inside_scratch' if its centroid is located inside the scratch polygon. The 'at_border' feature describes a cell, that is one of the cells of the leading edge of the tissue. In order to verify if a cell is positioned at the leading edge of the left and right tissue, polygons of the leading edges (borders) were created for every time step. The border polygons were created using the coordinates of the leading edges as well as the shifted coordinates of the leading edges accordingly in x-direction by 5 cell layers, each of a width of 16 px, meaning by $5 \cdot 16 \text{ px} = 80 \text{ px}$. The 'bulk' polygons were calculated using

the image edges as well as the leading edge coordinates, shifted by 20 cell layers, meaning by $20 \cdot 16 \text{ px} = 320 \text{ px}$.



Figure 5: Polygons created for extraction of positional features of cells. The cell centroids categorized by a polygon are colored accordingly.

2.3.2 Calculation of the closing wound area

To calculate the area of the scratch of each closing wound at each time point, the coordinates of the scratch were calculated using the scratch detection algorithm described in 2.3.1. With these, a polygon, containing all image pixels owned by the scratch, was constructed using the 'geometry.Polygon()' function of the Python module 'shapely' [67]. Thereby, to calculate the scratch area, the amount of pixels contained by the scratch polygon was summed up.

2.3.3 Estimation of velocity of the leading edge

In order to estimate the velocity of the leading edge, the scratch detection algorithm described in 2.3.1 was used to calculate the coordinates of the leading edges of the left and right tissue closing the wound over time (Fig. 6a). To each leading edge a linear function was fitted (Fig. 6b). The distance between the leading edges was calculated every hour for 100 points each on the linear functions fitted to the left and right tissue edges (Fig. 6c). Thereby, calculating the mean of the distance values of 100 points for the leading edges and taking the mean over all time point differences, the overall velocity v of the leading edge is obtained as shown in Eq. 1:

$$v = \sum_{t=0}^{17} \left(\frac{1}{2} \left[\sum_{N=1}^{100} \frac{1}{N} |(x_{left}^i(t+1) - x_{left}^i(t))| + \sum_{N=1}^{100} \frac{1}{N} |(x_{right}^i(t+1) - x_{right}^i(t))| \right] \right) \quad (1)$$

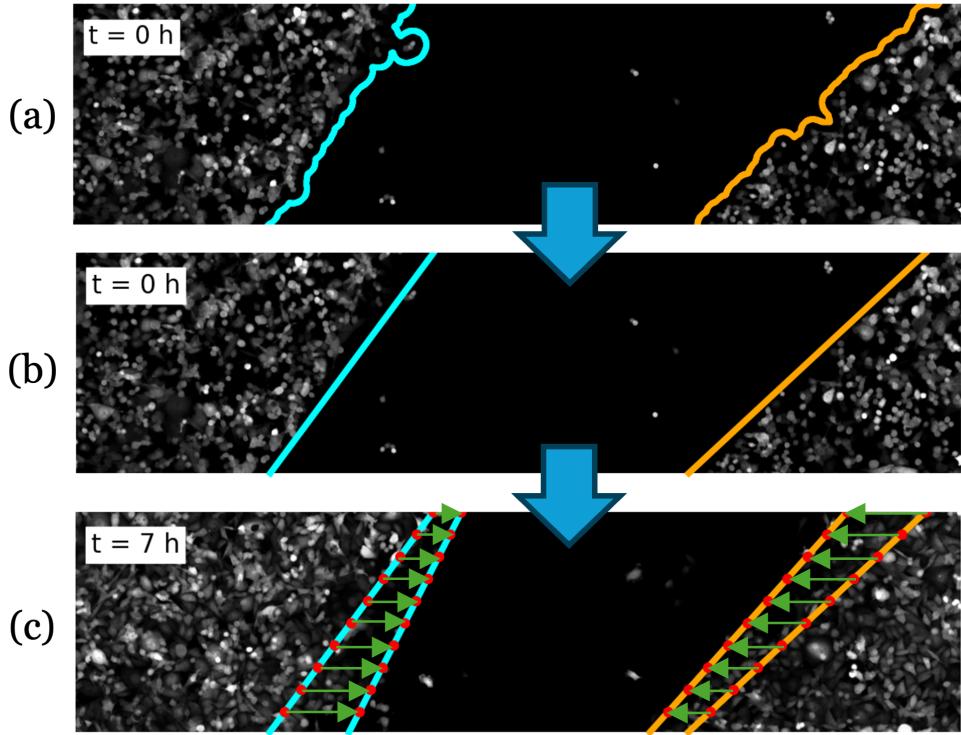


Figure 6: Estimation of velocity of leading edge cells. The distance of points on linear functions fitted to the (a) leading edge was calculated for every hour, but for displaying reasons depicted here for the distance between leading edges of (b) $t=0$ h and (c) $t=7$ h.

2.3.4 Calculation of nuclei displacement polarity

The labels of the cells were connected with the labels of the corresponding nuclei by matching nuclei centroids to the location of cell areas. By utilizing both of their cell center coordinates (cell centroids), a measure of nuclei displacement polarity can be obtained inspired by Polarity-JaM [15]. For each cell contained in the leading edge, a vector \mathbf{v}_{cell} from the cell center to the nucleus center was drawn by Eq. 2:

$$\mathbf{v}_{cell}^i = (\mathbf{x}_{nuc}^i - \mathbf{x}_{cyto}^i) \quad (2)$$

Thereby, \mathbf{x}_{cyto}^i are the cell centroid coordinates and \mathbf{x}_{nuc}^i denotes the nucleus centroid coordinates. In order to calculate the polarity angle of nuclei displacement towards the scratch, a perpendicular bisector of the scratch $\mathbf{v}_{scratch}$ was estimated. That was done by taking the leading edge coordinates found by the scratch detection algorithm, performing linear fitting and taking the mean of the two linear functions (Fig. 7). The nuclei displace-

ment polarity angle can be obtained by Eq. 3:

$$D(i, t) = \frac{\cos(\mathbf{v}_{\text{scratch}} \cdot \mathbf{v}_{\text{cell}}^i)}{|\mathbf{v}_{\text{scratch}}| \cdot |\mathbf{v}_{\text{cell}}^i|} \quad (3)$$

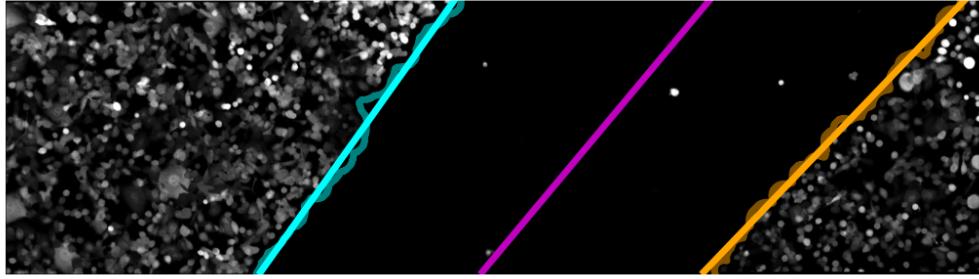


Figure 7: Estimation of perpendicular bisector of scratch. The left (cyan) and right (orange) leading edges of the tissue were detected using the scratch detection algorithm. The vector of the estimated perpendicular bisector of the scratch $\mathbf{v}_{\text{scratch}}$ is colored in magenta.

2.3.5 Cell trajectory analysis

The next step was to gain information on the single-cell positions over time, thus on their movement. Using the cell centroids of all cell over time, a tracking algorithm can follow each cell centroid along the time dimension. The result of that tracking algorithm is a cell movement trajectory that can be analysed further with different measures, like mean squared displacement (MSD) and persistence (P).

2.3.5.1 Single-cell tracking

The basis for a tracking algorithm was produced by segmenting the nucleus channel using Cellpose 2.0 described in 2.2. The first approaches tried out for this task were the three dimensional segmentation with Cellpose 2.0 [41], the python module 'trackpy' [68] and the Fiji plugin 'TrackMate' [69]. The final and working approach used for this task was DeepCell Tracking [35]. Its application 'CellTracking' was implemented in Python in order to use on the data set. It takes the three dimensional image stack as well as the three dimensional mask stack containing the segmented and labeled cells, each where one dimension is the time-axis, and gives as output mask stacks labeled correctly throughout the time dimension. Cells were tracked throughout the whole 16 hours covered by the data set, which is further defined as a track or trajectory, but cells may also only be tracked during a shorter period of time, referred to as a sub-track or sub-trajectory.

2.3.5.2 Trajectory analysis measures

To analyse the tracks and sub-tracks of the moving single cells calculated by DeepCell Tracking the measures MSD and persistency were implemented in Python, inspired by the R package CellTrackR [70].

Mean Squared Displacement

It is defined as an ensemble average [70] and calculated for each time point as shown in Eq. (4):

$$MSD(t) = \frac{1}{N} \sum_{i=1}^N |\mathbf{x}^i(t) - \mathbf{x}^i(0)|^2 \quad (4)$$

Thereby, \mathbf{x}^i denotes the centroid vector of the corresponding cell i at the time t and N represents the number of cells.

Persistency

For each time point the persistency (P) it is calculated as shown in Eq. (5-7).

$$\mathbf{v}_1 = (\mathbf{x}^i(t+1) - \mathbf{x}^i(t)) \quad (5)$$

$$\mathbf{v}_2 = (\mathbf{x}^i(t+2) - \mathbf{x}^i(t+1)) \quad (6)$$

$$P(t) = \frac{\cos(\mathbf{v}_1 \cdot \mathbf{v}_2)}{|\mathbf{v}_1| \cdot |\mathbf{v}_2|} \quad (7)$$

Thereby, \mathbf{x}^i denotes the centroid vector of the corresponding cell i at the time t .

2.4 Single-cell shape feature analysis

In order to analyse single-cell shape features, the image data was transformed into matrices for every image with rows representing cells and columns representing their cell shape features. Thus, in this chapter single-cell shape features will be named and defined. Examples of all feature measure definitions displayed on one model cell can be found in the appendix. Furthermore, this chapter introduces filtering and pre-processing steps of data for dimensionality reduction and clustering.

2.4.1 Definitions of single-cell shape features

In order to extract information about single-cell shapes, calculations of features were implemented and performed on the cell segmentation masks, generated as described in 2.2. The thereby obtained new features were integrated into the already existing matrices of the positional features described in 2.3.1.1. In total, 22 shape fea-

tures [24], [45]–[47] with most discriminatory power were selected and implemented for further analysis and are named and defined in Tab. 1:

Table 1: Extracted cell shape features.

nr.	feature	definition	unit	source
1	area	total number of pixels converted into mm ²	mm ²	[45]
2	axis_major_length	length of major axis	mm	[45]
3	axis_minor_length	length of minor axis	mm	[45]
4	eccentricity	focal distance / axis_major		[45]
5	equivalent_diameter_area	diameter of a circle with same area	mm	[45]
6	extent	area / (area of total bounding box)		[45]
7	feret_diameter_max	longest distance between points around convex hull contour	mm	[45]
8	orientation	angle between 0th axis and major axis	degrees	[45]
9	perimeter	tracing contour pixels and summing up distances	mm	[45]
10	perimeter_crofton	perimeter but approximated by Crofton formula	mm	[45]
11	solidity	area / area_convex_hull		[45]
12	convexity	perimeter / perimeter(convex_hull)		[46]
13	form_factor	area / perimeter ²		[46]
14	circularity	4 · π · area / perimeter ²		[46]
15	aspect_ratio	axis_major / axis_minor		[46]
16	perimeter_curl	perimeter / π · (1 - sqrt(1 - 4 · π · area/perimeter ²))	mm	[46]
17	curl	axis_major_length / feret_diameter_max		[46]
18	inscribed_area	axis_major_length ² · π / equivalent_diameter_area		[46]
19	sphericity	2 · equivalent_diameter_area / axis_major		[46]
20	roundness	4 · π · area / (perimeter(convex_hull)) ²		[47]
21	concavity	number of times the curvature's sign changes		[47]
22	protrusion	number of local maxima of curvature		[47]

An explanation the in Tab. 1 mentioned measures, used to calculate the features, can be found in the appendix in Fig. (28). Features with number 1-11 were extracted using the scikit-image 'measure.regionprops()' function [45]. Furthermore, features with number 12-19 were inspired by [46], where the focus was on understanding morphodynamics of MDA-MB-231 cells. There, extracting and analysing these shape features over time showed significant cell shape fluctuations, corresponding with different migration modes. Finally, features owning numbers 20-22 were implemented inspired by [47] as shown in Fig. (8).

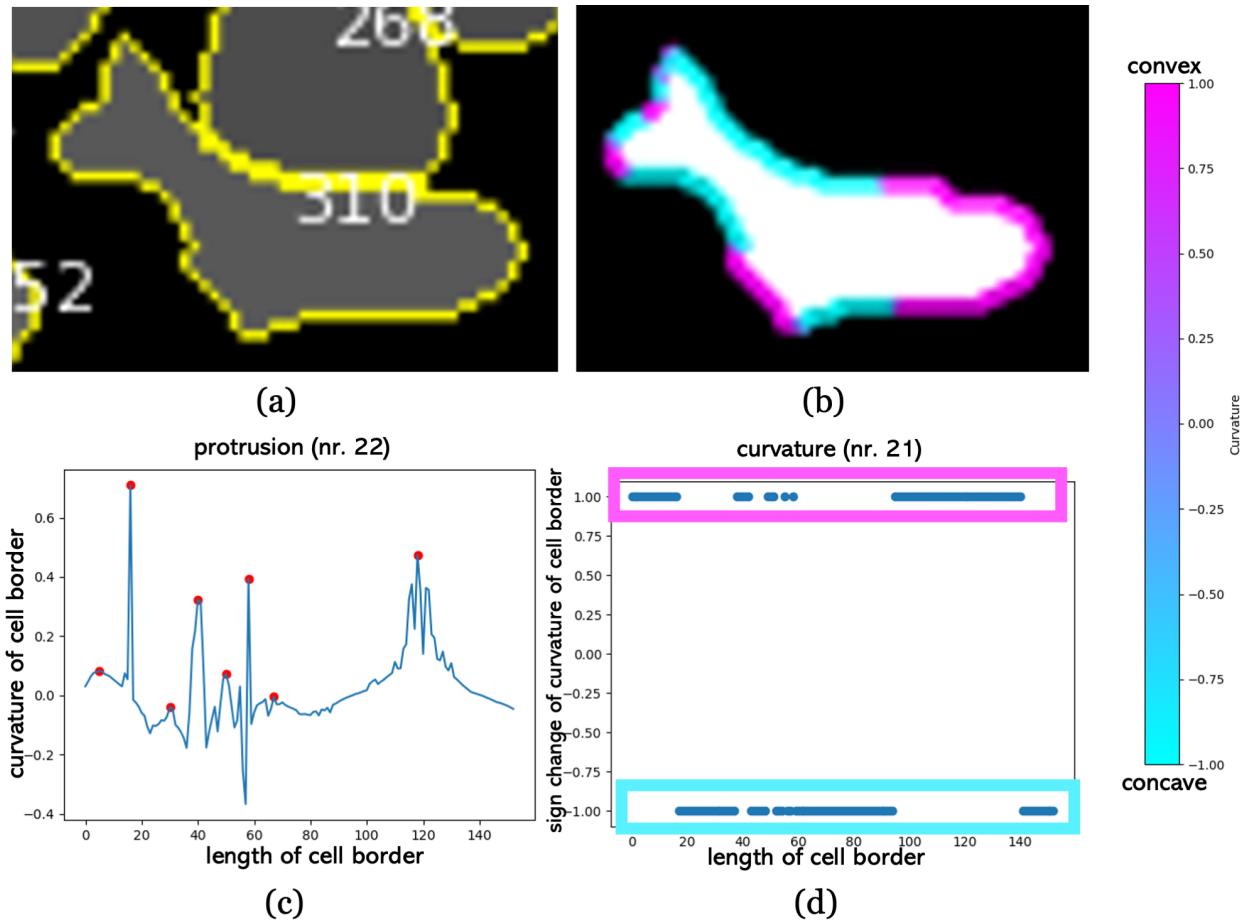


Figure 8: Estimation of curvature of cell border along its length. (a) represents the segmented cell with its predicted cell border generated by Cellpose 2.0. In (b) pink indicates convex regions while cyan indicates concave ones along the cell border length. (c) Plot of intensity of curvature along cell border length. Values greater zero infer convex regions and values smaller than zero correspond to concave regions. The red dots mark the local maxima of the curvature, also referred to as the feature 'protrusion' (feature nr. 22). (d) displays the calculation of the feature 'curvature' (feature nr. 21).

Overall by that workflow the 384 image-stacks of size (17, 440, 1772) per experimental replicate were thereby transformed into matrices where rows are representing cells and columns display their positional and shape related features as well as labeling of belonging to a certain replicate, knockdown and time point.

2.4.2 Filtering of cell shape features

In order to do the clustering with reasonable data, filtering steps were performed based on the feature distributions and experience with the data. Data of cells was kept under the conditions displayed in Tab. 2:

Table 2: Filtering conditions of cell shape features.

nr.	filtered feature	accepted range
1	area	$\geq 10 \text{ & } \leq 1500$
2	circularity	≤ 1.5
3	roundness	≤ 1.5
4	convexity	$\geq 0.97 \text{ & } \leq 1.03$
5	protrusion	≤ 10
6	inscribed_area	≤ 1500
7	at_border	True
8	inside_scratch at t=0	False

2.4.3 Dimensionality reduction and leiden clustering of cell shape features

For dimensionality reduction techniques such as UMAP, scaling of data was needed in order to create features that are comparable, each enabling to have an impact on the dimensional embedding. For that, 'StandardScaler()' by sklearn.preprocessing [71] was utilized. Since the scaled data accumulated around zero, further filtering was performed in order to generate significant data. Thus, only data showing values greater than 0.25 was kept, while the rest was set to zero. The UMAPs and leiden clusters were calculated exploratively using the Python module 'scanpy' [72] and its pre-processing option 'neighbors' as well as its tools 'umap' and 'leiden'. To show the influence of certain features on cluster development, significance tests on the clusters using the t-test were performed.

3 Results

As a general workflow, the research included in this thesis moves from processing raw plate images, to cell segmentation, over the analysis of cell migration detectable in this dataset, to the analysis of single-cell shapes. Therefore, it is structured around four primary topics, each addressing key aspects of the analysis pipeline.

3.1 Stitching and drift-correction

Every raw image gained in the wound healing assay process consists of three .flex sub-tiles. Due to very low information contained in the area of the image where the scratch is located, which is where the sub-tiles overlapped, especially in the earlier time points, stitching the sub-tiles to one whole image proofed to be problematic. For stitching, at first, the Fiji plugin 'Grid-Collection-Stitching' was tried out. Despite applying various Fiji smoothing-filters like 'Smooth' or 'Gaussian Blur' before using the stitching tool, the whole images were mostly stitched in wrong y-positions. Therefore, another Fiji plugin called 'Microscopy Image Stitching Tool' (MIST) [39] was used, yielding a much higher success in stitching the images correctly. After stitching, the drift-correction Fiji plugin 'NanoJ-Core' [40] was applied. The .ijm macro written to perform these tasks can be found in the appendix in 7.1 and on <https://github.com/olxssa/bachelorproject/>.

3.2 Cell segmentation

The combination of the pre-trained model 'cyto' and the own annotated data yields segmentation results in a quality displayed as example in Fig. (9), an essential step to enable the further analysis pipeline.

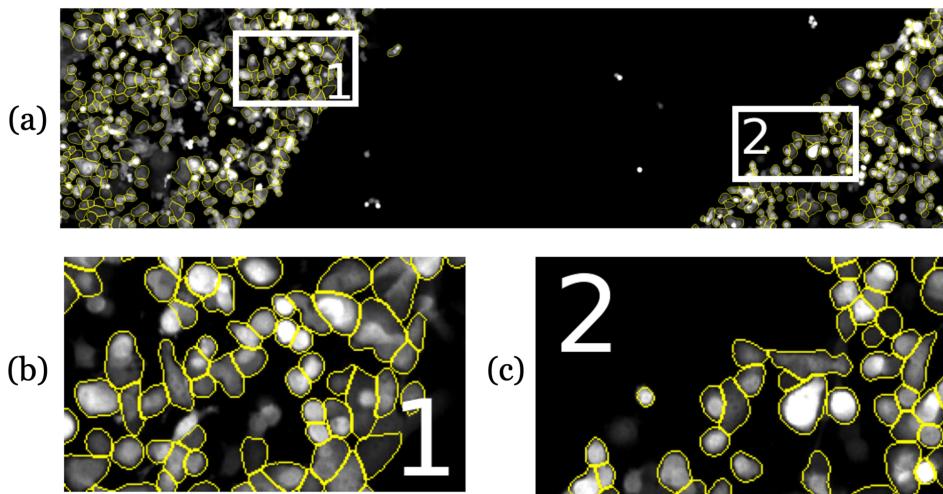


Figure 9: Cell segmentation after using Cellpose 2.0. In (b) and (c) the detected cell outlines colored in yellow can be inspected in more depth.

3.3 Cell migration analysis

In this section the cell migration analysis will be presented with a focus on the four control siRNA knockdown cells NTC1 (normal/WT migration), MYH9 (increased migration wrt. normal), POU5F1 (decreased migration wrt. normal) and PLK1 (no migration wrt. normal).

3.3.1 Area of closing wound

For every well of the multi-well plates of the three experimental replicates the area of the closing wound was calculated automated by Python scripts and as described in 2.3.2. Throughout the different wells with slightly varying wound inclinations and widths at t=0 h, a consistent and reproducible wound closure was observed.

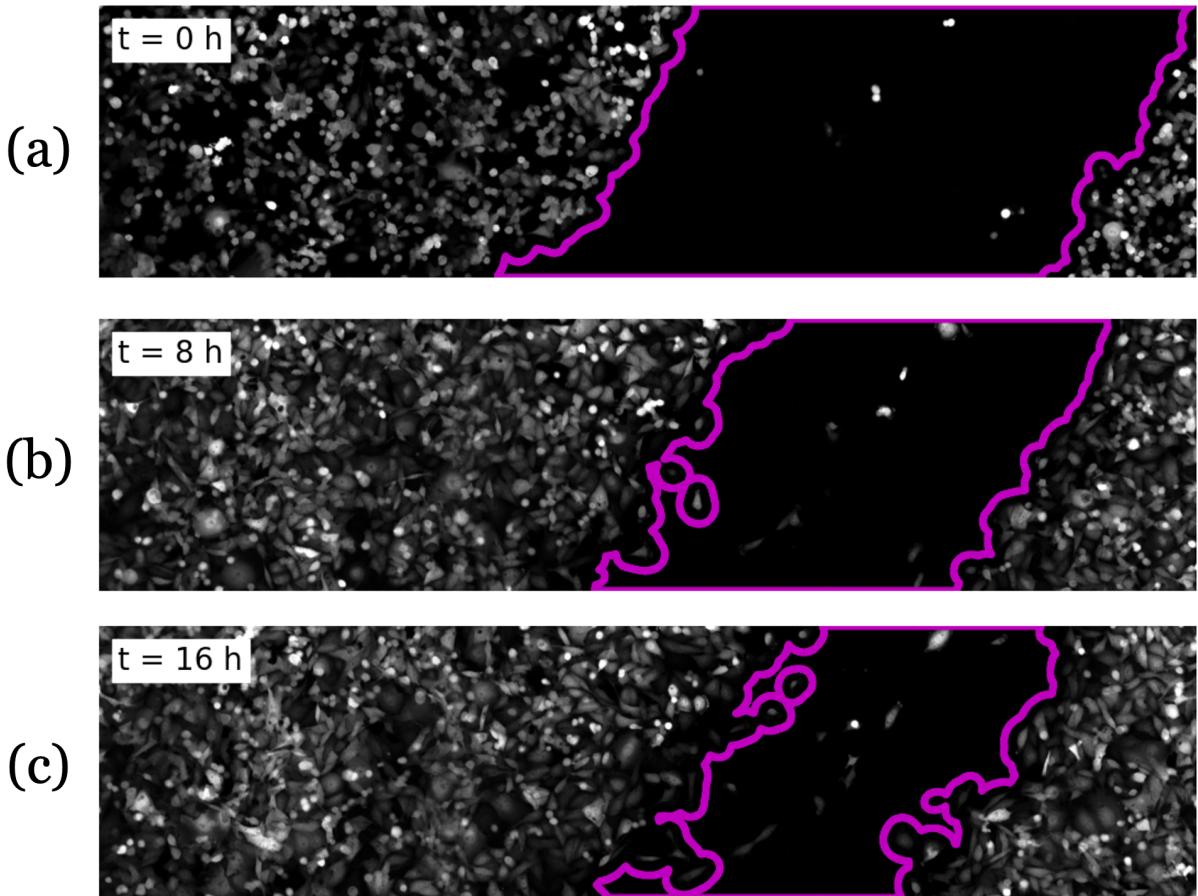


Figure 10: Scratch detection algorithm applied to NTC1 after (a) 0 h, (b) 8 h and (c) 16 h. Marked in magenta is the detected scratch contour.

The automated scratch detection method applied in this thesis proved to be robust for most of the time-series data stacks as for instance in Fig. (10). Due to noisy scratches, generated by cells drifting into the wound area after

executing the scratch, in some wells false scratch outlines were detected. An example for that case is depicted in the appendix in Fig. (25).

3.3.1.1 Reproducibility of three experiment replicates

After the calculation of the scratch area for all wells in all experimental replicates, reproducibility can be inspected for example by Fig. (11). There, the scratch area was normalized for every well at $t=0\text{ h}$ to area equals 100 %. Respectively the scratch area decreased over time. A mean was calculated for every well throughout the experimental replicates. The standard deviation was computed using Python's function '`.std()`'.

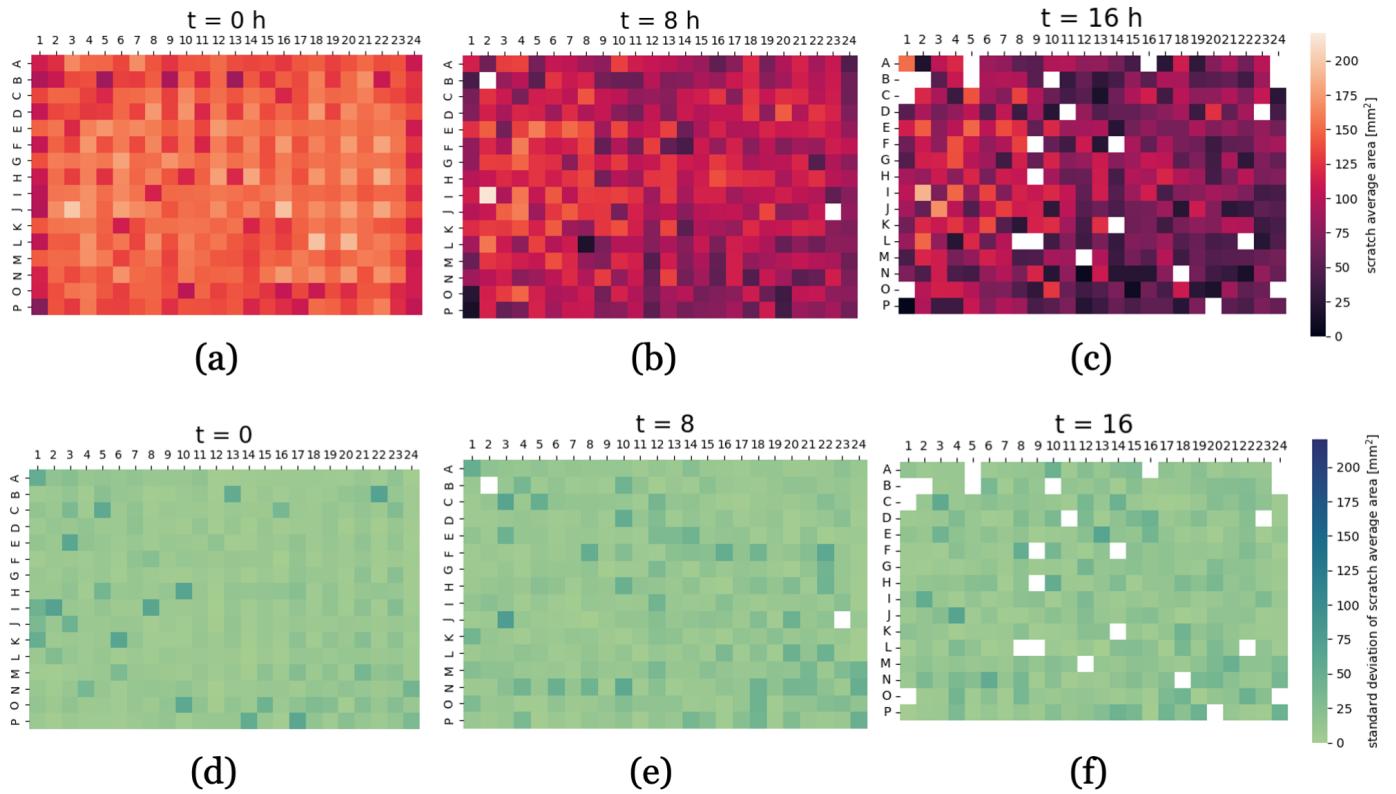


Figure 11: Plate heatmap of scratch area ((a)-(c)) and its standard deviation ((d)-(f)) over time for $t=0\text{ h}$, $t=8\text{ h}$ and $t=16\text{ h}$. Each square represents the mean of three replicates of a well in the multi-well plate. White squares represent wells where the scratch contours could not be detected, hence scratch area could not be determined.

Scratch area closes over time, indicated by colors getting darker. White squares occurring over time show wells where the scratch detection was not able to detect any scratch outlines which is why no scratch area could be determined. At $t=16\text{ h}$ the white squares appear prominently at the edges of the plate, at a total number of 9 times ($9/76 = 11.8\%$), while inside the plate the white squares appear at a total number of 14 times ($14/308 = 4.6\%$).

For the control knockdowns a clear trend of smooth wound closing is visible for a concatenated dataset of all replicates in Fig. (12). It was calculated using means of the eight wells present for each control knockdown in each experimental replicate plate.

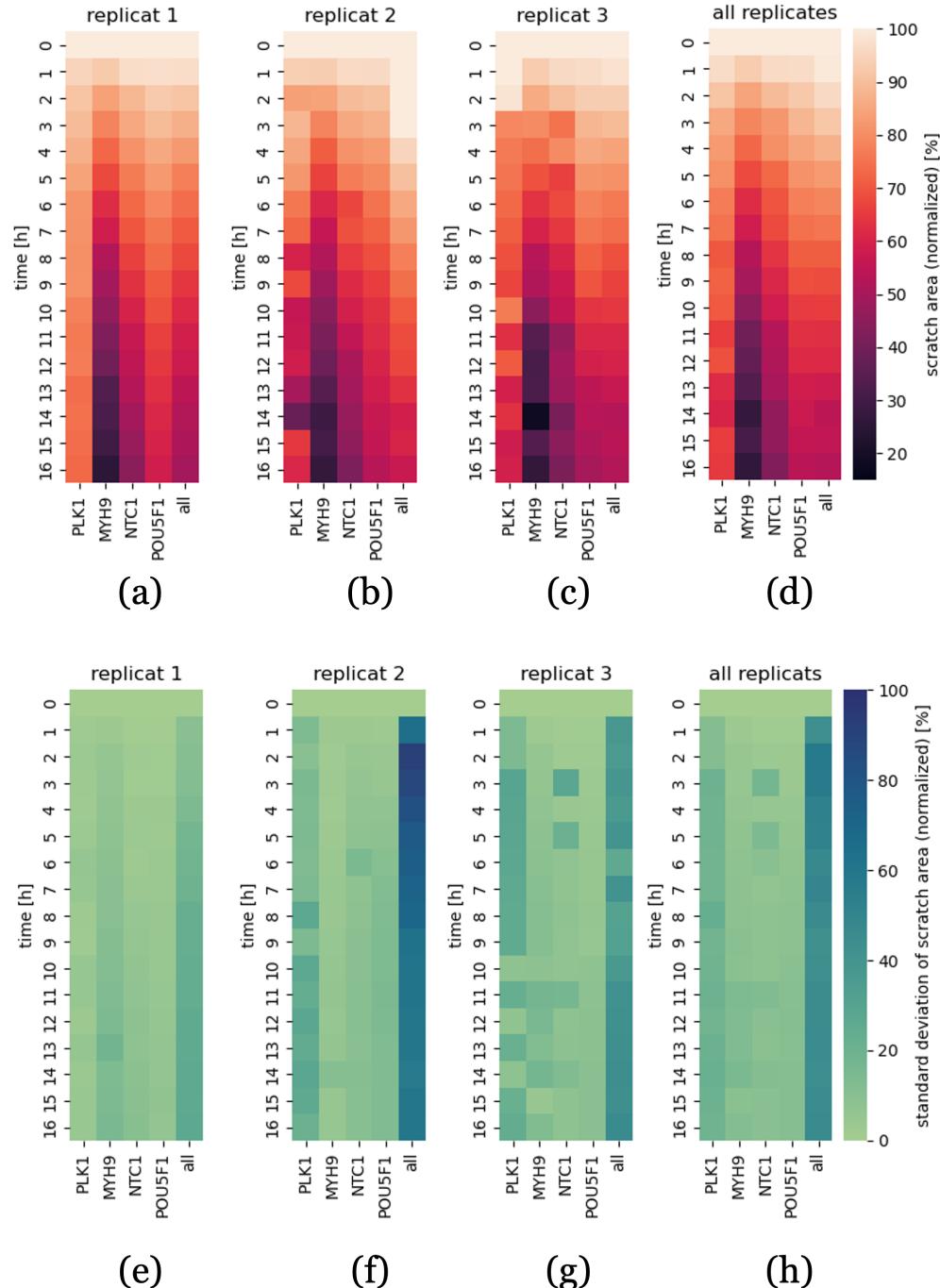


Figure 12: Calculated mean scratch area of control knockdowns ($n=8$) and of all knockdowns ($n=352$) ((a)-(c)) and their standard deviations ((e)-(g)) over 16 hours. In (d) the mean and in (h) the standard deviation has been taken over all time-series area data over all experimental replicates for each knockdown ($n=32$).

Fig. (12a) displays a trend of smooth gap closure as well, but (b) and (c) indicate variability since wound healing appears not as consistent as in (a). Especially in (12f) the standard deviation for 'all' knockdowns is very high.

3.3.2 Estimation of velocity of leading edges

Wound closing can be also referred to as velocity of the tissue, concerted cell movement realising the closing of the scratch. To compare the leading edge velocity of the different control siRNA knockdowns, the scratch detection algorithm was modified in order to calculate the global velocity of the leading edges moving towards each other as shown in 2.3.3. Fig. (13) shows the mean leading edge velocities for every knockdown ($n=8$) per replicate.

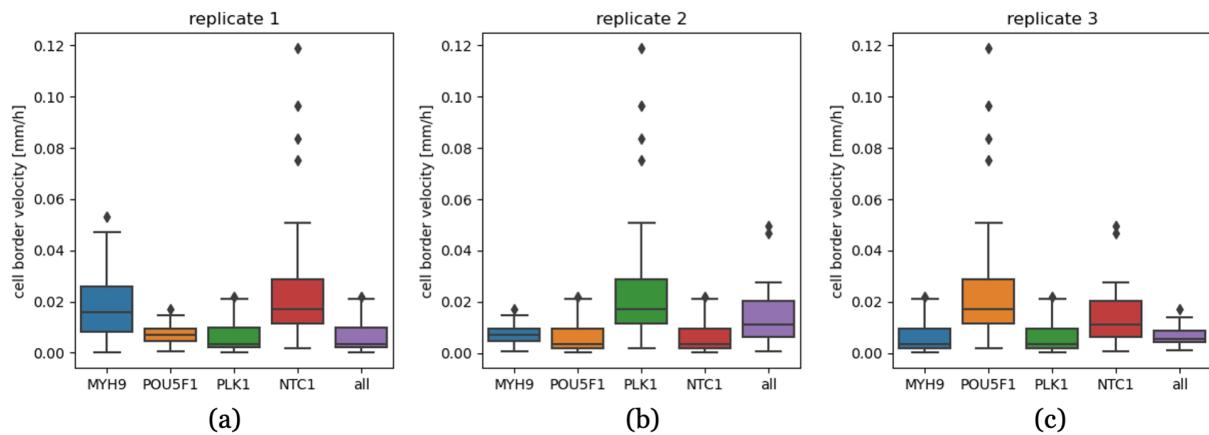


Figure 13: Estimated mean leading edge velocities of the control knockdowns and of all knockdowns for the three experimental replicates. (a) replicate 1. (b) replicate 2. (c) replicate 3.

While Fig. (13a) reflects the characteristics of the control knockdowns in a good manner by MYH9 and NTC1 being the fastest, (b) and (c) show totally different velocity distributions. In particular in (b) PLK1 is indicated to be the fastest, while actually being the 'non-migration' control.

3.3.3 Nuclei displacement polarity angle

In order to gain information on a measure of cell organelle polarity, at first, the cells of the leading edge, which are the most affected by the wound, and their nuclei were selected for each image as depicted in Fig. (14) using the positional feature extraction explained in 2.3.1.1.

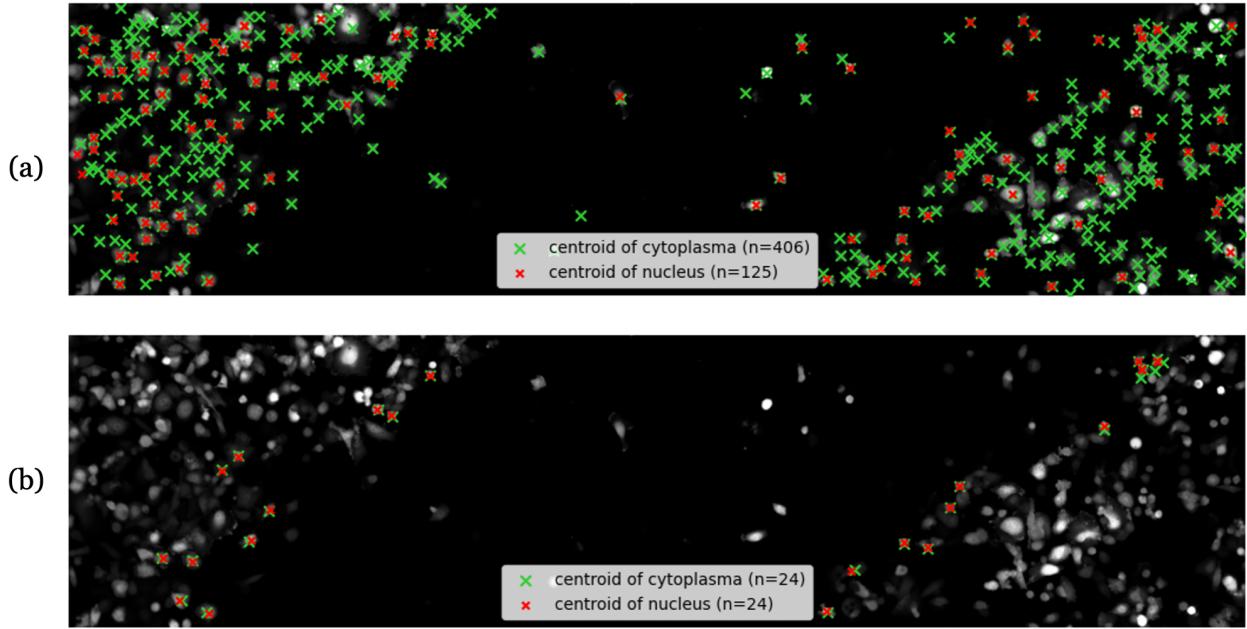


Figure 14: Cell centroids and nuclei centroids at $t=16$ h. Centroids of cells (cytoplasm) are marked in green and centroids of nuclei are marked in red. In (a) all detected cell centers ($n=406$) and their corresponding nuclei ($n=125$) are marked. The marked cells in (b) are labeled to be located the leading edge cells.

As shown in Fig. (14b) the number of leading edge cells with annotated nucleus decreases rapidly compared to cell number in total.

After choosing the correct cells and their nuclei, the nuclei displacement polarity angle towards the perpendicular bisector of the scratch was calculated as described in 2.3.4. Examples of the nuclei centroid displacement with respect to the corresponding cytoplasm centroid are displayed in Fig. (15).

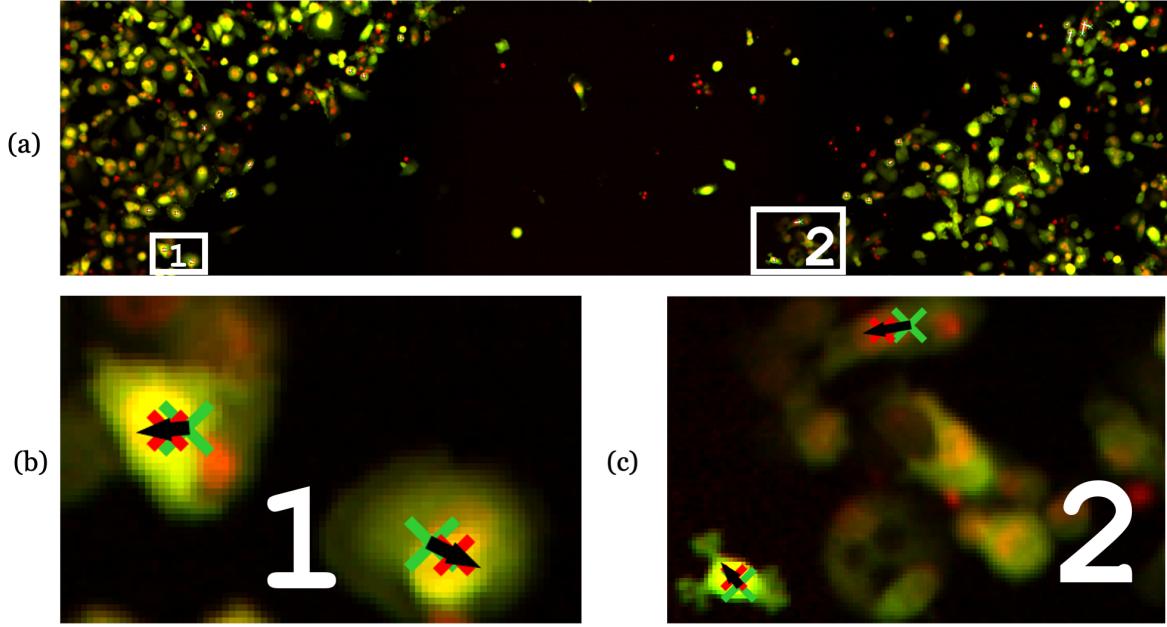


Figure 15: Nuclei displacement at $t=16$ h. The centroids of the cells (cytoplasm) are marked with green 'x', while the centroids of the nuclei are marked with red 'x'. In (a) the white arrows symbolise the calculated displacement of the nuclei with respect to the cytoplasm. In (b) and (c) the detected nuclei displacements can be observed in more detail. For better visibility, in (b) and (c) the displacement arrows are colored in black.

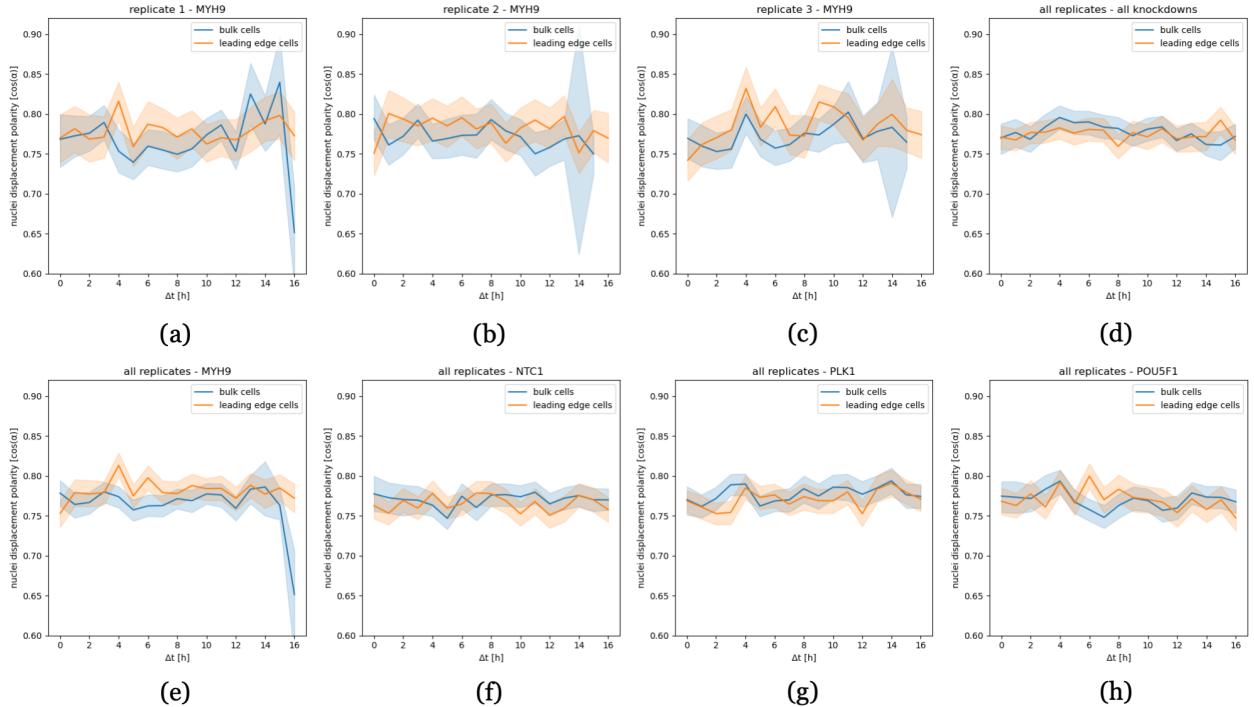


Figure 16: Nuclei displacement polarity angle over time of leading edge cells compared to bulk cells. (a)-(c) Variability of the polarity angles for knockdown MYH9 for the different replicates. (d) Angle progress over time for all knockdowns in all experimental replicates. ((e)-(h)) Angle evolvement over 16 h for means of control knockdowns.

Fig. (16) shows the calculated polarity angle evolvements over time. The angles and their developments seem very consistent throughout the different controls and all knockdowns. Further, against the expectations, the angles of the bulk cells and of the leading edge cells appear to be similar as well. Polarity angles of nuclei displacement for all control knockdowns as well as for all cells can be found in the appendix in Fig. (27).

3.3.4 Cell trajectory analysis

In order to analyse cell trajectories, single cells were tracked utilizing the CellTracker by DeepCell-Tracker and analysed by two measures both implemented in Python.

3.3.4.1 Single-cell tracking

The CellTracker makes it possible to follow some single cells for the whole experimental time of 16 h and most of them for a fraction of that time, generating data of tracks and sub-tracks. Using cytoplasm-channel data yielded a much higher ratio in tracked cells than the nucleus-channel data. Fig. (17) displays the results for one well of one experimental replicate of the tracking of cytoplasm-channel data between the time steps $t=0$ h and $t=1$ h.

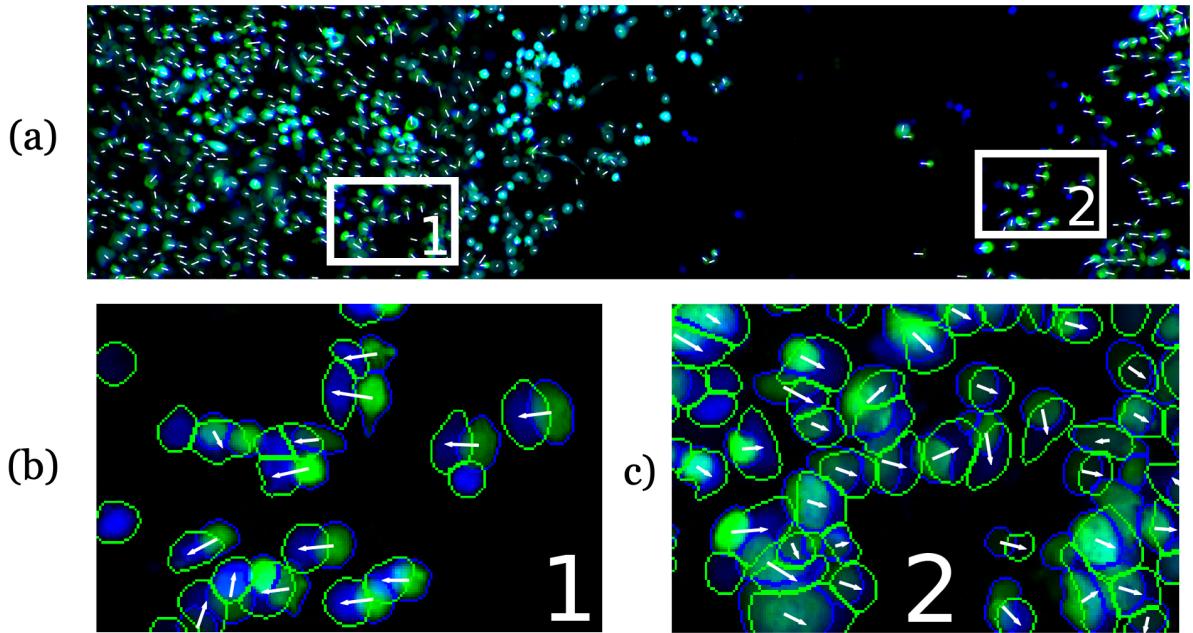


Figure 17: Tracked cells for a time difference of one hour. Cells colored in green (blue borders) are of $t=0$ h and cells colored in blue (green borders) are of $t=1$ h. The white arrows start at the centroid of the green cell at $t=0$ h and end with their tip at the centroid of the blue cell at $t=1$ h.

When combining the sub-tracks in order to cover the whole time of the wound healing experiment, most of the cells get filtered out. Examples for whole tracks are being shown in Fig. (18).

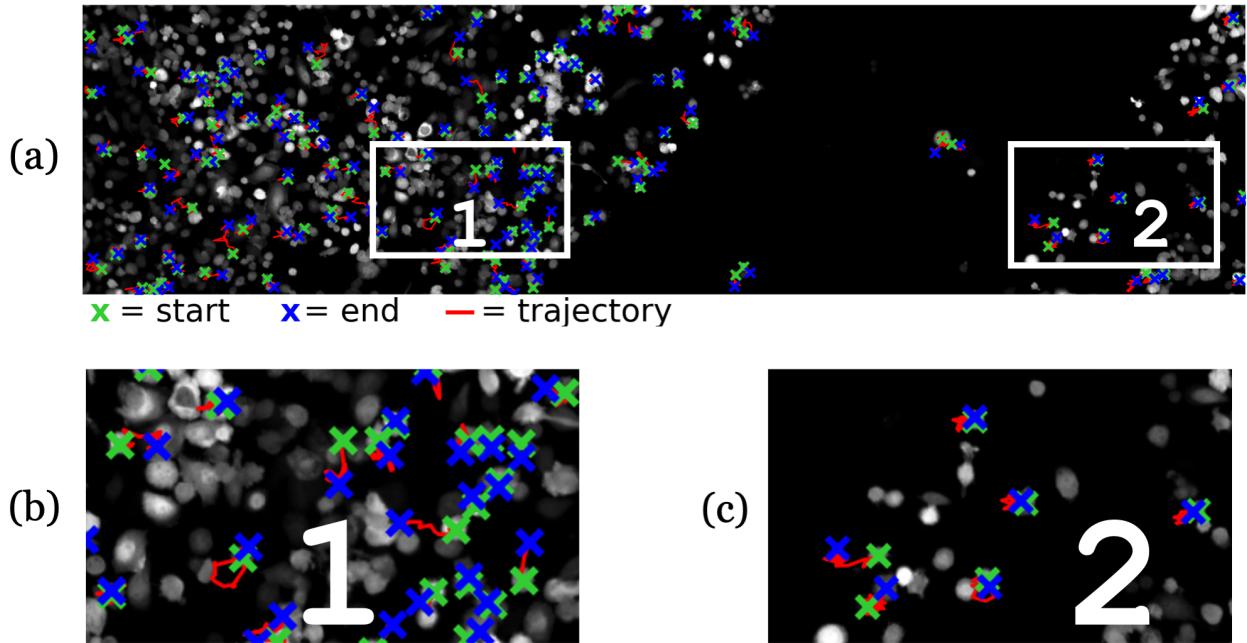


Figure 18: Whole cell tracks of one experiment plotted onto the cells at $t=0$ h. The tracking algorithm was able to follow 136 cells from $t=0$ h until $t=16$ h, while 762 cells were present at $t=0$ h, meaning that approximately 18 % of the cells were tracked for the whole time for this well. In (b) and (c) the cell trajectories can be observed in more depth.

The first experimental replicate showed an overall percentage of 3.3 % wholly tracked cells throughout the whole 17 hours of the experiment, while replicate 2 showed 3.6 % and replicate 3 showed 4.4 % completely followed cells. In total that results in a percentage of (3.8 ± 0.8) % of entirely tracked cells.

3.3.4.2 Trajectory analysis measures

In contrast to the size of the image, single-cell movements appear rather small. To analyse them, certain measures are applied for the cells starting their movement at the leading edge. Fig. (19) displays the mean squared displacement of the means of the controls and of all cells throughout replicates as well as combined for all replicates. A clear trend appears to be the robust high MSD of MYH9 and NTC1 compared to the other knockdowns throughout the replicates 1 and 2, reflecting the characteristic of the aggressively migrating knockdown MYH9.

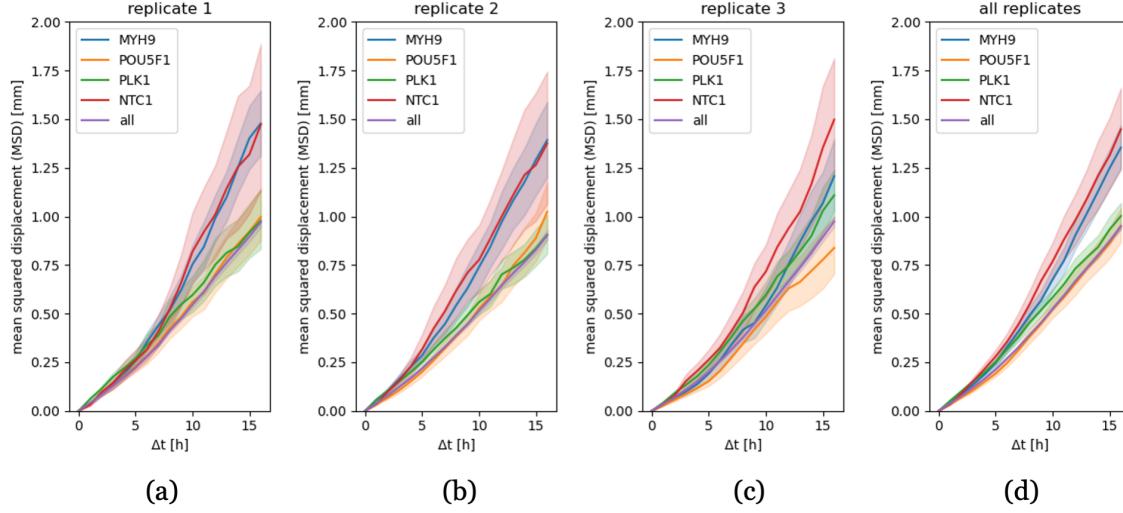


Figure 19: Mean squared displacement of whole tracks of leading edge cells of control knockdowns. The graph is plotted with a confidence interval width of 95 %. ((a)-(c)) Plots for each replicate. (d) All replicates.

The other defined trajectory analysis measure, namely presistence P , depicted in Fig. (20), shows little variability throughout the experimental replicates and similar values and evolvements for all controls.

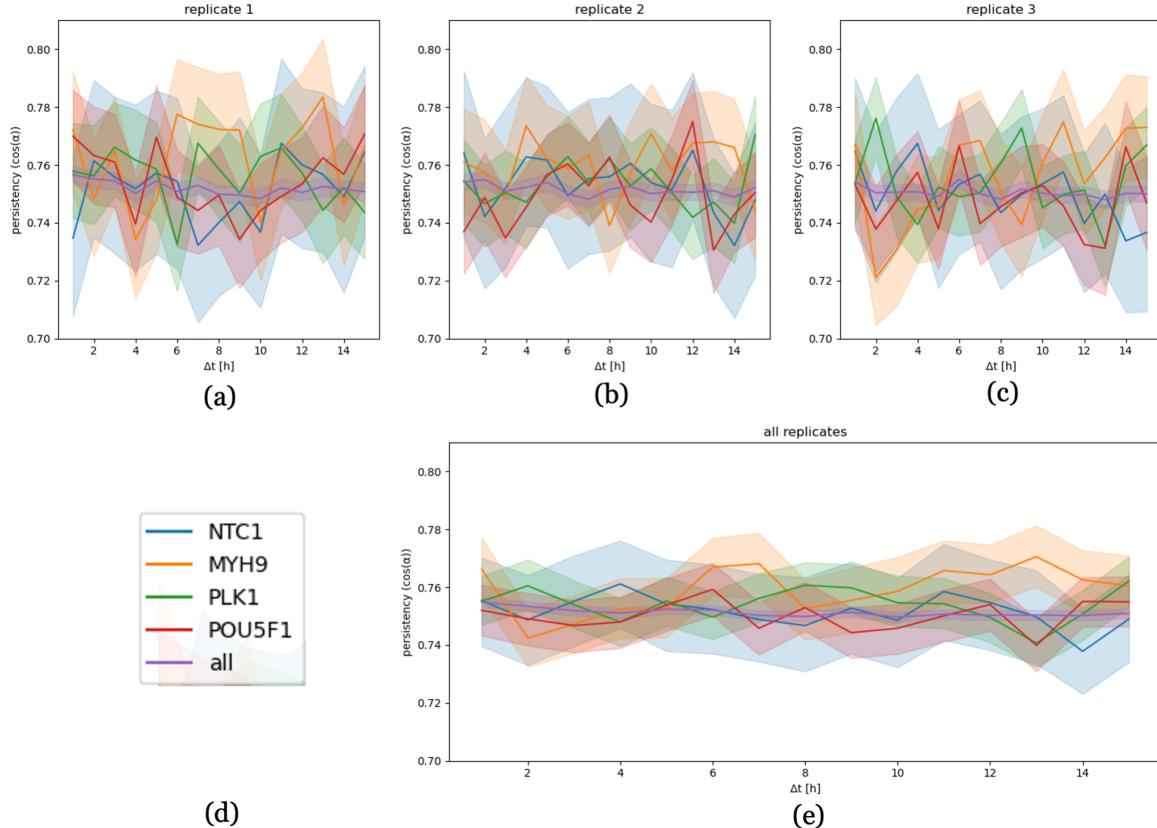


Figure 20: Persistency of whole tracks of leading edge cells of control knockdowns. All graphs are plotted with a confidence interval width of 95 %. ((a)-(c)) Plots for each replicate. (d) Legend. (e) All replicates.

3.4 Single-cell feature analysis

3.4.1 Feature distributions

In order to compare the variability of single-cell shape feature extraction throughout the three experimental replicates, histograms were calculated and plotted in top of each other in Fig. (21).

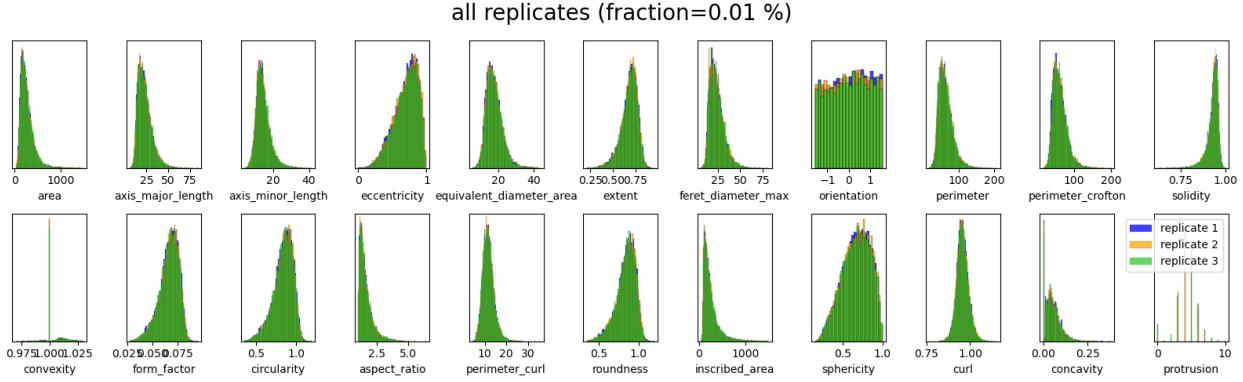


Figure 21: Cell shape feature distributions for $t=16$ h of all three replicates plotted on top of each other. Replicate 1 is colored in blue, replicate 2 in orange and replicate 3 in green. All of the distributions are overlapping very well. Displayed is a sample fraction of 0.01 % of the filtered cells for visibility.

The single-cell shape feature distributions overlap beautifully, indicating comparable extracted cell shape feature data to do high-dimensionality embedding via UMAP and Leiden clustering in the next steps of the analysis pipeline.

3.4.2 Clustering

When performing UMAP and Leiden clustering on the extracted single-cell shape feature data as described in 2.4, different UMAPs and various numbers of clusters appear, highly dependent on amount of data used, as well as highly dependent on the parameters entered as shown in Fig. (22).

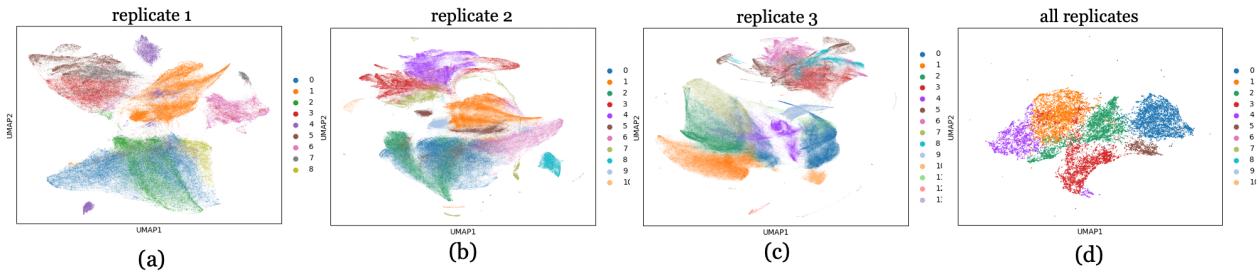


Figure 22: UMAPs with Leiden clustering of all experimental replicates at $t=16$ h in ((a)-(c)). In (d) using a sample fraction of 30 % of all concatenated replicate data for $t=16$ h an UMAP was calculated.

In the process of evaluation of generated clusters it appeared that each cluster was dominated by the cells of the control knockdowns in an approximate even distribution. That indicates that the clustering based on the in this workflow extracted single-cell shape features may not reflect to suspected migration phenotypes after all.

4 Discussion

This thesis centered on exploring the provided image data set. It contains a working image analysis pipeline that was implemented and displayed, focusing on image pre-processing and extracting migration data as well as cell shape features. Due to the limited scope of thesis it was not able to present results in the form of biological insights gained by the analysis. In order to use the image analysis pipeline to gain such insights, several problems have to be addressed.

Experiment

The workflow of the image analysis pipeline was based partly on the assumption that EMT is happening and therefore significant concerted tissue shape transitions, movements and single-cell shape changes occur. In this thesis it could not be confirmed if the EMT is well covered by the experiment that has been done, since the cells are tumor cells, based on the chosen cell line. The tumor cells already contain mutations and the knockdown was ongoing for 48 hours. That leads to the question if the cells were very mesenchymal to begin with. But besides the cell line being well-researched for the purpose of studying metastasis, distinct and characteristic differences in behaviour could also be observed by eye. For instance, in comparison to NTC1, POU5F1 or PLK1, the migration of control knockdown MYH9 was throughout all experimental replicates noticed as fast spreading, and therefore more aggressively.

Stitching and drift-correction

Single-cell analysis relies on perfectly stitched images representing cells in a biological reasonably way and the result would be distorted even by a sub-tile displacement of a few pixels. Both stitching and drift-correction was done using Fiji [38]. By randomly selecting approximately 30 image files, that have been stitched and drift-corrected, and doing a control of stitching and drift-correction by eye, the pre-processing was declared as successful. The control was done by comparing the overlapping regions of the image sub-tiles and checking if intersected cells were aligned in a biologically correct manner. To be absolutely sure for every single image that stitching does not alter the biological correctness, future research could use only the middle sub-tile of the images, containing most of the scratch area, for the analysis, to avoid the process of stitching.

Cell segmentation

Compared to simple threshold-segmentation techniques used in the past, algorithms like Cellpose 2.0 and Deep-Cell have revolutionized the field by providing precise and detailed segmentation results, enhancing our understanding of cellular processes and enabling more sophisticated analyses of biological samples. Automatic border

detection of cells enabled objective and reproducible quantification of assays. The implemented segmentation method utilizing Cellpose 2.0 yields a high success in segmenting the cells in a biologically correct manner, which was controlled by eye by selecting approximately 10 image stacks and comparing the raw data with the segmented cell outlines. In order to account for the high variability in cell shapes contained by this image data set, even more training data could be generated by annotating by hand. That would lead to more precision in the cell shape outline detection and ensure that further biological analysis relies on high quality extracted single-cell shapes.

Scratch detection of closing wound

In order to automatically distinguish between scratch and tissue, the scratch detection algorithm is needed. By comparing scratch closing using a method that detects the scratch outlines based on firm rules, wound area closure over time and velocity calculations can be more comparable. The scratch detection algorithm was implemented in a way to perform more robustly on heterogeneous tissue by virtually closing small areas in the tissue without cells. Since there are cases in scratch detection where the algorithm fails to detect the in the eye of the human observer 'correct' scratch border, it remains not entirely comparable for all image stacks. Due to the limited scope of this thesis it could not be improved by making it more robust against cells drifting into the scratch area and therefore interfering with the mask generation of the tissue areas. Improving the scratch detection algorithm as well as comparing it to existing tools like Bowhead [64] would be one great project for future research. For that task, besides trying out another marching-squares algorithm, calculated distance fields as depicted in Fig. (26) could be applied. Cells of the leading edges of the tissue could be detected by a small distance field value as well as possessing many neighbor cells. In that way, cells inside the scratch could be filtered out and banished from the further scratch detection. According to [18], cells that exist in scratch area at $t=0$ h can determine migration of cells intensely, which is why filtering out these could also produce more comparable results. Also, scratch detection worked best for time points around $t=0$ h, where the scratch outlines appear the most even. Another approach for scratch detection would be to keep in mind the result of scratch coordinates from $t=0$ h and check if sudden extreme changes, in scratch shape variation or leading edge velocity, happen. Then parameters for sensibility of scratch detection could be adjusted. Furthermore, for calculating leading edge velocity, filtering based on high variability or high standard deviation as depicted in Fig. (12) of knockdowns could be applied before the velocity calculations in order to filter out image stacks with failed scratch detection, that currently has a big impact on wound area calculations as well. Furthermore, curvature profiles of the scratch could be analysed over time and a scratch shape classifier could be developed, enabling to characterize siRNA knockdowns based on their scratch contour. In addition, a mathematical model of the movement of the leading edge could be developed and compared to the work in [44].

Nuclei displacement

For assessing cell organelle polarity towards the closing wound, the nuclei displacement method was implemented. As displayed in Fig. (15), the correctly placed arrows indicate a well implemented and working method to assess cell organelle polarity in a wound healing experiment as a measure of directionality of cell movement towards the scratch in order to close it. In further research the measure of the nuclei displacement angle could be improved by filtering out cells that are located in the area of the stitching regions in order to ensure that nuclei displacement happens due to a biological phenomenon and not by stitching errors. Furthermore, a measure to characterise dead or destroyed cells could be developed, enabling filtering them out of further analysis.

Single-cell tracking

Despite the image data set not being explicitly designed for single-cell analysis, the CellTracker by DeepCell enabled generation of single-cell trajectories and their analysis. Due to the big time steps of 1 hour in between image acquisition, the ratio of cells that were tracked throughout the whole experiment time of 16 hours was rather small compared to the total number of cells present in each well since the algorithm was not able to find the same cell again in the frame of the next time step. But as depicted in Fig. (17), tracking appears to be correct for the sub-tracks looked at for control purposes. In order to generate complete trajectories throughout the whole experiment duration for approximately 90 % of cells, the image acquisition must take place every 5 to 10 minutes [35]. In further research, a neural network could be built, trained to generate data for time steps in between image acquisition.

Single-cell trajectory measures

Trajectory measures such as MSD estimates may exhibit bias in applications involving a finite field of view, such as microscopy, due to slower objects remaining within the field of view for extended durations. This can introduce challenges when comparing different populations. Persistence is an important parameter in characterizing different types of cell migration.

Single-cell shape features and clustering

A major challenge in the creation of every image analysis pipeline is identifying and measuring informative features. This search has a large iterative component and depends on the precise and accurate measurements of the relevant images. The extraction of 22 cell shape features was successfully implemented and characterises the cells controlled by eye well. Even though normalization was applied throughout the extracted cell shape features, they are not as comparable as for instance gene expression, for what UMAP was originally developed. No filtering steps or filtering out cells with least significant cell shape features for UMAP computation made a big

difference in cluster composition. The obtained clusters featured approximately even distributions of control knockdowns as well as all knockdowns, indicating that the morphological single-shape differences extracted out of the image data was not the basis for cluster annotation. To be able to draw conclusions on migratory behaviour of knockdowns based on single-cell shapes, the clustering presented here needs to be improved further. In that way characteristic clusters may be gained, each dominated by a control knockdown. In that way, knockdowns in the same annotated cluster as a control may indicate similar migratory behaviour which can be further aligned with information from the STRING database [73].

Furthermore, other clustering methods may be implemented. A VAE can be developed, using its latent space representation of the input data in order to draw conclusions on knockdowns expressing similar migratory behaviour. Utilizing that method, various kinds of input data, like extracted single-cell shape features, extracted scratch outlines, single-cell tracks or even raw images or image-tiles may be compared. What has to be kept in mind when using images as input data, is that the varying shape of the scratch between image stacks may be dominating the clustering approach.

For further research another tool called 'vampireanalysis' [74], a robust unsupervised machine-learning method to quantify the morphological heterogeneity of cells and nuclei, could be utilized to draw biological conclusions of migratory behaviour from the results.

Handling of large data sets

In order to handle large data sets and draw information out of them, calculations have to be optimized and may be performed on super computers or on the A100 cluster of the Humboldt-Universität of Berlin that is now accessible via Jupyter Notebooks. Especially the computations executed in order to generate single-cell trajectories and UMAPs would not be feasible on 'normal' university servers anymore when handling even larger image data sets. For instance, single-cell trajectory computation time for three experimental replicates of 384 image stacks and corresponding masks took approximately 76 hours on a 96 Cores/512GB RAM server.

Concluding, this thesis contributed to exploring the image data set provided by the Linding Lab [43] by implementing and assessing various image analysis techniques.

5 References

- [1] A. J. Ridley, M. A. Schwartz, K. Burridge, *et al.*, "Cell migration: Integrating signals from front to back," *Science*, vol. 302, 1704–1709, 2003. DOI: [10.1126/science.1092053](https://doi.org/10.1126/science.1092053).
- [2] Z. Xiong, G. Deng, X. Huang, *et al.*, "Bone metastasis pattern in initial metastatic breast cancer: A population-based study," *Cancer Management and Research*, vol. Volume 10, pp. 287–295, 2018. DOI: [10.2147/cmar.s155524](https://doi.org/10.2147/cmar.s155524).
- [3] E. Svensson, C. F. Christiansen, S. P. Ulrichsen, M. Rørth, and H. T. Sørensen, "Survival after bone metastasis by primary cancer type: A danish population-based cohort study," *BMJ Open*, vol. 7(9), e016022, 2017. DOI: [10.1136/bmjopen-2017-016022](https://doi.org/10.1136/bmjopen-2017-016022).
- [4] Y. Kasakura, M. Fujii, F. Mochizuki, T. Suzuki, and T. Takahashi, "Clinicopathological study of brain metastasis in gastric cancer patients," *Surgery Today*, vol. 30(6), pp. 485–490, 2000. DOI: [10.1007/s005950070112](https://doi.org/10.1007/s005950070112).
- [5] J. E. York, J. Stringer, J. A. Ajani, D. M. Wildrick, and Z. L. Gokaslan, "Gastric cancer and metastasis to the brain," *Annals of Surgical Oncology*, vol. 6(8), pp. 771–776, 1999. DOI: [10.1007/s10434-999-0771-3](https://doi.org/10.1007/s10434-999-0771-3).
- [6] W. K. Kang, J. K. Lee, S. T. Oh, S. H. Lee, and C. K. Jung, "Stromal expression of mir-21 in t3-4a colorectal cancer is an independent predictor of early tumor relapse," *BMC Gastroenterology*, vol. 15(1), 2015. DOI: [10.1186/s12876-015-0227-0](https://doi.org/10.1186/s12876-015-0227-0).
- [7] C. Chen and X. Zhang, "Evaluation of ptprt mutations as biomarkers for cancer metastasis across multiple cancer types," 2020. DOI: [10.21203/rs.3.rs-16367/v1](https://doi.org/10.21203/rs.3.rs-16367/v1).
- [8] W. Sun, S. Huang, Y. Zu, and Y. Deng, "Klf3 is a crucial regulator of metastasis by controlling stat3 expression in lung cancer," *Molecular Carcinogenesis*, vol. 58(11), pp. 1933–1945, 2019. DOI: [10.1002/mc.23072](https://doi.org/10.1002/mc.23072).
- [9] M. Kohno and J. Pouyssegur, "Targeting the erk signaling pathway in cancer therapy," *Annals of Medicine*, vol. 38, 200–211, 2006. DOI: [10.1080/07853890600551037](https://doi.org/10.1080/07853890600551037).
- [10] I. J. Fidler, "The pathogenesis of cancer metastasis: The 'seed and soil' hypothesis revisited," *Nat Rev Cancer*, vol. 3, 453–458, 2003. DOI: [10.1038/nrc1098](https://doi.org/10.1038/nrc1098).
- [11] A. W. Lambert, D. R. Pattabiraman, and R. A. Weinberg, "Emerging biological principles of metastasis," *Cell*, vol. 168, 670–691, 2017. DOI: [10.1016/j.cell.2016.11.037](https://doi.org/10.1016/j.cell.2016.11.037).
- [12] J. Fares, M. Y. Fares, H. H. Khachfe, H. A. Salhab, and Y. Fares, "Molecular principles of metastasis: A hallmark of cancer revisited," *Sig Transduct Target Ther*, vol. 5 (28), 2020. DOI: [10.1038/s41392-020-0134-x](https://doi.org/10.1038/s41392-020-0134-x).
- [13] S. Bakhoum, B. Ngo, and A. L. et al., "Chromosomal instability drives metastasis through a cytosolic dna response," *Nature*, vol. 553, 467–472, 2018. DOI: [10.1038/nature25432](https://doi.org/10.1038/nature25432).

- [14] R. Kalluri and R. A. Weinberg, "The basics of epithelial-mesenchymal transition," *J Clin Invest*, vol. 119(6):1420–1428, 2009. DOI: [10.1172/JCI39104](https://doi.org/10.1172/JCI39104).
- [15] W. Giese, J. P. Albrecht, O. Oppenheim, *et al.*, "Polarity-jam: An image analysis toolbox for cell polarity, junction and morphology quantification," *bioRxiv*, 2024. DOI: [10.1101/2024.01.24.577027](https://doi.org/10.1101/2024.01.24.577027).
- [16] C.-C. Liang, A. Y. Park, and J.-L. Guan, "*In vitro* Scratch assay: A convenient and inexpensive method for analysis of cell migration *in vitro*," *Nat Protoc*, vol. 2, 329–333, 2007. DOI: [10.1038/nprot.2007.30](https://doi.org/10.1038/nprot.2007.30).
- [17] K. Tanaka, D. Joshi, S. Timalsina, and M. A. Schwartz, "Early events in endothelial flow sensing," *Cytoskeleton*, vol. 78(6):217–231, 2021. DOI: [10.1002/cm.21652](https://doi.org/10.1002/cm.21652).
- [18] P. Friedl and K. Wolf, "Tumour-cell invasion and migration: Diversity and escape mechanisms," *Nat Rev Cancer*, vol. 3, 362–374, 2003. DOI: [10.1038/nrc1075](https://doi.org/10.1038/nrc1075).
- [19] J. E. Sero, H. Z. Sailem, R. C. Ardy, H. Almuttaqi, T. Zhang, and C. Bakal, "Cell shape and the microenvironment regulate nuclear translocation of nf- κ b in breast epithelial and tumor cells," *Mol Syst Biol*, vol. 11, 790, 2015. DOI: [10.15252/msb.20145644](https://doi.org/10.15252/msb.20145644).
- [20] J. Fares, M. Y. Fares, H. H. Khachfe, H. A. Salhab, and Y. Fares, "Molecular principles of metastasis: A hallmark of cancer revisited," *Sig Transduct Target Ther*, vol. 5, 28, 2020. DOI: [10.1038/s41392-020-0134-x](https://doi.org/10.1038/s41392-020-0134-x).
- [21] K. M. Yamada, "Immunological characterization of a major transformation-sensitive fibroblast cell surface glycoprotein. localization, redistribution, and role in cell shape..," *The Journal of Cell Biology*, vol. 78(2), pp. 520–541, 1978. DOI: [10.1083/jcb.78.2.520](https://doi.org/10.1083/jcb.78.2.520).
- [22] R. J. Petrie and K. M. Yamada, "At the leading edge of three-dimensional cell migration," *J Cell Sci*, vol. 125(24), pp. 917–5926, 2012. DOI: [10.1242/jcs.093732](https://doi.org/10.1242/jcs.093732).
- [23] D. Hong, T. L. Messier, C. E. Tye, *et al.*, "Runx1 stabilizes the mammary epithelial cell phenotype and prevents epithelial to mesenchymal transition," *Oncotarget*, vol. 8(11), pp. 17 610–17 627, 2017. DOI: [10.18632/oncotarget.15381](https://doi.org/10.18632/oncotarget.15381).
- [24] Z. Mousavikhamene, D. J. Sykora, M. Mrksich, and N. Bagheri, "Morphological features of single cells enable accurate automated classification of cancer from non-cancer cell lines," *Sci Rep*, vol. 11, 24375, 2021. DOI: [10.1038/s41598-021-03813-8](https://doi.org/10.1038/s41598-021-03813-8).
- [25] J. Konen, D. J. Brat, A. I. Marcus, and L. Cooper, "Tasi: A software tool for spatial-temporal quantification of tumor spheroid dynamics," *Scientific Reports*, vol. 8(1), 2018. DOI: [10.1038/s41598-018-25337-4](https://doi.org/10.1038/s41598-018-25337-4).
- [26] K. I. Hulkower and R. L. Herber, "Cell migration and invasion assays as tools for drug discovery," *Pharmaceutics*, vol. 3(1), 107–124, 2011. DOI: [10.3390/pharmaceutics3010107](https://doi.org/10.3390/pharmaceutics3010107).

- [27] G. Pan, X. Wang, Y. Wang, *et al.*, “Helicobacter pylori promotes gastric cancer progression by upregulating semaphorin 5a expression via erk/mmp9 signaling,” *Molecular Therapy - Oncolytics*, vol. 22, pp. 256–264, 2021. DOI: [10.1016/j.omto.2021.06.002](https://doi.org/10.1016/j.omto.2021.06.002).
- [28] S. Back and J. J. Manfredi, “Knockdown of target genes by sirna in vitro,” *Methods Mol Biol.*, vol. 2267, 159–163, 2021. DOI: [10.1007/978-1-0716-1217-0_10](https://doi.org/10.1007/978-1-0716-1217-0_10).
- [29] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, “A programmable dual-rna-guided dna endonuclease in adaptive bacterial immunity,” *Cytoskeleton*, vol. 337(6096), pp. 816–821, 2012. DOI: [10.1126/science.1225829](https://doi.org/10.1126/science.1225829).
- [30] M. T. M. Michael Boettcher, “Choosing the right tool for the job: Rnai, talen, or crispr,” *Molecular Cell Review*, vol. 58, pp. 575–585, 2015. DOI: [10.1016/j.molcel.2015.04.028](https://doi.org/10.1016/j.molcel.2015.04.028).
- [31] Q. Yan, G. Lou, Y. Qian, *et al.*, “Spag9 is involved in hepatocarcinoma cell migration and invasion via modulation of elk1 expression,” *Oncotargets and Therapy*, p. 1067, 2016. DOI: [10.2147/ott.s98727](https://doi.org/10.2147/ott.s98727).
- [32] Q. Wang, C. Yang, Z. Ma, *et al.*, “Inhibition of prostate cancer du145 cell growth with small interfering rna targeting the satb1 gene,” *Experimental and Therapeutic Medicine*, 2018. DOI: [10.3892/etm.2018.5792](https://doi.org/10.3892/etm.2018.5792).
- [33] J. E. Zuckerman, T. Hsueh, R. C. Koya, M. E. Davis, and A. Ribas, “Sirna knockdown of ribonucleotide reductase inhibits melanoma cell line proliferation alone or synergistically with temozolomide,” *Journal of Investigative Dermatology*, vol. 131(2), pp. 453–460, 2011. DOI: [10.1038/jid.2010.310](https://doi.org/10.1038/jid.2010.310).
- [34] H. Huang, Z. Jiang, Q. Li, *et al.*, “Inhibition of human breast cancer cell invasion by sirna against urokinase-type plasminogen activator,” *Cancer Investigation*, vol. 28(7), pp. 689–697, 2010. DOI: [10.3109/07357901003735642](https://doi.org/10.3109/07357901003735642).
- [35] M. S. Schwartz, E. Moen, G. Miller, *et al.*, “Caliban: Accurate cell tracking and lineage construction in live-cell imaging experiments with deep learning,” *bioRxiv*, 2023. DOI: [10.1101/803205](https://doi.org/10.1101/803205).
- [36] D. A. V. Valen, T. Kudo, K. Lane, *et al.*, “Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments,” *PLOS Computational Biology*, vol. 12(11), e1005177, 2016. DOI: [10.1371/journal.pcbi.1005177](https://doi.org/10.1371/journal.pcbi.1005177).
- [37] M. H. Lee, P. H. Wu, J. R. Staunton, R. Ros, G. D. Longmore, and D. Wirtz, “Mismatch in mechanical and adhesive properties induces pulsating cancer cell migration in epithelial monolayer,” *Biophysical Journal*, vol. 102(12), pp. 2731–2741, 2012. DOI: [10.1016/j.bpj.2012.05.005](https://doi.org/10.1016/j.bpj.2012.05.005).
- [38] J. Schindelin, I. Arganda-Carreras, E. Frise, *et al.*, “Fiji: An open-source platform for biological-image analysis,” *Nat Methods*, vol. 9, 676–682, 2012. DOI: [10.1038/nmeth.2019](https://doi.org/10.1038/nmeth.2019).
- [39] J. Chalfoun, M. Majurski, and T. Blattner, “Mist: Accurate and scalable microscopy image stitching tool with stage modeling and error minimization,” *Sci Rep*, vol. 7, 4988, 2017. DOI: [10.1038/s41598-017-04567-y](https://doi.org/10.1038/s41598-017-04567-y).

- [40] R. F. Laine, K. L. Tosheva, N. Gustafsson, and R. D. M. Gray, “Nanoj: A high-performance open-source super-resolution microscopy toolbox,” *J Phys D Appl Phys*, vol. 52(16): 163001, 2019. DOI: [10.1088/1361-6463/ab0261](https://doi.org/10.1088/1361-6463/ab0261).
- [41] M. Pachitariu and C. Stringer, “Cellpose 2.0: How to train your own model,” *Nat Methods*, vol. 19, 1634–1641, 2022. DOI: [10.1038/s41592-022-01663-4](https://doi.org/10.1038/s41592-022-01663-4).
- [42] N. F. Greenwald, G. Miller, E. Moen, *et al.*, “Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning,” *Nat Biotechnol*, vol. 40, 555–565, 2020. DOI: [10.1038/s41587-021-01094-0](https://doi.org/10.1038/s41587-021-01094-0).
- [43] G. Sağınç, J. Longden, R. Linding, and et al., “Dataset: High-content screening experiment of wound healing assay of mda-mb-231 cells,” *in preparation*, 2016.
- [44] S. Johnston, M. J. Simpson, and D. L. S. McElwain, “How much information can be obtained from tracking the position of the leading edge in a scratch assay?” *Journal of the Royal Society Interface*, vol. 11(97), p. 20140325, 2014. DOI: [10.1098/rsif.2014.0325](https://doi.org/10.1098/rsif.2014.0325).
- [45] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, *et al.*, “Scikit-image: Image processing in python,” *ArXiv e-prints*, 2014. DOI: [10.48550/arXiv.1407.6245](https://doi.org/10.48550/arXiv.1407.6245).
- [46] C. Z. Eddy, H. Raposo, A. Manchanda, R. Wong, F. Li, and B. Sun, “Morphodynamics facilitate cancer cells to navigate 3d extracellular matrix,” *Sci Rep*, vol. 11, 20434, 2021. DOI: [10.1038/s41598-021-99902-9](https://doi.org/10.1038/s41598-021-99902-9).
- [47] B. Möller, Y. Poeschl, R. Plötner, and K. Bürstenbinder, “Pacequant: A tool for high-throughput quantification of pavement cell shape characteristics,” *Plant Physiology*, vol. 175, 998–1017, 2017. DOI: [10.1038/s41598-021-03813-8](https://doi.org/10.1038/s41598-021-03813-8).
- [48] M. B. Woodworth, K. M. Girsakis, and C. A. Walsh, “Building a lineage from single cells: Genetic techniques for cell lineage tracking,” *Nature Reviews Genetics*, vol. 18(4), pp. 230–244, 2017. DOI: [10.1038/nrg.2016.159](https://doi.org/10.1038/nrg.2016.159).
- [49] G. Didier and K. Zhang, “The asymptotic distribution of the pathwise mean squared displacement in single particle tracking experiments,” *Journal of Time Series Analysis*, vol. 38, pp. 395–416, 3 2016. DOI: [10.1111/jtsa.12208](https://doi.org/10.1111/jtsa.12208).
- [50] P. Allegrini, J. F. Douglas, and S. C. Glotzer, “Dynamic entropy as a measure of caging and persistent particle motion in supercooled liquids,” *Physical Review E*, vol. 60, pp. 5714–5724, 5 1999. DOI: [10.1103/physreve.60.5714](https://doi.org/10.1103/physreve.60.5714).
- [51] É. Becht, L. McInnes, J. J. Healy, *et al.*, “Dimensionality reduction for visualizing single-cell data using t-SNE,” *Nature Biotechnology*, vol. 37, pp. 38–44, 1 2018. DOI: [10.1038/nbt.4314](https://doi.org/10.1038/nbt.4314).

- [52] J. Tenenbaum, V. d. Silva, and J. Langford, “A global geometric framework for nonlinear dimensionality reduction,” *Science*, vol. 290(5500), pp. 2319–2323, 2000. DOI: [10.1126/science.290.5500.2319](https://doi.org/10.1126/science.290.5500.2319).
- [53] I. T. Jolliffe and J. Cadima, “Principal component analysis: A review and recent developments,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374(2065), p. 20150202, 2016. DOI: [10.1098/rsta.2015.0202](https://doi.org/10.1098/rsta.2015.0202).
- [54] C. X. Hernández, H. K. Wayment-Steele, M. M. Sultan, B. E. Husic, and V. S. Pande, “Variational encoding of complex dynamics,” *Physical Review E*, vol. 97, 6 2018. DOI: [10.1103/physreve.97.062412](https://doi.org/10.1103/physreve.97.062412).
- [55] V. A. Traag, L. Waltman, and N. J. van Eck, “From louvain to leiden: Guaranteeing well-connected communities,” *Sci Rep*, vol. 9, 5233, 2019. DOI: [10.1038/s41598-019-41695-z](https://doi.org/10.1038/s41598-019-41695-z).
- [56] S. Chandrasekharan, M. Zaka, S. A. Gallo, *et al.*, “Finding scientific communities in citation graphs: Articles and authors,” *Quantitative Science Studies*, vol. 2(1), pp. 184–203, 2021. DOI: [10.1162/qss_a_00095](https://doi.org/10.1162/qss_a_00095).
- [57] R. Shemirani, G. M. Belbin, K. Burghardt, *et al.*, “Selecting clustering algorithms for ibd mapping,” 2021. DOI: [10.1101/2021.08.11.456036](https://doi.org/10.1101/2021.08.11.456036).
- [58] A. Sabt, W. M. Eldehna, T. Al-Warhi, *et al.*, “Discovery of 3,6-disubstituted pyridazines as a novel class of anticancer agents targeting cyclin-dependent kinase 2: Synthesis, biological evaluation and in silico insights,” *Journal of Enzyme Inhibition and Medicinal Chemistry*, vol. 35(1), pp. 1616–1630, 2020. DOI: [10.1080/14756366.2020.1806259](https://doi.org/10.1080/14756366.2020.1806259).
- [59] J. E. Price, A. Polyzos, R. D. Zhang, and L. M. Daniels, “Tumorigenicity and metastasis of human breast carcinoma cell lines in nude mice,” *Cancer Res*, vol. 50(3), 717–21, 1990. DOI: [PMID:2297709](https://pubmed.ncbi.nlm.nih.gov/2297709/).
- [60] T. Zhou, L. Yong, L. Yang, L. Liu, Y. Ju, and C. Li, “Silencing of anxa3 expression by rna interference inhibits the proliferation and invasion of breast cancer cells,” *Oncology Reports*, vol. 37(1), pp. 388–398, 2016. DOI: [10.3892/or.2016.5251](https://doi.org/10.3892/or.2016.5251).
- [61] S. Even-Ram, A. D. Doyle, M. A. Conti, K. Matsumoto, R. S. Adelstein, and K. M. Yamada, “Myosin iia regulates cell motility and actomyosin–microtubule crosstalk,” *Nat Cell Biol*, vol. 9, 299–309, 2007. DOI: [10.1038/ncb1540](https://doi.org/10.1038/ncb1540).
- [62] Y.-h. Xin, B.-s.-j. Bian, X.-j. Yang, *et al.*, “Pou5f1 enhances the invasiveness of cancer stem-like cells in lung adenocarcinoma by upregulation of mmp-2 expression,” *PLoS ONE*, vol. 8(12), e83373, 2013. DOI: [10.1371/journal.pone.0083373](https://doi.org/10.1371/journal.pone.0083373).
- [63] Z. Liu, Q. Sun, and X. Wang, “Plk1, a potential target for cancer therapy,” *Translational Oncology*, vol. 10(1), 22-32, 2017. DOI: [10.1016/j.tranon.2016.10.003](https://doi.org/10.1016/j.tranon.2016.10.003).

- [64] M. Engel, J. Longden, J. Ferkinghoff-Borg, X. Robin, G. Saginc, and R. Linding, “Bowhead: Bayesian modelling of cell velocity during concerted cell migration,” *PLoS Comput Biol*, vol. 14(1): e1005900, 2018. DOI: [10.1371/journal.pcbi.1005900](https://doi.org/10.1371/journal.pcbi.1005900).
- [65] C. Bakal, J. Aach, G. Church, and N. Perrimon, “Quantitative morphological signatures define local signaling networks regulating cell morphology,” *Science*, vol. 316, 5832, 2007. DOI: [10.1126/science.1140324](https://doi.org/10.1126/science.1140324).
- [66] P. Virtanen, R. Gommers, T. E. Oliphant, *et al.*, “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python,” *Nature Methods*, vol. 17, pp. 261–272, 2020. DOI: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).
- [67] S. Gillies *et al.*, *Shapely: Manipulation and analysis of geometric objects*, *toblerity.org*, 2007–. [Online]. Available: <https://github.com/Toblerity/Shapely>.
- [68] A. Ahmadi, F. Boulogne, R. P. Zoey S. Davidson, *et al.*, “Trackpy 0.6.2,” 2024. DOI: [10.5281/zenodo.1213240](https://doi.org/10.5281/zenodo.1213240).
- [69] D. Ershov, M.-S. Phan, J. W. Pylvänäinen, *et al.*, “Trackmate 7: Integrating state-of-the-art segmentation algorithms into tracking pipelines,” *Nat Methods*, vol. 19, 829–832, 2022. DOI: [10.1038/s41592-022-01507-1](https://doi.org/10.1038/s41592-022-01507-1).
- [70] I. M. Wortel, A. Y. Liu, K. Dannenberg, J. C. Berry, M. J. Miller, and J. Textor, “Celltrackr: An r package for fast and flexible analysis of immune cell migration data,” *ImmunoInformatics*, vol. 1–2, 100003, 2021. DOI: [10.1016/j.immuno.2021.100003](https://doi.org/10.1016/j.immuno.2021.100003).
- [71] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [72] F. Wolf, P. Angerer, and F. Theis, “Scanpy: Large-scale single-cell gene expression data analysis,” *Genome Biol*, vol. 19, 15, 2018. DOI: [10.1186/s13059-017-1382-0](https://doi.org/10.1186/s13059-017-1382-0).
- [73] D. Szklarczyk, R. Kirsch, M. Koutrouli, *et al.*, “The string database in 2023: Protein-protein association networks and functional enrichment analyses for any sequenced genome of interest,” *Nucleic Acids Res*, vol. 6;51(D1):D638–D646, 2023. DOI: [10.1093/nar/gkac1000](https://doi.org/10.1093/nar/gkac1000).
- [74] J. M. Phillip, K.-S. Han, W.-C. Chen, D. Wirtz, and P.-H. Wu, “A robust unsupervised machine-learning method to quantify the morphological heterogeneity of cells and nuclei,” *Nat Protoc*, vol. 16, 754–774, 2021. DOI: [10.1038/s41596-020-00432-x](https://doi.org/10.1038/s41596-020-00432-x).

6 Acknowledgments

I am sincerely grateful to Prof. Dr. Dr. h.c. Edda Klipp for the ability to work in her group on my bachelor project and thesis. Throughout the entire journey the close collaborative work was inspiring and I could rely on invaluable guidance, support, and encouragement. Her expertise, patience and insightful feedback have been instrumental in shaping this work.

To Prof. Dr. Uwe Ohler I express my sincere gratitude for agreeing to be the second supervisor of my thesis without any hesitation.

Furthermore, I would like to thank Dr. Rune Linding for providing the data set analysed in this thesis, which enabled me to pursue this research.

A special mention is reserved for Lukas Kiwitz. It is absolutely no bold claim that his ideas and support were essential right from the beginning, where not even image displaying worked, to the end, where he needed to stop me from further trying to get new things to work. I am very thankful for his time, dedication and expertise. Also, I am confident that he will be Bundeskanzler some day.

Also, I would like to express my thanks to Dominik Kranz of the Cardiovascular Physics AG for all of his awesome project ideas and his time.

Moreover, I am deeply thankful to my study group members, especially Anna Truckenbrodt, Jorin Diemer, Gabriele Schreiber, Hanno Fandrich and Patrick Brunner, for their inspiring distractions, collaborative mood, lectures on understanding irony and shared insights that have enriched my understanding of and passion for the topic enormously.

Furthermore, I would like to thank my parents, my brother and my grandparents. Your support gives me the possibility to fulfill my dreams and become a scientist.

7 Appendices

7.1 Fiji macro for image stitching and drift-correction

For experimental replicate 2 this macro takes 3 .flex files per timepoint (n=17) and knockdown (n=384) and stitches them using MIST. After that it creates one stack and then one hyperstack per knockdown (n=384). Afterwards it performs drift-correction on every hyperstack.

```

1 // Experiment 2: started June 22, 2017 at 19:53:39
2 // This macro takes 3 files per timepoint (17) and knockdown (384) and stitches them using MIST. After that it creates one stack and then one hyperstack per knockdown (384). Afterwards it performs dr
3 // The list containing the knockdown numbers was created using Python and pasted into the macro script.
4
5 list = newArrayList("001001", "001002", "001003", "001004", "001005", "001006", "001007", "001008", "001009", "001010", "001011", "001012", "001013", "001014", "001015", "001016", "001017", "001018", "001019", "001020", "001021"
6
7 path_input = '/home/basar/Olyssa/Exp2_2017_06_22/'
8 path_output = '/home/basar/Olyssa/exp2'
9 path_done = '/home/basar/Olyssa/exp2_done/'
10
11 for (k=0; k<list.length; k++) {
12     for (j=0; j<17; j++) {
13         // import image sub-title 001 (mid), split into the two color channels C=0 (nuc) and C=1 (cyto), enhance image contrast, merge both channels, transform the nuc-cyto stack to RGB and save as 002
14         run("Bio-Formats Importer", "open=" + path_input + list[k] + "001_" + j + ".flex autoscale color_mode=Default rois_import=[ROI manager] split_channels view=Hyperstack stack_order=XYCZT");
15         selectImage("+" + list[k] + "001_" + j + ".flex", C=0");
16         run("Enhance Contrast", "saturated=0.35");
17         selectImage("+" + list[k] + "001_" + j + ".flex", C=1");
18         run("Enhance Contrast", "saturated=0.35");
19         run("Merge Channels...", "C1=[" + list[k] + "001_" + j + ".flex - C=1] C2=[" + list[k] + "001_" + j + ".flex - C=0] create");
20         run("Stack to RGB");
21         saveAs("Tiff", path_output + list[k] + "002_" + j + "_RGB.tif");
22         run("Close All");
23
24         // import image sub-title 002 (left), split into the two color channels C=0 (nuc) and C=1 (cyto), enhance image contrast, merge both channels, transform the nuc-cyto stack to RGB and save as 001
25         run("Bio-Formats Importer", "open=" + path_input + list[k] + "002_" + j + ".flex autoscale color_mode=Default rois_import=[ROI manager] split_channels view=Hyperstack stack_order=XYCZT");
26         selectImage("+" + list[k] + "002_" + j + ".flex", C=0");
27         run("Enhance Contrast", "saturated=0.35");
28         selectImage("+" + list[k] + "002_" + j + ".flex", C=1");
29         run("Enhance Contrast", "saturated=0.35");
30         run("Merge Channels...", "C1=[" + list[k] + "002_" + j + ".flex - C=1] C2=[" + list[k] + "002_" + j + ".flex - C=0] create");
31         run("Stack to RGB");
32         saveAs("Tiff", path_output + list[k] + "001_" + j + "_RGB.tif");
33         run("Close All");
34
35         // import image sub-title 003 (right), split into the two color channels C=0 (nuc) and C=1 (cyto), enhance image contrast, merge both channels, transform the nuc-cyto stack to RGB and save as 003
36         run("Bio-Formats Importer", "open=" + path_input + list[k] + "003_" + j + ".flex autoscale color_mode=Default rois_import=[ROI manager] split_channels view=Hyperstack stack_order=XYCZT");
37         selectImage("+" + list[k] + "003_" + j + ".flex - C=0");
38         run("Enhance Contrast", "saturated=0.35");
39         selectImage("+" + list[k] + "003_" + j + ".flex - C=1");
40         run("Enhance Contrast", "saturated=0.35");
41         run("Merge Channels...", "C1=[" + list[k] + "003_" + j + ".flex - C=1] C2=[" + list[k] + "003_" + j + ".flex - C=0] create");
42         run("Stack to RGB");
43         saveAs("Tiff", path_output + list[k] + "003_" + j + "_RGB.tif");
44         run("Close All");
45
46         // execute image stitching with the tool 'MIST' (Microscopy Image Stitching Tool)
47         run("MIST", "gridwidth=3 gridheight=1 starttile=1 imagedir=/home/basar/Olyssa/exp2 filenamepattern=" + list[k] + "00(p)_" + j + ".RGB.tif filenamepatterntype=SEQUENTIAL gridorigin=UL assemblefromm
48
49         // save stitched image
50         saveAs("Tiff", path_output + list[k] + "_" + j + "_fused_MIST.tif");
51         run("Close All");
52     }

```

(a)

```

53 // opening stitched images of one time series
54 run("Close All");
55 for (i=0; i<17; i++) {
56   run("Bio-Formats Importer", "open=" + path_output + list[k] + "_" + i + "_fused_MIST.tif autoscale color_mode=Composite rois_import=[ROI manager] view=Hyperstack stack_order=XYCZT");
57 }
58 // concatenate the time series images, transform to hyperstack and save the hyperstack
59 run("Concatenate...", "open=image1=" + list[k] + "_0_fused_MIST.tif image2=" + list[k] + "_1_fused_MIST.tif image3=" + list[k] + "_2_fused_MIST.tif image4=" + list[k] + "_3_fused_MIST.tif image5=" + list[k] + '_4_fused_MIST.tif" stack_order=XYCZT");
60 run("Stack to Hyperstack...", "order=xyzt(default) channels=3 slices=1 frames=17 display=Composite");
61 saveAs("Tif", path_output + list[k] + "_hyperstack_MIST.tif");
62 run("Close All");
63
64 // open the hyperstack
65 run("Bio-Formats Importer", "open=" + path_output + list[k] + "_hyperstack_MIST.tif autoscale color_mode=Colorized rois_import=[ROI manager] split_channels view=Hyperstack stack_order=XYCZT");
66 selectImage("0" + list[k] + "_hyperstack_MIST.tif - C<2");
67 close();
68 // perform drift estimation and correction using the channel C=0, save the nuc-channel time series image stack
69 selectImage("0" + list[k] + "_hyperstack_MIST.tif - C<0");
70 run("Estimate Drift", "times=max0 reference=previous frame (better for live) show_cross-correlation show_drift_plot show_drift_table apply choose=" + path_output + list[k] + "_hyperstack_MIST_C=0.njt");
71 saveAs("Tif", path_done + list[k] + "_hyperstack_drift_corrected_nuc_MIST.tif");
72 close();
73 selectImage("0" + list[k] + "_hyperstack_MIST.tif - C<0");
74 close();
75 // select the cyto-channel time series image stack and perform the same drift correction as applied to the nuc-channel stack, save cyto-channel image stack
76 selectImage("0" + list[k] + "_hyperstack_MIST.tif - C<1");
77 run("Correct Drift", "choose=" + path_output + list[k] + "_hyperstack_MIST_C=0DriftTable.njt");
78 saveAs("Tif", path_done + list[k] + "_hyperstack_drift_corrected_cyto_MIST.tif");
79
80 run("Close All");
81
82
83
84
85 //other stitching plugin:
86
87 //run("Grid/Collection stitching", "type=[Grid: row-by-row] order=[Right & Down
88 //saveAs("Tif", path_output + list[k] + "_" + j + "_RGB_stitched.tif");
89 //run("Close All");

```

(b)

Figure 23: Script of Fiji macro for image stitching with MIST and drift-correction with NanoJ-Core.

7.2 Own cell segmentation trained onto model 'cyto' from Cellpose 2.0

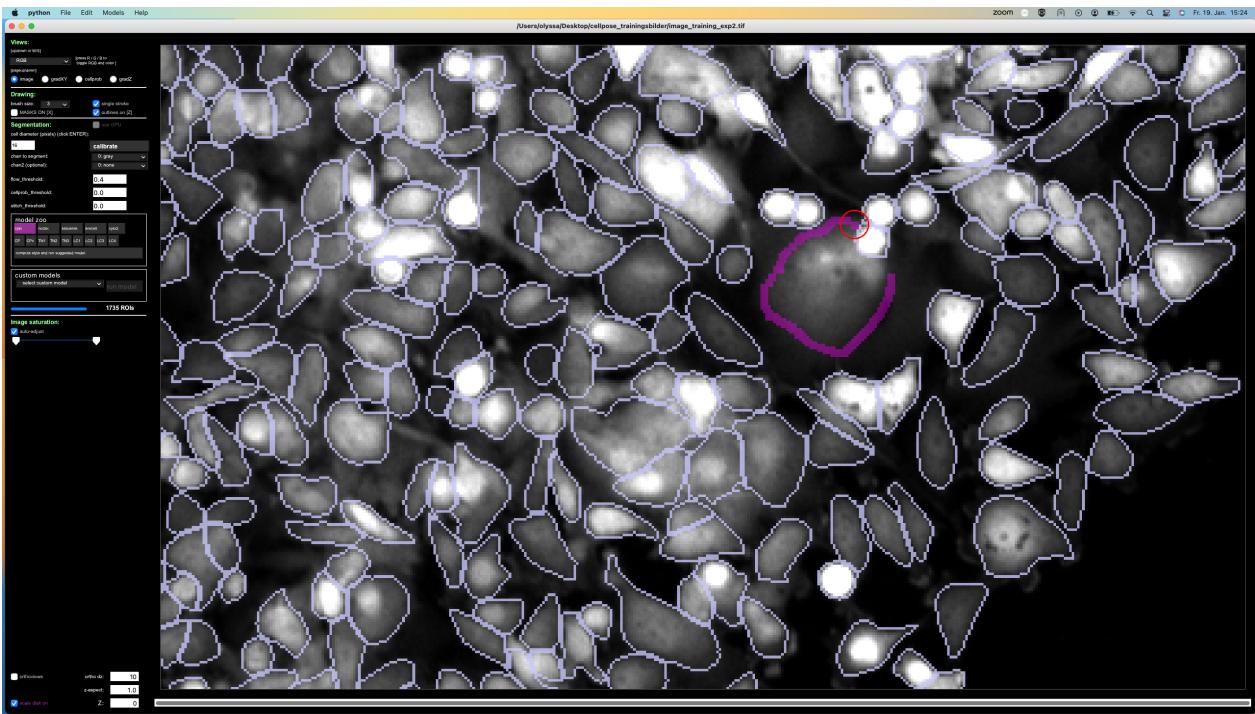


Figure 24: Correcting cell segmentations generated by model 'cyto' from Cellpose 2.0 and thereby creating a data set for training on some random images from own data in order to create a segmentation model adapted to own data.

7.3 Scratch detection algorithm fail case

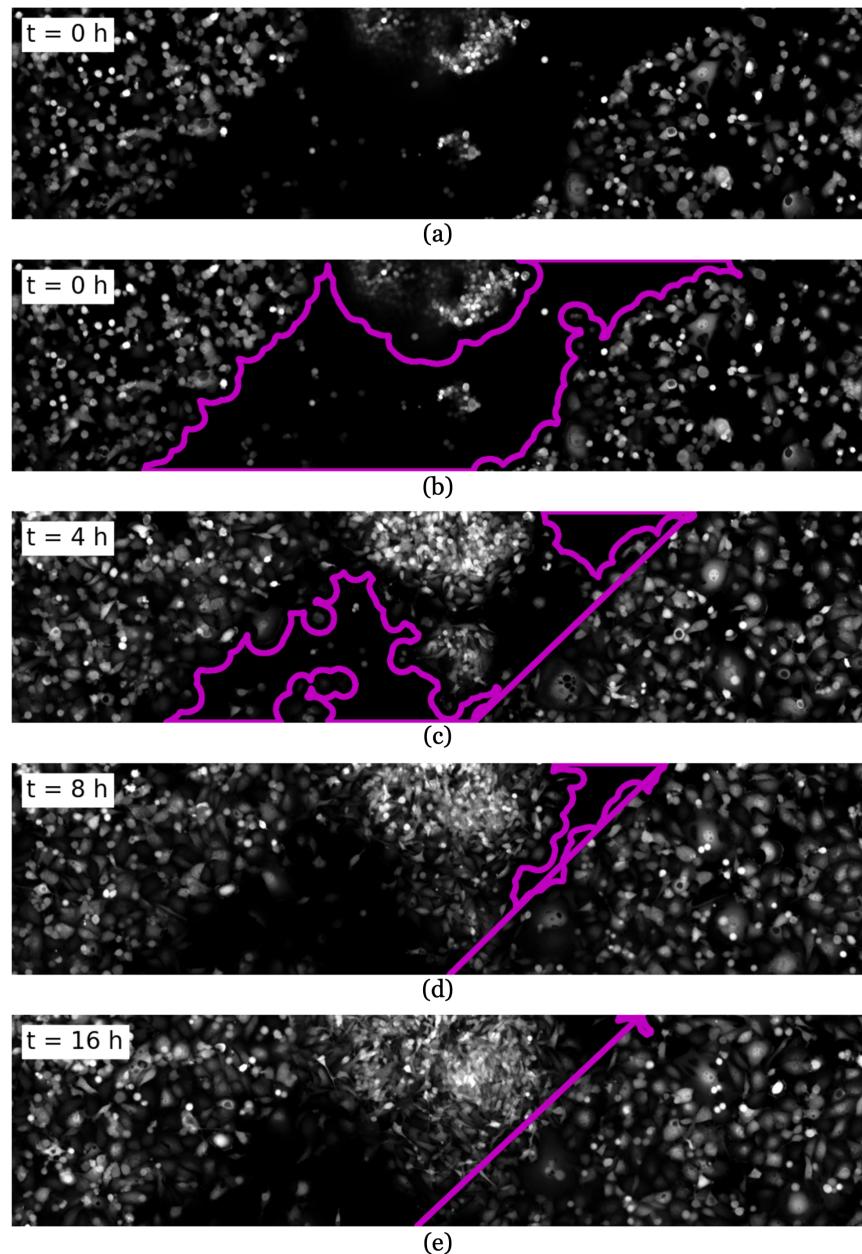


Figure 25: Case of failing of the automatic scratch detection due to not enough robustness of the algorithm when employed on a scratch with cells contained inside the scratch area at $t=0$ h.

7.4 Distance fields

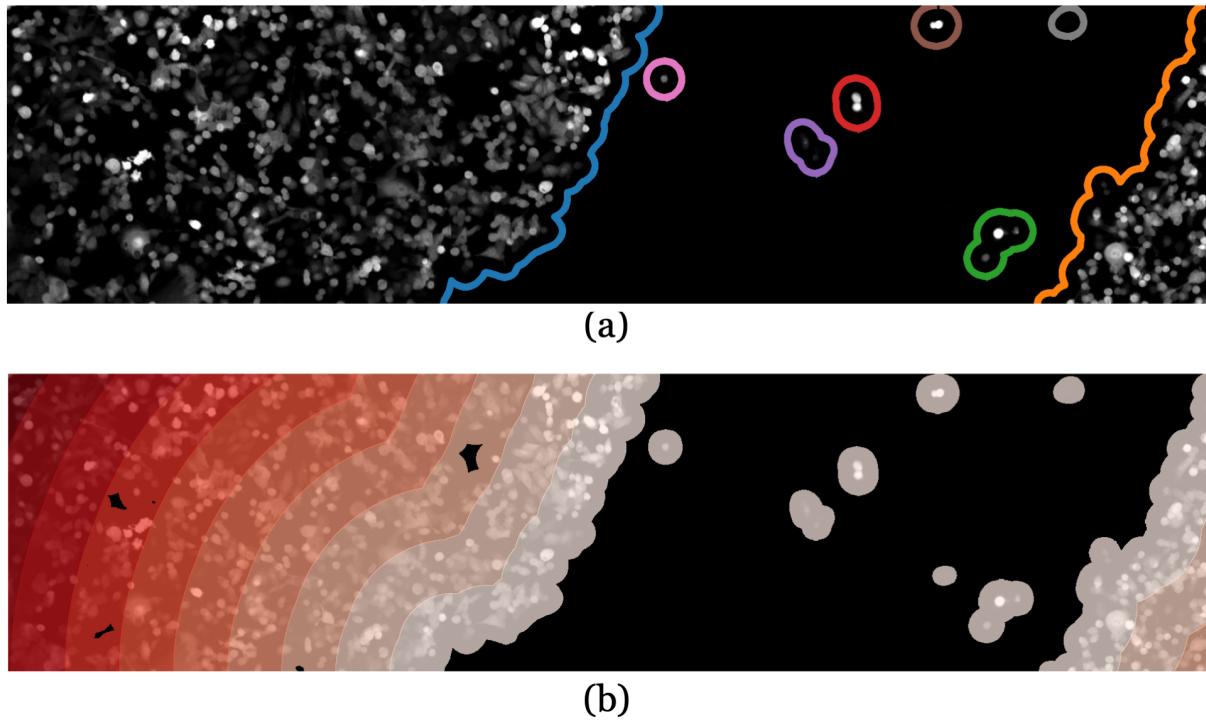


Figure 26: Comparison of positional feature extraction methods. (a) Results from the in this thesis used scratch detection algorithm are showed. (b) Distance fields colored by distance to the scratch. The distance field was computed with the Python package `skfmm` and depicts another approach to categorize leading edge cells and bulk cells, especially for untidy scratches.

7.5 Nuclei displacement polarity angle

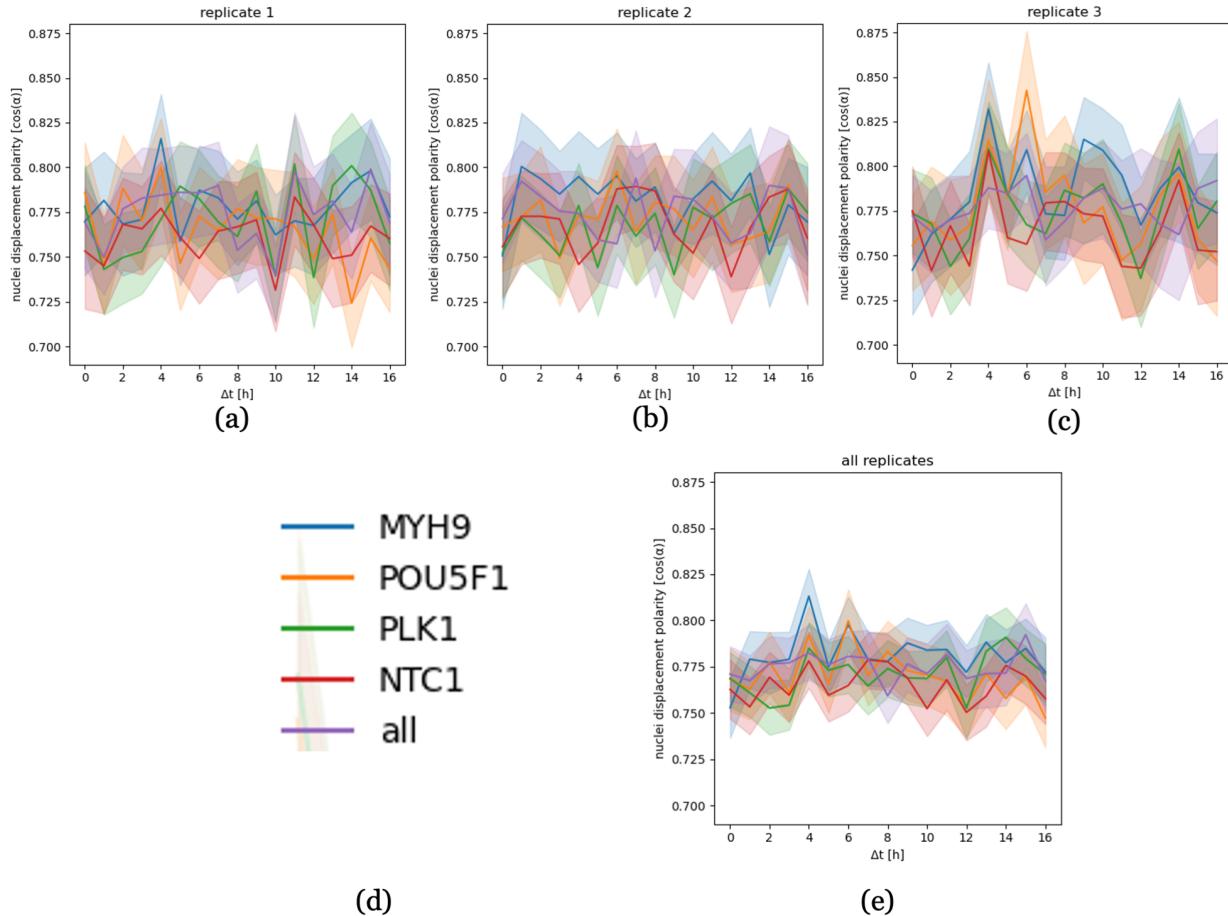


Figure 27: Nuclei displacement polarity angle of leading edge cells towards perpendicular bisector of the scratch, displayed for the control knockdowns MYH9, PLK1, POU5F1 and NTC1 and for all knockdowns over time. (a)-(c) depict the evolution of the nuclei displacement polarity angle for the three experimental replicates, (d) shows the legend for the different knockdowns and (e) displays the angle evolution for all experimental replicates combined. All graphs are plotted with a confidence interval width of 95 %.

7.6 Explanation of single-cell shape feature measures

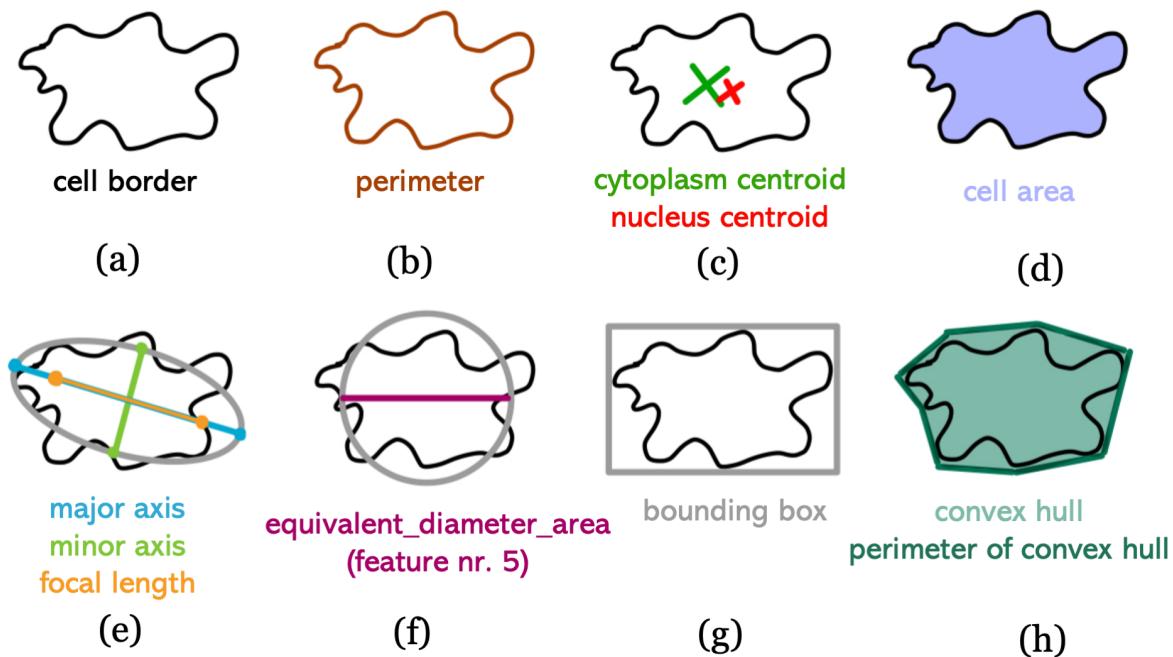


Figure 28: Explanation of measures for calculation of cell shape features.

7.7 Single-cell shape features

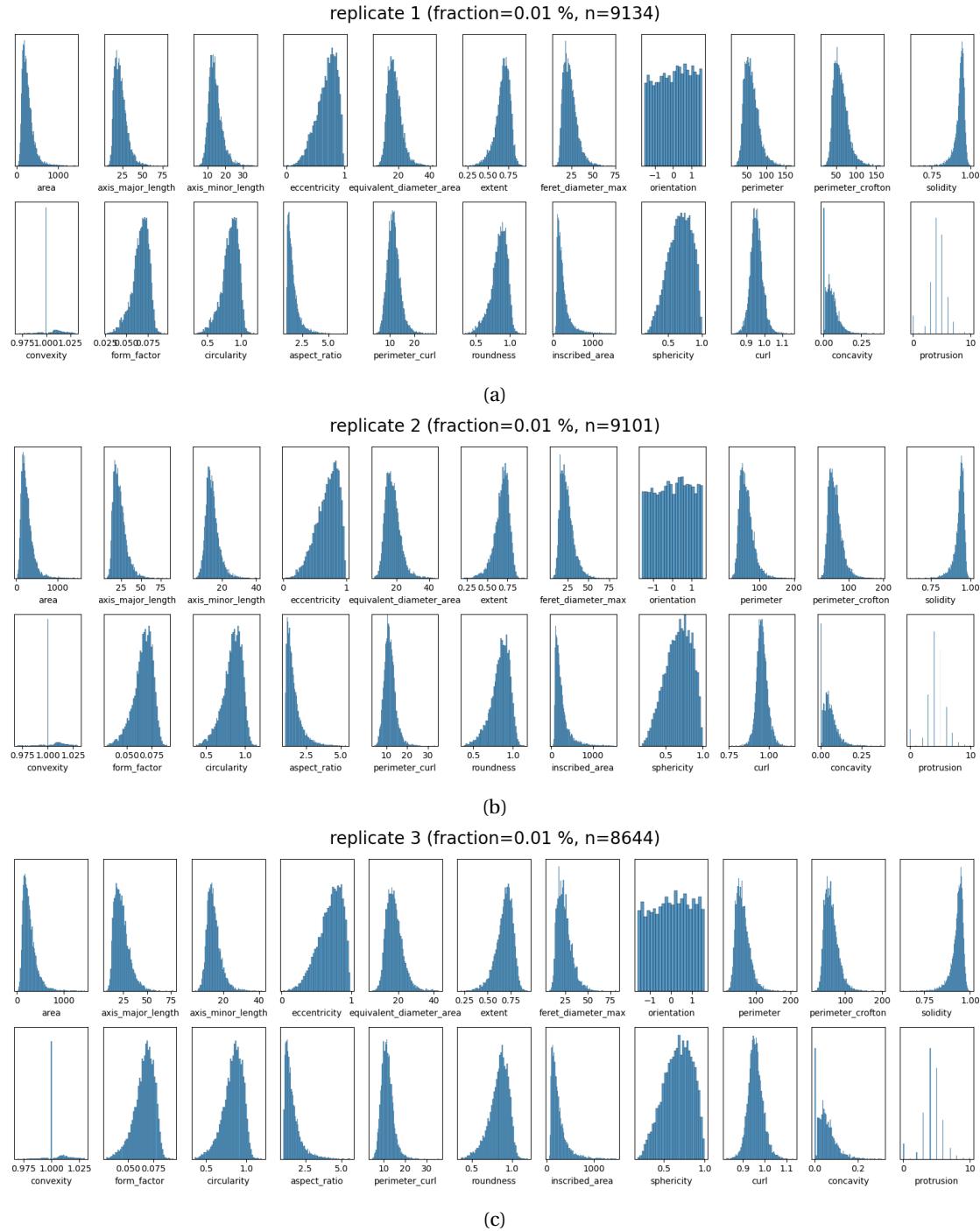


Figure 29: Cell shape feature distributions of all experimental replicates. Displayed is a sample fraction of 0.01 % of the filtered cells for visibility. All of the y-axes are normed throughout all replicates.

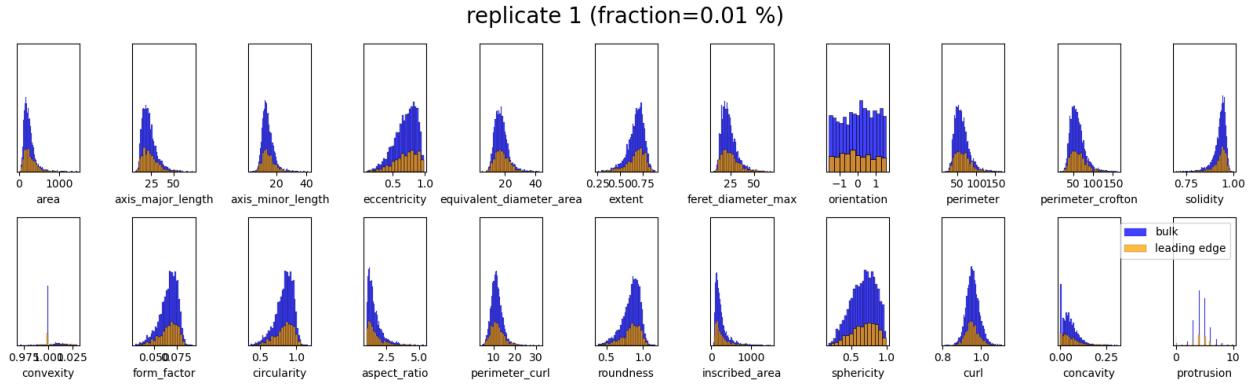


Figure 30: Cell shape feature distributions of replicate 1 of bulk cells (blue) compared to leading edge cells (orange). Displayed is a sample fraction of 0.01 % of the filtered cells for visibility.

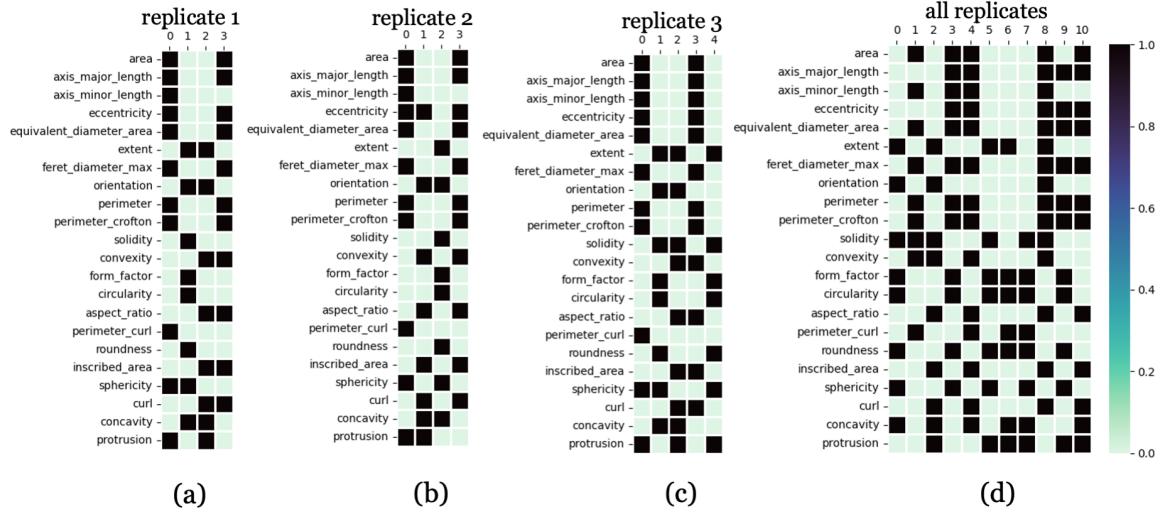


Figure 31: Heatmaps of feature impact on cluster. Significant features for a cluster are marked by the value 1 (black), while insignificant clusters are marked by the value 0 (light blue).