Image Analysis and Platform Development for Automated Phenotyping in Cytomics

Yan, K.

Chapter 4

A Study to Cell Migration Analysis

This chapter is based on the following publications

Le Dévédec, S.E., Yan K., de Bont H., Ghotra V., Truong H.,Danen E., Verbeek F.J., and vande Water B.(2010), "A Systems Microscopy Approach to Understand Cancer Cell Migration and Metastasis", *Cellular and Molecular in Life Science*, Vol. 67, Issue 19, pp.3219-3240

Damiano, L., Le Dévédec, S.E., Di Stefano, P., Repetto, D., Lalai, R., Truong, H., Xiong, J.L., Danen, E.H., Yan, K., Verbeek, F.J., De Luca, E., Attanasio, F., Buccione, R., Turco, E., van de Water, B. and Defilippi, P.(2011), "p140Cap suppresses the invasive properties of highly metastatic MTLn3-EGFR cells via impaired cortactin phosphorylation", *Oncogene*, Vol. 31, Issue 5, pp.624-633. doi: 10.1038/onc.2011.257.

Chapter Summary

In this chapter, we focus on estimating the practical performance of an image analysis solution within the scope of a live cell HT/HC screen study of tumor cell morphology and motility. In detail, we address the following research question:

1. Can the unique phenotypical profile of cells treated with different growth factors be characterized using newly developed algorithms for image segmentation (cf. Ch.2) and object tracking (cf. Ch. 3) in HT/HC screen?

Following the workflow of HT/HC screen (cf. Figure 1-2), this chapter is divided into three major sections. First, the experimental design and image acquisition procedure are illustrated. The selected growth factors and their expected cell responses are explained. Then the design of the image analysis solution is demonstrated and the motivation behind this design is explained. Next, the measurements of cell morphology and motility extracted through image analysis are illustrated. Finally, the variation in morphology and motility are derived from numerical measurements using data analysis. These results are compared to morphology and motility as described in the literature.

4.1. Workflow of Growth Factor Analysis

4.1.1. Experiment Design

Cell morphology and motility are critical parameters in many physiological as well as pathophysiological processes. Recent progress in the technology of live cell enables the capture of detailed information on morphology and motility for cell biology analysis. This chapter aims to use functional genomics and compound screening to unravel the mechanisms of tumor cell migration in the context of breast cancer metastasis. To that end, an image analysis solution is often used to provide an automated analysis solution for such HT/HC *in vitro* live cell screen. Central to this solution is the segmentation algorithm [62] (cf. Ch. 2). From the segmentation, a per-cell tracking over the time-lapse sequence is accomplished via tracking algorithm [9] (cf. Ch. 3). From the tracked cells, a set of numeric measurements related to tumor cell behavior is extracted.

We validated this method through a pilot experiment that assays the random migration of highly motile tumor cells with growth factor regulation. As results, we obtained a sensitive, time-resolved profiling for every condition. During invasion and metastasis, migration is driven by growth factors, such as Epidermal Growth Factor (EGF), Fibroblast Growth Factor (FGF), Hepatocy Growth Factor (HGF) and Transforming Growth Factor β1 (TGF-β1). All growth factors (GFs) are well known chemoattractants or cytokines involved in various cellular processes. However, the response of tumor cells at the level of cell migration to those growth factors, as supplemented individually or in combination, has not yet been quantified. To understand how those different GFs control the tumor cell motility, we performed an experiment that includes single exposure to the different GFs and combinations of GFs.

Cell Culture

Two subgroups of MTLn3 rat breast cancer cells [115] were cultured as previously described [116]. The MTLn3-plGFP is a standard MTLn3 cell line while the MTLn3-GB1 is an MTLn3 cell line with overexpression of the EGF-receptor. These MTLn3 cell-lines were previously described in other tumor migration studies [115][80]. They were maintained in αMEM (Life Technologies, Inc., Gaithersburg, MD) supplemented with 5% fetal bovine serum (Life Technologies). During the random cell migration assay, the cells are exposed to different growth factors known to be involved in tumor progression in general and in tumor cell motility in particular. We analyzed the motility behavior of MTLn3 cells when exposed to 12 different treatments including: DMSO, EGF, FGF, HGF, TGF-β1, EGF+FGF, EGF+HGF, EGF+ TGF-β1, FGF+HGF, FGF+ TGF-β1, HGF+ TGF-β1, EGF+FGF+HGF+TGF-β1. The growth factors used are as follows: Epidermal Growth Factor (EGF, 100 µg/ml), Fibroblast Growth Factor (FGF, 100 µg/ml), Hepatocy Growth Factor (HGF, 20 µg/ml) and Transforming Growth Factor β 1 (TGF-β1, 20 µg/ml). The overall plate design is given in Table 4-1. The DMSO treatment is considered the negative control group.

Table 4-1 Plate design of GF regulation experiment (plGFP-yellow, GB1-green), alphabet represents column index, number represents row index

	plGFP				GB1			
	C	D	E	F	G	H	I	J
2	DMSO	HGF	EGF+HGF	FGF+TGFβ	DMSO	HGF	EGF+HGF	FGF+TGFβ
3	DMSO	HGF	EGF+HGF	FGF+TGFβ	DMSO	HGF	EGF+HGF	FGF+TGFβ
4	EGF	TGFβ	EGF+TGFβ	HGF+TGFβ	EGF	TGFβ	EGF+TGFβ	HGF+TGFβ
5	EGF	TGFβ	EGF+TGFβ	HGF+TGFβ	EGF	TGFβ	EGF+TGFβ	HGF+TGFβ
6	FGF	EGF+FGF	FGF+HGF	all	FGF	EGF+FGF	FGF+HGF	all
7	FGF	EGF+FGF	FGF+HGF	all	FGF	EGF+FGF	FGF+HGF	all

Live Cell Imaging

Glass bottom 96-well plates (PAA) were coated with 20 µg/µl collagen type 1 (isolated from rat tails) for 1 hr at 37 ºC. Cells were plated directly onto the coated glass coverslips and imaged 24 hrs after plating. Prior to the imaging, the medium was refreshed using serum free medium to starve the cells for at least 4 hrs. Cells were then placed on a NIKON Eclipse TE2000-E widefield microscope using a 20x objective lens (0.75 NA, 1.00 WD), perfect focus system, and a 37 ^{o}C incubation chamber. These cells are directly exposed to the different growth factors at the beginning of the experiment. The image acquisition was done with 3 locations per well and GFP signal was acquired every 5 min for a total imaging period of 12 hrs. The images are captured using a NIKON DQC-FS EMCCD camera with 16-bit pixel depth in 512x512 pixels. For each plate, we collected 144 time-lapse image sequences and this experiment was triplicated.

4.1.2. Image Analysis

For each plate, 144 time-lapse sequences are acquired. With a total number of three replicas, the MTLn3 Growth Factor Regulation experiment produced over 432 image sequences, each containing 145 frames. The whole experiment produced over 50k images. Therefore, the use of automated image analysis is essential. The major goal of image analysis solution is to automatically extract both motility and morphology quantifications from these images. To that end, we need to both extract individual cells and track them using an integrated solution combining both image segmentation and object tracking. These HT/HC images are first uploaded to the image management server for the sake of efficient data management. The pipeline is illustrated as four consecutive steps: (1) **image preprocessing**, (2) **image segmentation**, (3) **object labeling**, and (4) **object tracking**.

4.1.2.1. Image Preprocessing

For each image sequence, an image preprocessing procedure is first applied to improve image quality for segmentation using noise suppression or signal enhancement algorithms. The image preprocessing procedure contains four sequential steps (1) **intensity normalization**, (2) **median filter**, (3) **Gaussian filter,** and (4) **Rolling ball filter**. Each step removes or reduces image quality issues during the image acquisition. In the images (cf. Figure 4-1a), we observe an uneven illumination. In the cells, we observe an intensity variation. In addition, a high level Poisson noise [28][57][109] is observed in the images.

Intensity normalization (cf. Figure 4-1b) is first employed to rescale the intensity values over the complete dynamic range. Intensity normalization is used to ensure that the entire dynamic

range is used by all images. The most common intensity normalization formula is described as following:

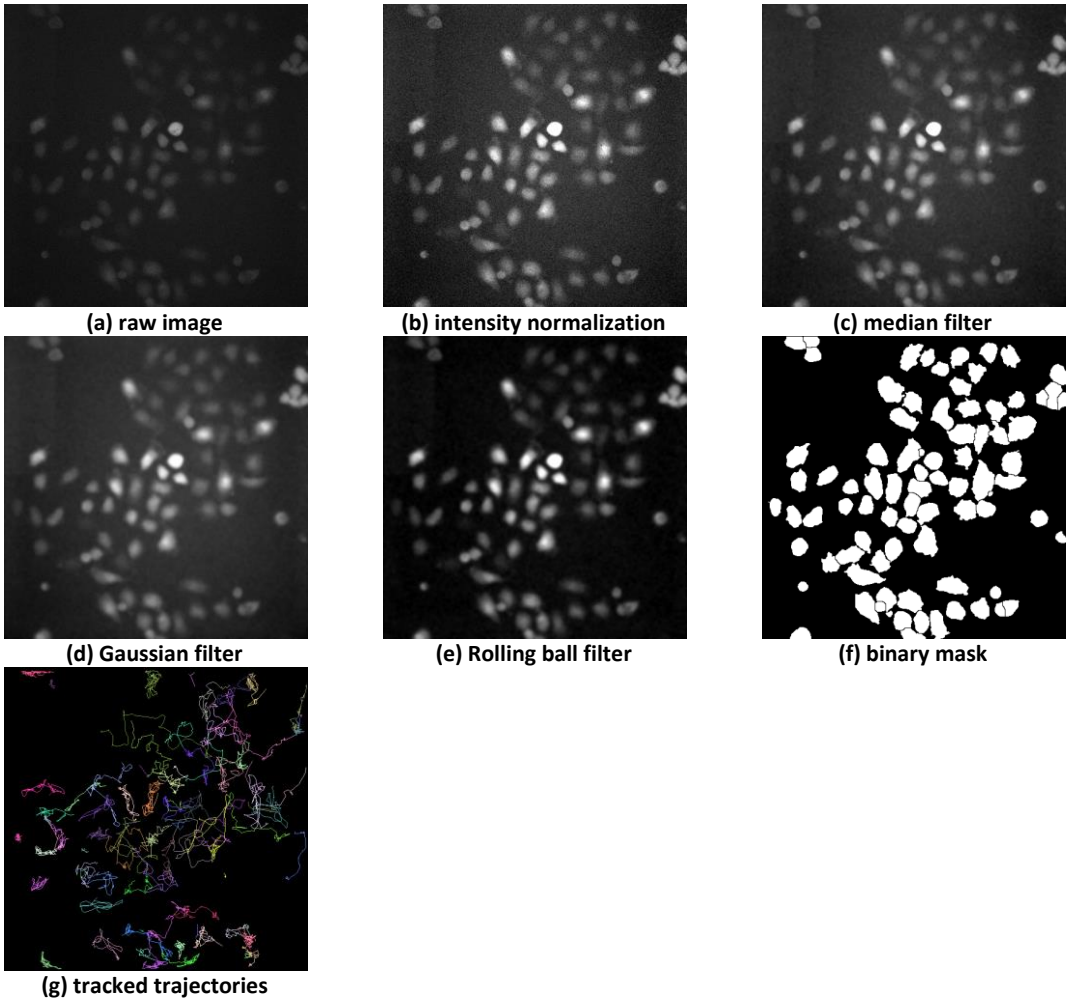
$$I' = \big(I - min(I)\big) \cdot \frac{2^n - 1}{max(I) - min(I)}$$ **Equation 4-1**

, where I is the intensity values in the image and n is the bit-depth of the image.

After the intensity normalization, a **median filter** [85] is applied to the image (cf. Figure 4-1c). A 3x3 median filter kernel [6] can effectively removes the Poisson noise [28][57][109].

Subsequently, a **Gaussian filter** is applied ($\sigma = 2$, just above the Poisson noise) to the image (cf. Figure 4-1d) to create a smoother and more continuous intensity landscape for the seeded watershed algorithm.

Finally, a **rolling ball filter** is applied to the image (cf. Figure 4-1e) to remove the uneven illumination due to autofluorescence [117][101] (cf. Figure 4-1b) from the culture medium. The current rolling ball filter uses a spherical kernel with a radius of 128 pixels so that the kernel is approximately little larger than a cell, leaving the dominant foreground signal. After preprocessing, the images are ready for segmentation.



(a) raw image **(b) intensity normalization** **(c) median filter**

(d) Gaussian filter **(e) Rolling ball filter** **(f) binary mask**

(g) tracked trajectories

Figure 4-1 intermediate results of image analysis

4.1.2.2. Image Segmentation

An image quality assessment with 32 randomly selected images from the experiment (cf. Figure 4-2 raw images) shows that:

1. Regardless the image preprocessing, object intensities are not consistent within the same image sequence. The object intensities are not consistent even for an object at different time points. Such variability seems to be related to different level of GFP expression in the cell [118][21].

As a result, this image set has an average **coefficient of variation (CV) of 4.3** (cf. Ch. 2 Equation 2-8); consequently it is difficult to ensure a high score for both sensitivity and specificity. Therefore, the dedicated solution **WMC segmentation algorithm** (cf. Ch. 2) is used since it is designed to handle intensity variation. From the WMC algorithm (cf. Figure 4-1f), each object is labeled[119][120]. All labeled objects will be passed to the object tracking solution.

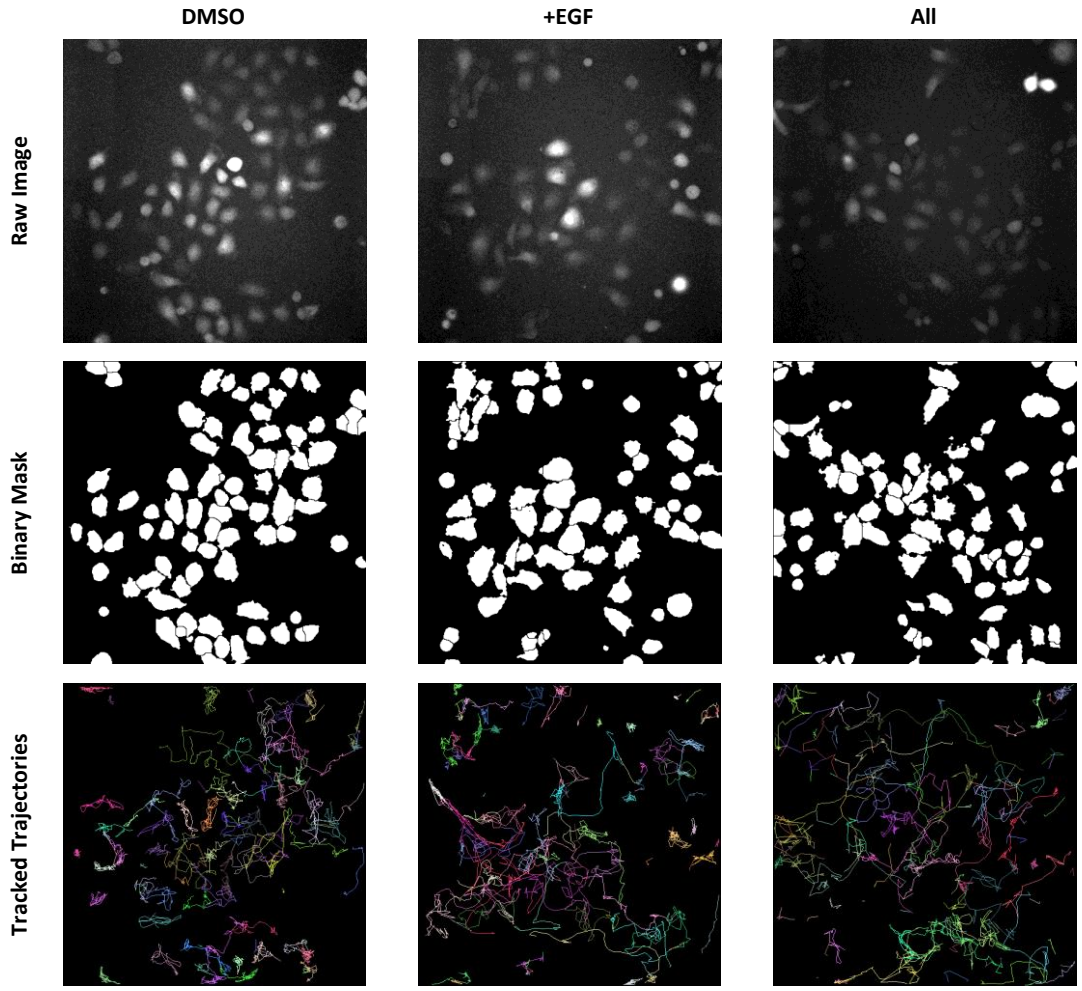


Figure 4-2 segmentation and tracking results for the different GF treatments

4.1.2.3. Object tracking

Cell tracking with this image set faces several limitations. First, since this experiment is a study of cancer cell migration behavior under the influence of a growth factor, the motility of these cells (1.2~4.2 μm/min) [9][10] is higher compared to existing studies [121] (0.1~0.4 μm/min). Confidence measurement based tracking solutions such as overlap tracking cannot be easily adapted since the confidence measurement no longer applies. The variation in the cell motility

under different treatments complicates the performance of motion model based tracking. Moreover, during cell migration a cell is exposed to non-rigid deformations. As a result, the image set has an average **consecutive displacement rate (CDR) of 0.3** (cf. Ch. 3 Equation 3-16), which for the temporal-resolution is barely sufficient to capture the continuity of cell body deformation and position shift during cell migration. Therefore, here we employ the robust tracking solution **KDE mean shift tracking algorithm**. The KDE algorithm can produce a higher overall performance and robustness in tracking cells with constantly changing migration patterns (cf. Figure 4-2 tracked trajectories). Once consecutive objects are tracked, measurements will be further extracted from both binary masks and trajectories.

4.1.2.4. Morphology and Motility Measurement

For image analysis two categories of measurements are introduced: the morphology measurements and motility measurements [9][10]. The morphology measurements are quantifications describing the shape change of each object. They are derived from the binary mask and the associated intensity information of the original image. The motility measurements are quantifications describing the object motion pattern. They are derived from trajectories obtained from the object tracking algorithm. With both motility and morphology measurements, we will further quantitatively compare cell behavior under different treatments (cf. Figure 4-2).

4.1.3. Data Analysis

We designed a data analysis pipeline (cf. Figure 4-3) to extract comprehensive information from the data. The data analysis pipeline converts morphology and motility measurements into a representation to support the research question. The data analysis pipeline consists of: (1) overview, (2) target identification, (3) individual comparison, and (4) manual verification.

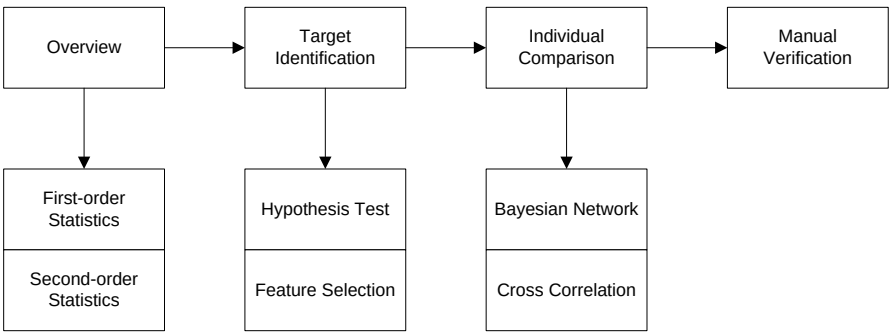


Figure 4-3 data analysis workflow of cell migrations

Overview

In the overview, HT/HC measurements from different conditions and locations are combined and visualized to provide an overall pattern for the whole data set. The combination of the data set in the overview step is accomplished using low-order statistics such as mean or median, and the second-order statistics such as variance or correlation [14][102][122][12][123]. In practice, they are considered an efficient and fast representation over a large amount of data [10][15][124]. As a generalization of cell behavior, it is still rough.

Target Identification

In the target the identification step, measurements are coupled for statistical comparison in order to verify whether these measurements are different by chance. Statistical significance tests [124][125][126] [127] or feature selection solutions (cf. Figure 4-4) [10][15][33][12] are used in this step.

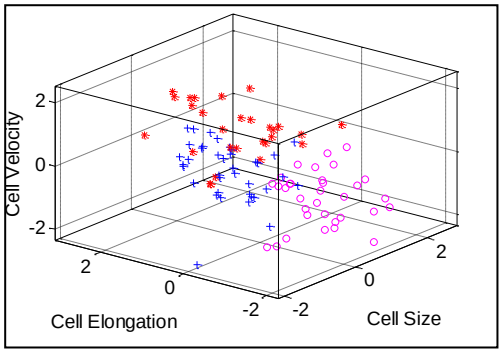


Figure 4-4 top-3 measurements selected by feature selection between ctrl (blue), EGF treated (red), and HGF treated (magenta) cells.

Temporal Profiling

The low-order statistics do not reveal the temporal tendencies in the measurements. Therefore, temporal-order statistics [39][12] are further used as dynamic modeling solution for cell behavior. The typical solutions of cell behavior, i.e. a Dynamic Bayesian Network (DBN)[128][129][130] or a Hidden Markov model (HMM) [33][12], are frequently used in the modeling of spatio-temporal patterns [33][111].

User Verification

The manual verification is the final step in data analysis. The researchers perform a check of the image analysis results of the interesting targets. They also perform a similar check over the control condition. In HT/HC, a behavior that is significantly altered may suggest an interesting target. User verification over these targets will provide further understanding of the data. It is possible that the image and data analysis may result in errors due to experimental bias or design flaws. The user verification of image analysis in both control condition and treated is considered an obvious necessity. We currently use the following checklist during user verification:

- Too few foreground pixels being captured?
- Too many background pixels being captured as foreground?
- Most individual objects separated correctly?
- Most objects captured?
- Most object trajectories tracked?
- Too many trajectories overlapping?
- Too many early terminations of trajectories?
- Is change/difference of measurement corresponding to observation for controls?

In our approach, researchers are presented with a number of overlay images from each condition (cf. Figure 4-5) enabling a quick overview on the performance of segmentation and tracking frame-by-frame. In practice, users check the segmentation result by looking to how the overlay mask fits the user perception of the image content (cells) (cf. Figure 4-5 mask overlay). The trajectories are further overlaid (cf. Figure 4-5b) to the mask-overlay image (cf. Figure 4-5a) to allow users to check the tracking performance.

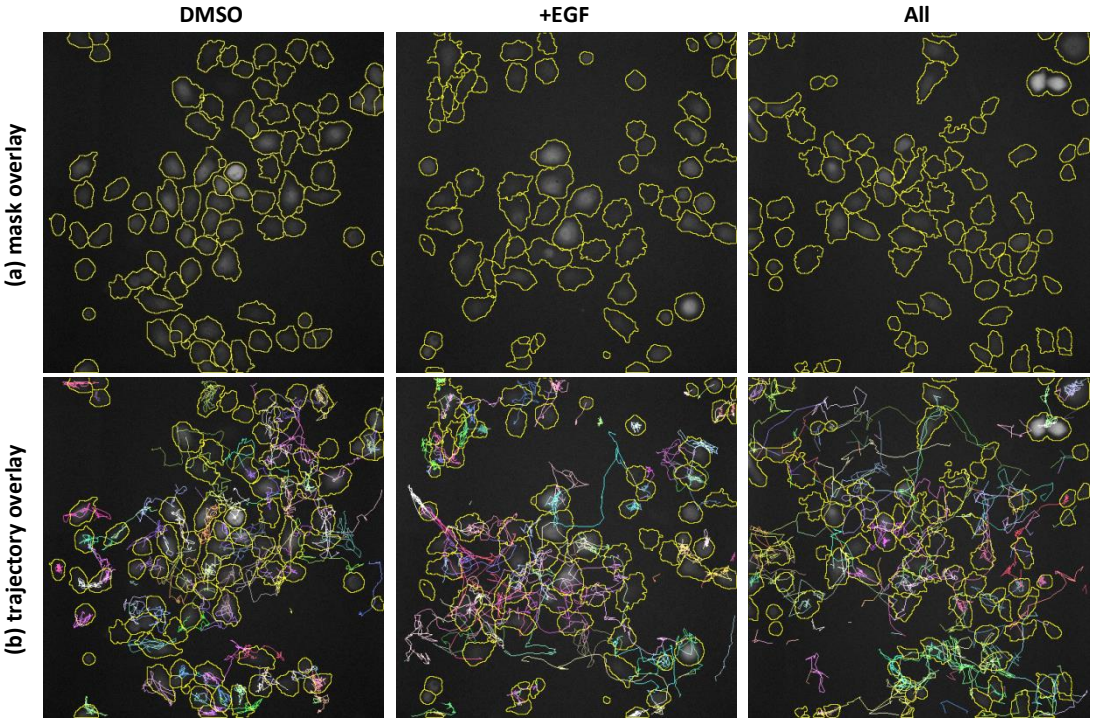


Figure 4-5 manual verification of segmentation and tracking results of the targets

4.2. Analysis of GF Regulation

4.2.1. Differential effect of individual and combined GF stimulation

In Figure 4-6, a heat map is generated from the mean values of all features and a hierarchical clustering is performed. From the heat map, it is clear that HGF treatment as a single growth factor or in combination with other GFs is the most potent regulator of the migration of MTLn3 cells. Indeed, HGF alone and other combinations are all clustered together and the clustering is based on an increase in velocity and motion linearity (cf. §4.2.2 persistence). While EGF is not directly inducing an increase in cell velocity, it seems to affect the directionality of the cells. Surprisingly, TGF- β1 does not seem to have any measurable effect on the migration of MTLn3 cells. Finally, FGF does affect cell size accompanied with a slight effect on motion linearity (cf. Figure 4-2).

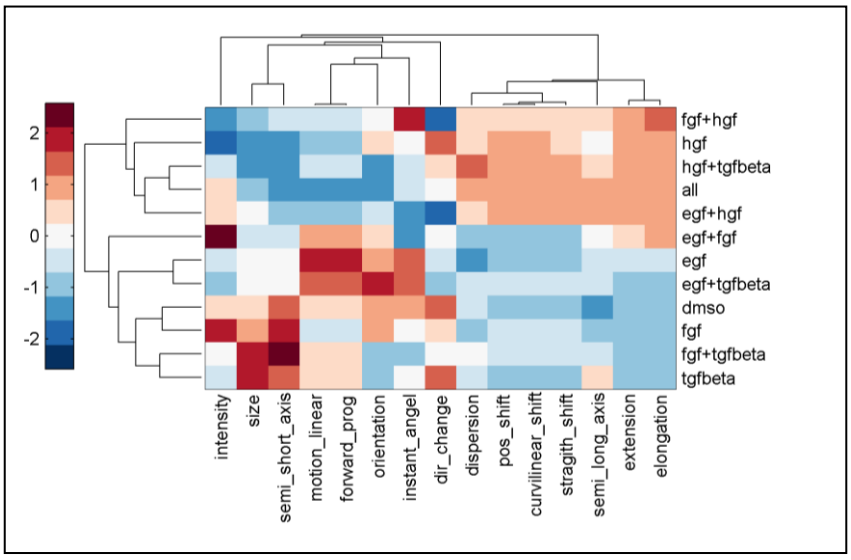
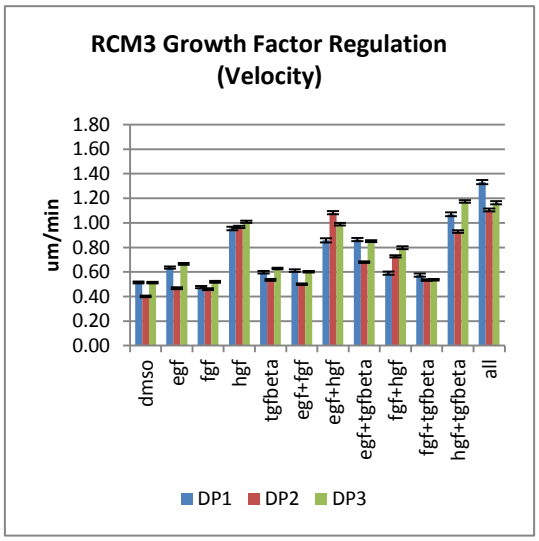


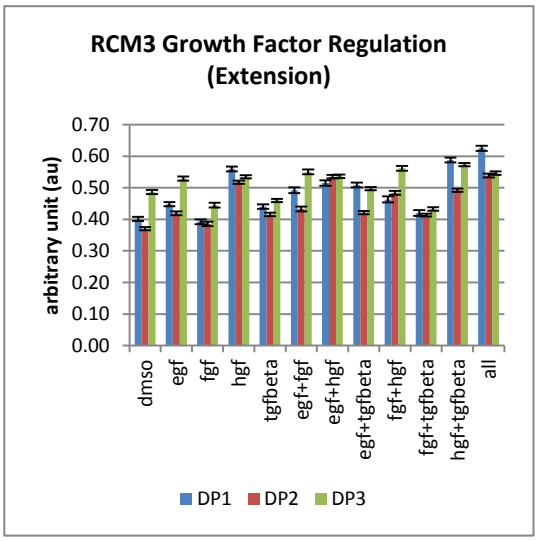
Figure 4-6 heatmap of the RCM3 data plGFP

4.2.2. HGF stimulation increases speed while EGF increases the migration persistence of MTLn3 cells

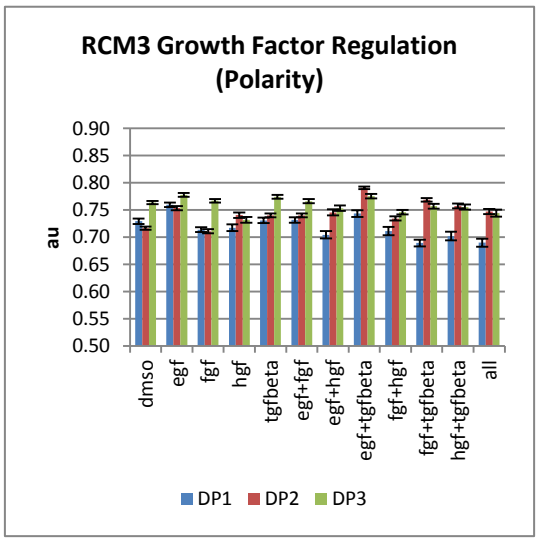
In this section, we introduce the persistence or motion linearity as another very important parameter for cell migration. Persistence is defined as a measure of how efficient the cells move [10]. The analysis result (cf. Figure 4-7) shows that HGF alone or in combination (except for HGF+FGF) does significantly enhance the migration speed of MTLn3 cells. The EGF only slightly affects the speed and interestingly FGF and TGF- β1 do not affect the speed at all. Furthermore, it seems that an increased velocity is always associated with an increased cell protrusion measured through extension [9]. So we can conclude that faster cells show an increased extension as a major feature. On the one hand, the EGF alone or combined with TGF- β1 always affects the motion linearity. On the other hand, the HGF alone does not change much of the motion linearity of the MTLn3 cells.



(a) velocity



(b) extension

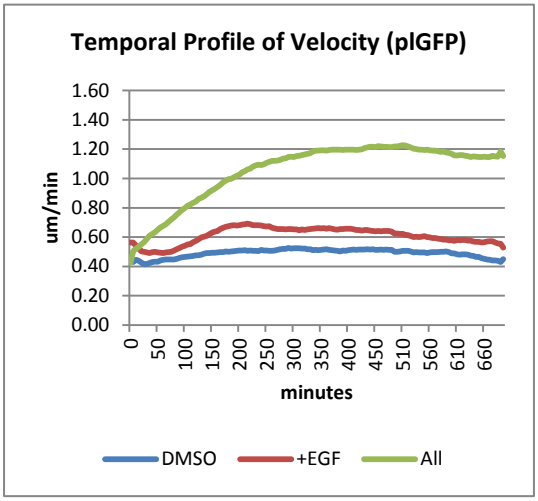


(c) polarity

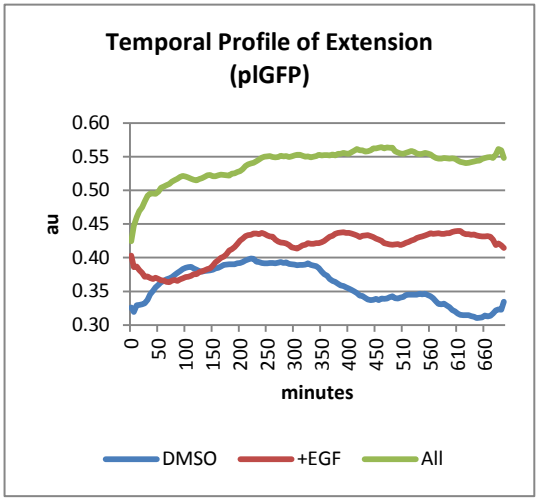
Figure 4-7 RCM3 Growth Factor Regulation plGFP

4.2.3. Temporal-Order Statistics

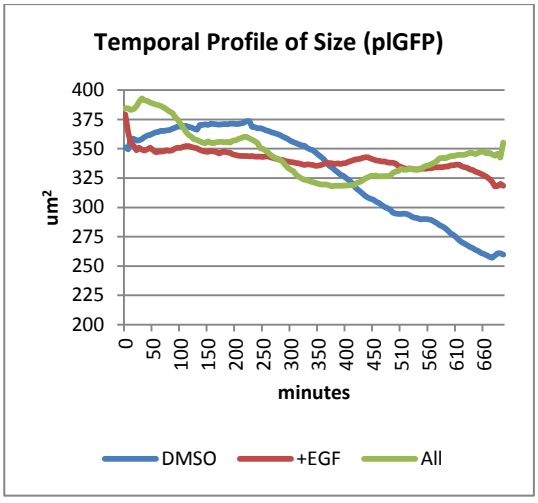
In many cell types, GFs have been reported to enhance the average migration speed. Especially in MTLn3 cells, EGF has been reported to increase the migration speed. The long term migration response of cells after exposure to EGF cannot be extrapolated from a known model. Basic first-order statistics do not reveal time related tendencies, temporal-order statistics [39] are further employed for the spatio-temporal analysis of cell migration. We plotted different parameters over the time-lapse (cf. Figure 4-8) and we ended up with descriptive time profile of the different stimulations. EGF stimluation did result in higher velocity through increased extension. Indeed, almost 1 hr after EGF stimulation the cells start to move faster until reaching a steady state (after approximately 5 hrs) of migration compared to the DMSO condition (negative control). On the other hand, a mixture of all GFs does greatly affect the migration speed of the MTLn3 cells since the temporal profile shows that the velocity increases continuously over the whole time-lapse sequence revealing a complex cascade of signaling that takes place at cellular level. This faster migration results in a continuous increase of cell extension.



(a) velocity



(b) extension



(c) size

Figure 4-8 temporal profile of plGFP

Cell tracks were generated and in our system it does reveal the heterogeneity in behavior within cell population. EGF stimulation indeed seems to increase the velocity only of a certain subpopulation, whereas the GF cocktail does affect all cells, resulting in a significant increase of the mean cell migration speed.

4.3. Conclusions and Discussion

Cell motility is an important event in many biological processes, such as tissue repair, metastatic potential, chemotaxis or analysis of drug performance. Cell migration and invasion are crucial aspects of tumor progression. It depends on the intrinsic capacity of the cells to move faster and more efficient. It is also depending on the tumor microenvironment that does produce all kind of cytokines or extra-cellular components. Understanding how tumor cells move and how their movement can be controlled by internal signaling pathways or external cues, in particular, contributes to our knowledge on tumor progression and metastasis formation. In order to obtain a better insight in the underlying processes leading to an efficient tumor cell migration, methods need to be developed that enable the study of migration at the individual cell level and in a high throughput fashion. Time-lapse imaging with fluorescent microscopy enables a thorough quantification of the cell dynamics at the level of migration. In this study, we successfully developed a computational approach that allows **(1)** reliable segmentation of migrating cells with a high degree of plasticity, **(2)** reliable tracking of fast and often dense migrating cells, **(3)** extraction of motility and morphology parameters over the time of imaging. In this chapter, we have illustrated a dedicated image analysis solution for HT/HC screen analysis combining robust **Watershed Masked Clustering (WMC)** segmentation algorithm and **Kernel Density Estimation (KDE) mean shift** tracking algorithm. This integrated solution produces an accurate profiling of cell migration behavior under the influence of growth factor regulation.

To ensure that our methodology was sufficiently competent for fast moving cells, we performed a pilot study on the migratory behavior of the rat breast carcinoma MTLn3 cells. They show a mesenchymal morphology and do not cluster with cells from the negative control. They are highly motile and propose a challenge for finding good segmentation and tracking solutions. In contrast to generic solutions, our dedicated solution of **WMC** segmentation and **KDE mean shift** tracking algorithm demonstrated a good performance in term of both algorithm efficiency and representation of the biology.

From the biological validation, we further conclude that robustness of both segmentation and tracking algorithm is a crucial factor behind successful analysis in a HT/HC screen study when cell responses from induced treatment are unknown. If the parameter set of an image analysis algorithm is tuned based on control cells demonstrating minimized cell protrusion and motility, with limited robustness, the same parameter set clearly is not optimized for treated cells such as +EGF cell that demonstrate an increasing cell protrusion and motility. The inflexibility not only produces errors in image analysis, but may lead to false conclusions.

On the workbench we resolve this dilemma via the employment of a pilot experiment consisting of (1) control condition in which no treatment is induced, (2) positive control condition with a treatment resulting in the maximum expected responses, and (3) negative

control condition with a treatment resulting in no responses. Parameters of image analysis algorithms and the robustness of selected solutions are verified using the pilot experiment before being applied to full-scale HT/HC screen study.