

RESEARCH ARTICLE

Cell-morphodynamic phenotype classification with application to cancer metastasis using cell magnetorotation and machine-learning

Remy Elbez¹ , Jeff Folz² , Alan McLean³, Hernan Roca⁴, Joseph M. Labuz⁵, Kenneth J. Pienta⁶, Shuichi Takayama⁷, Raoul Kopelman^{1,2,3*}

1 Applied Physics Program, University of Michigan, Ann Arbor, Michigan, United States of America, **2** Biophysics Program, University of Michigan, Ann Arbor, Michigan, United States of America, **3** Department of Chemistry, University of Michigan, Ann Arbor, Michigan, United States of America, **4** Department of Urology, University of Michigan School of Medicine, Ann Arbor, Michigan, United States of America, **5** Department of Biomedical Engineering, University of Michigan, Ann Arbor, Michigan, United States of America, **6** Department of Urology, The James Buchanan Brady Urological Institute, Johns Hopkins Hospital, Baltimore, Maryland, United States of America, **7** Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, Georgia, United States of America

 These authors contributed equally to this work.

* kopelman@umich.edu



OPEN ACCESS

Citation: Elbez R, Folz J, McLean A, Roca H, Labuz JM, Pienta KJ, et al. (2021) Cell-morphodynamic phenotype classification with application to cancer metastasis using cell magnetorotation and machine-learning. PLoS ONE 16(11): e0259462. <https://doi.org/10.1371/journal.pone.0259462>

Editor: Sushanta K Banerjee, University of Kansas Medical Center and VA Medical Center, UNITED STATES

Received: June 14, 2021

Accepted: October 19, 2021

Published: November 17, 2021

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0259462>

Copyright: © 2021 Elbez et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data and Python scripts used for the analysis can be found at the

Abstract

We define cell morphodynamics as the cell's time dependent morphology. It could be called the cell's *shape shifting ability*. To measure it we use a *biomarker free*, dynamic histology method, which is based on multiplexed *Cell Magneto-Rotation* and *Machine Learning*. We note that standard studies looking at cells immobilized on microscope slides cannot reveal their shape shifting, no more than pinned butterfly collections can reveal their flight patterns. Using cell magnetorotation, with the aid of cell embedded magnetic nanoparticles, our method allows each cell to move freely in 3 dimensions, with a rapid following of cell deformations in all 3-dimensions, so as to identify and classify a cell by its *dynamic morphology*. Using object recognition and machine learning algorithms, we continuously measure the real-time *shape dynamics* of each cell, where from we successfully resolve the inherent broad heterogeneity of the *morphological phenotypes* found in a given cancer cell population. In three illustrative experiments we have achieved clustering, differentiation, and identification of cells from (A) two distinct cell lines, (B) cells having gone through the *epithelial-to-mesenchymal transition*, and (C) cells differing only by their *motility*. This microfluidic method may enable a *fast screening* and identification of *invasive cells*, e.g., *metastatic cancer cells*, even in the absence of biomarkers, thus providing a rapid diagnostics and assessment protocol for effective personalized cancer therapy.

Introduction

Despite much progress over the last century, cancer remains one of the leading causes of death globally [1]. Its lethality is overwhelmingly due to metastasis, the process by which cells from

University of Michigan's DeepBlue data library
<https://doi.org/10.7302/513f-1h23>.

Funding: The authors wish to thank the National Institute of Health/National Cancer Institute IMAT program for financial support through an NIH-NCI (IMAT) grant R21 CA160157 (RK), as well as NIH grants CA136829 (ST), R01CA186769 (RK) and 1R01CA250499 (RK). JML gratefully acknowledges support from the University of Michigan Tissue Engineering and Regenerative Medicine Training Program (NIH T32-DE007057), a US Department of Education GAANN fellowship, and the University of Michigan Microfluidics in Biomedical Sciences Training Program (NIH T32 EB005582-05). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

the original cancerous tumor leave their micro-environment (TME) and disseminate to colonize new tissues [2]. During this *metastatic process*, separated single cells, or multi-cellular clusters, migrate through the *extra-cellular matrix* (ECM) surrounding the tumor, passing through the endothelium into the bloodstream [3, 4]. Upon entering the bloodstream, cells and clusters are buffeted by hemodynamic forces on the range of 4–30 dynes/cm² [5–7]. Additionally, these cells must contend with immunological insults and collisions with red blood cells. Having survived under these conditions, cancer cells must latch onto epithelial cells and extravasate into “foreign” tissue, so as to seed a secondary tumor.

To complete the aforementioned challenges, *metastatic cells* must express entirely *different phenotypes* than their stationary counterparts. Specifically, the *epithelial to mesenchymal transition* (EMT) permits the relatively stationary, epithelial cells of solid tumors to obtain the mobility required to *intravasate* and exit the primary tumor, and eventually intravasate at a new tissue location. The EMT may be induced without any gene mutations [8]. It has been observed that the post-EMT *amoeboid-like* cells can significantly increase the *metastatic potential* of the tumor [9]. It has also been reported that *morphological changes* can be used to identify cells having undergone the EMT [10, 11]. Morphology has been linked to cell cycle progression, cell-matrix adhesion properties, gene expression patterns, aging, chemo-sensitivity, and chemo-resistance [12, 13]. Morphology has also been used to predict the *metastatic potential* of both osteosarcoma [14, 15], and breast cancer [16]. Thus, morphology *with its dynamics* presents an attractive option for evaluating cancer progression. Here we emphasize the dynamic aspects, i.e., the *cancer cell's morphodynamics*, that is its time-dependent morphology (or shape shifting ability).

To date, most studies of cell morphology have focused on plated, adherent cell lines. Even though clear morphological distinctions can be discerned among cells when plated, the mere two-dimensional plating process might change their phenotypes and thus alter the quality of the diagnostics [17–22]. Furthermore, tumor grading by professional histologists is characterized by poor reproducibility and accuracy [23]. The goal of these studies has been to correlate genetic features with morphological ones. Wu et al. demonstrated that isolated morphological sub-phenotypes were predictive of tumorigenic and metastatic potentials [24]. In a separate study, the same group showed that metastatic cells possessed more homogenous heritable morphological traits [25]. Another study demonstrated that oncogenesis and metastasis were associated with characteristic changes in morphology [11, 26]. While studies have been conducted that differentiate cancerous from non-cancerous cells [27, 28], we have extended this analysis to compare metastatic with non-metastatic cell types. It would be of significant utility if a morphodynamic classification system could be built for suspended cells, such as circulating tumor cells (CTCs) or those harvested from a biopsy. To realize this goal, we use magnetic nanoparticles so as to trap, suspend, and rotate cells that are captured in a microfluidic chamber [29]. Magnetorotation prevents cells from adhering the microwells and permits exploration of their morphodynamic space.

To examine the cellular morphodynamics, we combine *cell magneto-rotation* with *machine learning* and we show that this approach may allow one to probe both cell motility as well as morphological expression. Machine learning has lent itself to many medical applications and removes the subjectivity of a histologist's analysis [24, 30]. In our approach, green fluorescent protein (GFP) expressing cancer cells are activated by endosomal uptake of magnetic nanoparticles, and are then loaded into a microfluidic device that contains an array of microwells where they remain non-adherent while rotating in an oscillating magnetic field [31, 32]. This enables 3-dimensional cell deformations in which the cells explore and express their morphological phenotype. Most of the device's microwells contain just one single cell; and in each such microwell the single cell is free to take *any* of the shapes that exist in its morphological

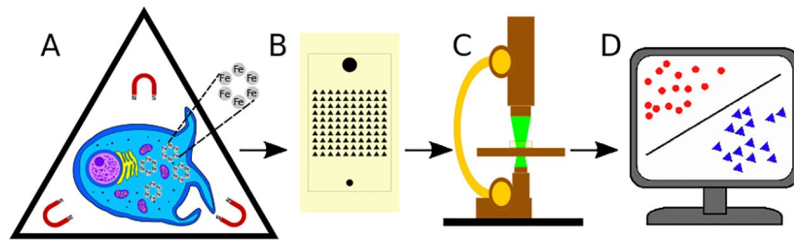


Fig 1. A schematic summary of the cell morphodynamics protocol. A: A captured cell expresses its morphodynamic phenotype while being gently rotated in a magnetic field, enabled by endocytosis of magnetic nanoparticles. B: The microfluidic device contains an array of triangular microwells designed so as to capture individual cells, in spaces large enough for cells to rotate freely. C: Rotating cells are fluorescently imaged on an environmentally controlled microscope stage. D: Cell images are converted by CellProfiler into parameters, used by Machine Learning algorithms to provide cellular clustering, classification, and analysis.

<https://doi.org/10.1371/journal.pone.0259462.g001>

space. After taking fluorescent images of these cells, we combine *object recognition* and *machine learning* algorithms, so as to first *reject* information from multi-occupancy or empty microwells, and then to *differentiate, cluster, and identify* each of the cells by its *morphological profile* (Fig 1). We found that cells having undergone the EMT could be distinguished from control cells, which demonstrates the *morphodynamic equivalent of a change in protein expression*. Furthermore, highly migratory cells were found to be morphodynamically distinct from a control population. This new *machine learning* (ML) based method appears to have the potential to map and classify the *morphodynamic distribution* of a given cell population, and thus to provide information on the degree of *morphological plasticity* of a tumor cell's population. The latter may be related to the tumor's lethality. We have thus used this method here to demonstrate the strong relationship between a cell's morphological and biological behaviors.

Results

To test our technique, we started with an example system: distinguishing between two well characterized breast cancer cell lines of unequal metastatic potential (MDA-MB-231 and MCF-7), using the *AdaBoost* algorithm [33]. The algorithm is capable of correctly identifying cells with a probability of 99%, both in terms of true positives and true negatives, for both phenotypes, as shown in Fig 2A. This means that, on average, our method has a false positive rate below 0.01, and a false negative rate below 0.01 as well (Fig 2A). Using *principal component analysis*, we can project the morphological measurements of the cells onto those principal components that explain most of the data's variance (Fig 2B). Doing so shows that distinct cell phenotypes are clustered together in a manner that allows for separation and classification. Naturally, this does not exclude partial overlap between distributions.

In order to test the capacity of our method, we also reduced the amount of training examples of the target cells (MDA-MB-231, being the most aggressive cells) in the general population (being represented by the MCF-7 cells). Surprisingly, even at a ratio of 1/1500 the algorithm is still capable of learning to differentiate the two cell lines, with an f1 score of 0.965. As scores for the reduced number of examples from one of the two categories remain high, it is instructive to look at how the *standard deviation* of the score changes (Fig 2C). We see that reducing the presence of one target population largely contributes to results with a higher variance; in other words, outliers, even small ones, have a larger effect for smaller sample sizes. One of the reasons for this is that having fewer examples for one class contributes to overfitting that particular class, and thus brings a poorer discrimination power when more diverse examples of the same class are presented to the computer.

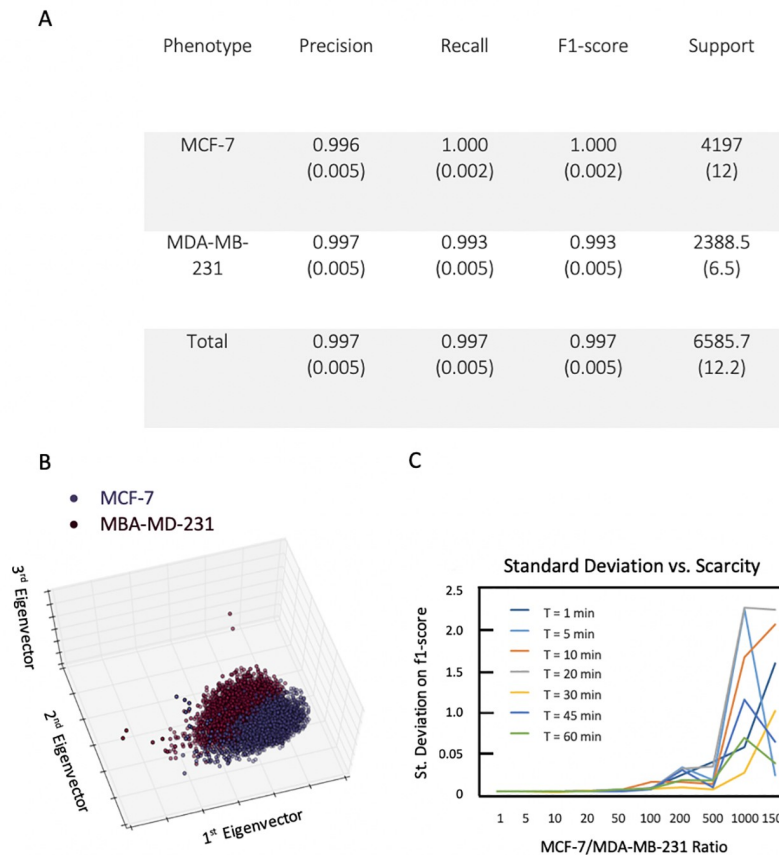


Fig 2. A ‘proof of concept’ classification task by morphodynamics with two distinct cell lines: MCF-7 (low metastatic potential, epithelial) and MDA-MB-231 (high metastatic potential, mesenchymal). A: The detection power for MCF-7 vs. MDA-MB-231 after 1 minute (one image is captured every minute) using Adaboost. We observe high precision and recall, in both classes, indicating that the algorithm is robust for the two phenotypes, that the rate of false positives is low, and that we are capturing nearly every cell in each group (support designates the number of cells in a given group). B: Projections of the measured data onto the first three eigenvectors (principal components) reveals distinguishable clusters for the two cell lines: MCF-7 (blue) and MDA-MB-231 (red). Each point in the eigenvector space represents a single cell. C: A plot of the f1-score’s standard deviation as a function of the MCF-7/MDA-MB-231 ratio. Colored lines indicate how long the cells were imaged. By increasing the number of scans, the classifier can become more confident in the phenotype of a particular cell, allowing us to distinguish the artificially rare subpopulation of MDA-MB-231 cells, even as their relative abundance becomes less than 0.1% of the population.

<https://doi.org/10.1371/journal.pone.0259462.g002>

Having distinguished cells from two distinct lineages, we next utilized our method to analyze cells that originally shared the same phenotype, but had undergone a significant change, i.e., the *epithelial to mesenchymal transition* (EMT). To do so, we forced a human prostate cancer cell line, PC-3, to go through the EMT, yielding the cell line HR-14 [34]. As can be seen in Fig 3A, when using the representation of the cell measurements in the eigenvector space after only one imaging sequence, the two populations of cells are readily separable. To determine if any *sub-phenotypes* existed among these populations, we ran unsupervised *k-means clustering*, in which we partition our observations into *k* clusters, with each observation falling into the cluster with the nearest mean [35]. We chose *k* so as to maximize cluster homogeneity, with a strict constraint that all clusters must have a homogeneity greater than 0.95. In doing so, we found seven (7) sub-phenotypes (Fig 3B). Therefore, with our microfluidics imaging method, just using Cell Magneto-Rotation (CMR) coupled with unsupervised clustering techniques, we could identify and discriminate cells that went through an EMT. Furthermore, we were able to

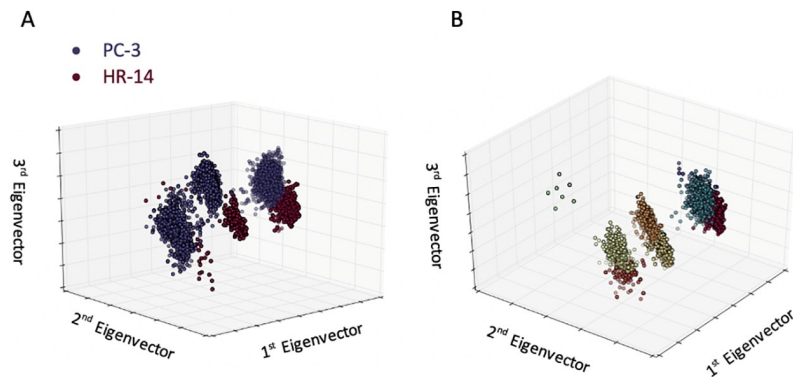


Fig 3. Unsupervised analysis of PC-3 (epithelial) and HR-14 (mesenchymal) cell lines using *k-means clustering*. A: The projection of PC-3 (blue) and HR-14 (red) cells onto the first three eigenvectors. Notably, when training the computer, it gets the information about each cell being either PC-13 or HR-14. Before any machine learning has been done, we can see that the two cell lines clearly cluster into distinguishable groups. Interestingly, we find that the clusters are not continuous and the emergence of *morphological sub-phenotypes* is apparent. B: The results of *k-means clustering* with the constraint that each cluster must maintain a homogeneity score greater than 0.95. We find that the constraint for highly pure clusters results in the identification of 7 sub-phenotypes.

<https://doi.org/10.1371/journal.pone.0259462.g003>

identify the presence of *sub-phenotypes*, though their physiological manifestation remains uncharacterized.

Finally, we sought to detect subpopulations of cells within a given population. To do so, we separated those MDA-MB-231 cells that had a *higher motility* than the rest of the population, using a *Boyden chamber* [36]. We compared the data collected from these highly motile cells with those from the cells that had failed to migrate (Fig 4). Our results show that, using just a *morphodynamic analysis* with *k-means clustering*, we were able to easily distinguish the highly motile subpopulation of MDA-MB-231 cells from that of the general population. Our homogeneity score stands at 0.96 for 7 clusters, which means that the clusters have a very high purity. As a consequence, each cluster is composed, almost exclusively, of cells that have the same

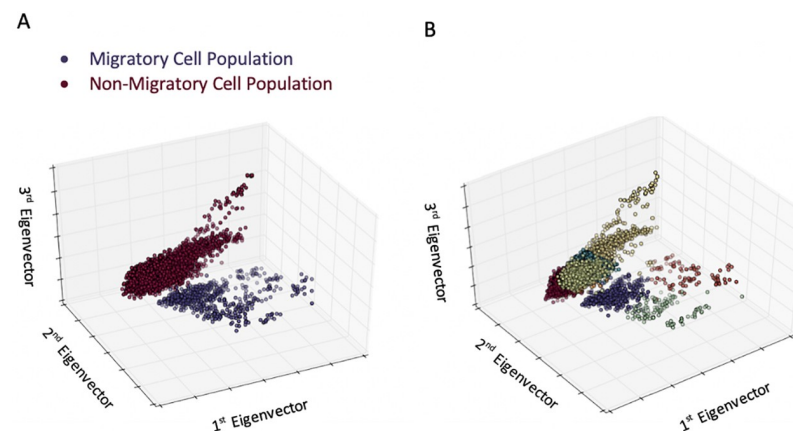


Fig 4. Unsupervised analysis of highly motile cells separated from the bulk population chamber via a *Boyden chamber*. A: Projections of migratory (blue) and non-migratory (red) MDA-MB-231 cells. The migratory and non-migratory fractions segregate into distinct clusters of cells, with many individual or small clusters of cells expanding into the periphery, away from the main clusters, indicating the presence of *morphological sub-phenotypes*. B: Using the *k-means clustering* algorithm with the strict criterion that clusters must have a homogeneity greater than 0.95, we find seven distinct clusters. In the absence of genetic profiling, we cannot confirm the biological role of these clusters, but have demonstrated that morphology alone is enough to distinguish, cluster, and analyze cell sub-phenotypes.

<https://doi.org/10.1371/journal.pone.0259462.g004>

phenotype (“normal” or invasive). We conclude that a simple morphodynamics test can reliably predict the results of a motility test, such as a *Boyden Chamber* test, i.e., segregate the motile from the non-motile cell populations. It appears, therefore, that using unsupervised clustering (and without any human input regarding classification into phenotypes), we can detect, with this morphodynamics based method, subpopulations of cells that are very different from the rest of the population, including their motility-related potential aggressiveness, i.e., *metastatic potential*. Thus, our technique augments, or obviates, the Boyden chamber assay, by allowing us to predict, using the morphodynamics analysis, which cells will migrate through the chamber. Additionally, we can detect sub-phenotypes of both migratory and non-migratory cells, though the origin of these sub-phenotypes remains uncharacterized. This preliminary evidence that new sub-phenotypes can be detected by our method will be a main item in our future studies.

Discussion

Our *morphodynamic histology* approach pushes the biological adage, “structure dictates function”. Using only a morphology analysis, we were able to cluster and differentiate (A) cells of separate and distinct lineages (e.g., MCF-7 vs. MDA-MB-231), (B) cells before and after having gone through the EMT (PC-3 vs. HR-14), and (C) cells from a single cell line (MDA-MB-231) that differed only in their motility. Our first classification experiment (Fig 2) is a proof of concept, as the ability to distinguish between the two genotypically distinct cell lines MCF-7 (epithelial-type) and MDA-MB-231 (mesenchymal-type) can be accomplished by eye. However, it is less trivial to delineate cells sharing a single genotype that have undergone the EMT. The second experiment (Fig 3) demonstrates that we can readily distinguish between pre- and post-EMT cells, suggesting that changes in phenotype are detectable as morphodynamic changes. In addition, the HR-14 cells have been shown, in animal models, to be much more metastatic than their epithelial counterparts [34]. In our final experiment (Fig 4), we used only a single population of cells, and we were able to identify, by morphodynamics alone, the highly motile cells within that population, as confirmed by a *Boyden* migration assay.

Importantly, we can track behaviors and morphological changes that are intrinsic to a *label free*, floating or circulating, cell, without relying on any biomarker or on any *a priori* knowledge of the genotype, which are often required for other CTC-capturing techniques [37, 38]. We note that *protrusions*, *blebs* and *amoeboid* morphologies have all been strongly correlated with an increase in malignancy [9, 34, 39, 40]. It thus follows that detecting such morphological cues could greatly help in the identification of those cells that are most responsible for the metastatic process. Importantly, early identification of cancerous or pre-cancerous morphodynamic phenotypes could enable early detection of cancers which produce no known biomarkers, such as is the case for pancreatic cancer [28, 41]. In future work, it would be beneficial to extend the dynamic morphodynamics platform to demonstrate sensitive differentiation of such cells, which could be of important diagnostic value.

Another advantage of the morphodynamic cell phenotyping approach is the ability to easily track thousands of individual cancer cells. Since the cells are loaded onto a grid of microfluidic wells, the location of each cell is known for the duration of the experiment, and the morphodynamic features of the cells can be tracked over time. While each present device is limited to a maximum capacity of 10,000 wells, the experimental approach is easily automatable, and there appears to be no difficulty in extending the analysis to other common model cell lines, or even to non-adherent cells.

To emphasize, in this study we have been able to start from a purely physical readout, i.e. the *shape features* of rotating cells, and end up segregating cells into morphodynamically unique clusters. While subtle differences in nanoparticle formulations have been shown to

effect cell responses to therapy [42, 43], our technique has been demonstrated to have minimal impact on cell viability (S2 Fig in S1 File). All along the analysis, our method maintains the ability to follow single cells individually. As such, this system lends itself well to serving as a single cell analysis assay. Importantly, each of the populations we studied (Figs 2–4) could be readily split into two distinct populations (e.g., MDA-MB-231 vs. MCF-7). Beyond that, our use of *unsupervised learning* demonstrated that there exist in these cell groups identifiable subpopulations of cells. The underlying genetic origin of these sub-clusters remains unresolved and their elucidation is beyond the scope of the present report. Indeed, their resolution may not be strictly fruitful as the inherent genetic instability of many late stage, metastatic cancers makes finding reliable, gene-based biomarkers quite challenging [44, 45]. Also, our technique, by probing morphology directly, avoids the difficulty of *attempting to associate unstable genetic codes with single, static cell states*.

Finally, we notice that by using just our morphodynamic analysis, we were able to easily distinguish the highly motile subpopulation of the MDA-MB-231 cells from that of the general population. As this entire cell line is considered to have a relatively high metastatic potential, and motility is often associated with intravasation, our analysis may have identified the more, or most, aggressive subpopulation of this cell line. Furthermore, we clearly demonstrate that morphology can be used to associate cells with real, physiological behaviors, such as crawling through the extracellular matrix (Boyden chamber). In our future work, our goal will be to establish the relationship between the biological heterogeneity and the morphological expression within a cell population, ultimately leading to the characterization of EMT activation within a cell population, without the need of any biomarker or genetic profiling.

In summary, we first introduced a new concept, *cell morphodynamics*, as well as the method for measuring it, based on the cell magneto-rotation (CMR) technique, which prevents cell adherence and allows 3-dimensional cell deformations, and on combining CMR with machine learning (ML) algorithms. This morphodynamic method is thus based on a *label-free testing* of non-fixed, minimally perturbed, live individual cells, kept in bio-mimetic micro-environments. Our massively parallel single cell analysis assay investigates the similarities and dissimilarities of cancer cells' morphological behaviors over time, and we could thus identify cells whose phenotype may be associated with a more, or most, malignant potential, including motility and invasiveness, and achieving this *without the use of any biomarker*. We further note that this approach does lend itself well to mapping the *heterogeneity characterizing a tumorous cell population*, as well as identifying the presence of both morphologically and biologically distinct subpopulations. We believe that these techniques could well present health-care providers with a new and inexpensive tool for evaluating and predicting the *plasticity* and *potential aggressiveness* of a population or subpopulation of cancer cells, and how it might behave, without a genetic screen. We do plan, in future work, to further develop the method introduced here for the characterization of cellular subpopulations. This may benefit from sequencing specific single cells. Overall, this effort will be geared toward monitoring changes in the magnitudes and ratios of subpopulations of cell groups, so as to better predict a tumor's metastatic potential. We hope that such a rapid and reliable estimate of a tumor's migration potential could become an important feature of *informed precision cancer medicine*; it would provide the caregiver, as early as possible, with the likelihood of metastasis.

Materials and methods

Preparation of Magnetic Nanoparticles (MNPs)

Amine-coated magnetic nanoparticles (Ocean Nanotech[®]) with a diameter of 30nm, are prepared in a 1mL stock solution of 200 μ g/mL in cell culture media. We then add 15 μ L of Poly-

L-Lysine at 0.1%w/v (Sigma-Aldrich[®]), and the solution is left for an hour on a rotator at room temperature. The solution is then filtered using a 0.2 μ m syringe filter.

Cell culture and magnetization

For these experiments, two lines of breast cancer cells, MCF-7 and MDA-MB-231, and one line of prostate cancer cells, PC-3, was used. These three cell lines were purchased from ATCC[®]. A fourth cell line, dubbed HR-14, which consisted of PC-3 cells that had undergone the EMT, was also used [34]. All cell lines were stably expressing Green Fluorescent Protein (GFP) and cultured in RPMI 1640 supplemented with 10% fetal bovine serum (FBS) and 1% Penicillin-Streptomycin-Glutamine (PSG) in a cell incubator at 37°C, with 5% CO₂ and 100% humidity. Media and supplements were all purchased from Life Technologies[®]. Cells' confluency before addition of the MNPs is around 20–30%. Cells are incubated for 24 hours with cell culture medium to which is added (see below) 20 μ g/mL of amine-coated magnetic nanoparticles. These particles are uptaken via endocytosis (S1 Fig in [S1 File](#)).

Microfluidic trapping system and cell loading

One hour before being exposed to fluorescent light, cells are washed with Hank's Balanced Salt Solution (HBSS) three times to remove traces of phenol red contained in the cell culture media, and then incubated for an hour in a colorless cell culture media that has been supplemented with the radical oxygen scavenger, Trolox (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid, Sigma-Aldrich), at 0.25nM. After an hour, cells are washed with HBSS, and gently detached using a cell scraper. Cell density is then adjusted by the help of a magnetic separator.

Cells are then gently pipetted into the microfluidic device. The microfluidic trapping device is made of polydimethylsiloxane, according to the protocol used by Park et al. [46]. Each well has a triangular shape, with a side size of 40 μ m and a depth of 35 μ m. The chip has two ports: An inlet port and an outlet port. Cells are loaded with a 100 μ L pipetter into the inlet, and gently introduced into the channel by applying negative pressure at the outlet. Once positioned, the device is put on top of a rare earth magnet to pull the cells down into the wells. We repeat these steps several times, until a sufficient loading ratio is achieved (above 60% of the traps occupied by single cells). These loading steps take around 3 minutes, and no more than 5 minutes. Finally, cells are washed with fresh media by gently pipetting fresh media into the device (fresh media is placed at the inlet port and pipetted from the outlet port). At the end of the imaging series (usually 60 minutes), propidium iodide is pipetted into the inlet port and the cells are imaged so that dead cells can be removed from the analysis.

Cell Imaging and rotation

Cells are imaged on an Olympus[®]IX71[™] microscope, equipped with an arc-mercury lamp (U-RX-T[™]) and a high definition monochromatic digital camera (Q-Imaging[®]Retiga 6000, 10 Megapixels). To image simultaneously multiple positions of the device, the microscope stage is replaced with a motorized stage (ASI MS-4400 XYZ Automated Stage). Images are captured with the software package Micro-Manager (extension of ImageJ), while the stage is programmed and controlled via a custom made script in Micro-Manager [47]. To protect cells from light exposure, a custom made shutter opens for 700ms at every position each minute. Only single cells are kept to be measured. Temperature and humidity are controlled using a homemade, on-stage system that keeps the cells at 37°C with 100% humidity. Cell media is supplemented with HEPES in order to maintain pH in the absence of CO₂. The oscillating magnetic field is generated via 4 solenoids positioned around and slightly above the

microfluidic device (S2 Fig in [S1 File](#)). All solenoids are driven by an alternating current with frequency of 15Hz; two solenoids are driven 90° out of phase. Suspended cells rotate with a frequency of 0.1Hz.

Image processing

Raw images consist of a grid of cells at regular intervals (each cell is sitting in regularly-spaced microwells). Each live, single cell is cropped from the original image into separate, smaller images, each consisting of a single cell. It is these cropped images of single cells that are analyzed. The basis for measuring cell morphology relies on the accurate delineation of a cell's contours (S4 Fig in [S1 File](#)). This task is performed by a pipeline with the image analysis software CellProfiler [48]. Once cells are delineated, CellProfiler measures and records the value of different morphological parameters, such as cell area, perimeter, extent, etc., as well as Zernike moments and Haralick features. For a single experimental run, over 1000 individual cells are processed, and each cell has over 100 measured features.

Data processing

To form a training set, the list of all the cells that have been monitored is established. Every time the classification function is called, the list of names (from different populations of cells that we want to distinguish) is re-shuffled and 70% of the cells are randomly selected to be part of the training set, the 30% left are kept as a testing set that is used to establish the efficiency of the algorithm; these percentages typically correspond to 1100 training cells and 400 tests cells. When a time limit is set, only the measurements at time points smaller than the time limit are kept to form the training and testing data sets. Once selected, the training set is normalized and its dimension reduced to 14 components with Principal Component Analysis (PCA). The parameters used for normalization and dimension reduction are kept and used to perform the same transformations on the testing set. To avoid over-fitting problems, we also used a Cross Validation technique, with a random shuffling of the data samples. A classifier is trained and results are calculated on the testing set. For a specific time limit and ratio between general and target population (i.e. MCF-7 vs MDA-MB-231), we repeat all these steps (from shuffling to testing) 30 times and average the results.

Adaboost method—Supervised learning

In order to perform the machine learning step of our method, we split our data (morphology measurements) into two distinct subsets: a training subset and a testing subset. The training subset is used to train the computer to make decisions, while the testing set is used to evaluate its performance. We used 70% of the data as a training set, and the rest as a testing set, as is customary for machine learning problems [49, 50]. To avoid over-fitting, we also used a cross validation technique, with a random shuffling of the data samples. Cross validation trains the computer by using different training sets and evaluating its resolving abilities on the corresponding testing sets. This helps avoiding the problem of training the computer on a subset where the samples are too similar, leading to over-fitting, because the algorithm will be able to detect and rightly recognize only small variations from the training subset. In that case, when tested, the algorithm would perform poorly, and other variations would be missed. Shuffling cells randomly reduces the likelihood of this issue occurring. Finally, before we commence learning, the data is normalized and we perform a principal components analysis to reduce the dimensionality of the data from 169 to 14.

Training the computer means that for each entry, or cell measurement at a specific time point, we let it know the phenotype to which this entry belongs (0 if MCF-7, and 1 if

MDA-MB-231). Using the AdaBoost algorithm [33], the computer builds a decision procedure. When presented with unlabeled data (testing set), the decision rules are used to make predictions on the labels to assign whether a cell is an MCF-7 or an MDA-MB-231 cell. A way to measure the efficiency of a classification is by measuring its precision and recall, which are defined below:

$$Precision = \frac{T_p}{T_p + F_p} \quad (1)$$

$$Recall = \frac{T_p}{T_p + F_N} \quad (2)$$

Where the F1 score is the geometric mean of precision and recall:

$$F_1 score = 2 \frac{Precision * Recall}{Precision + Recall} \quad (3)$$

Where T_p and F_p are true positives and false positives in the classification task, and F_N represents the false negatives.

K-means clustering—Unsupervised learning

Without indicating from which sample the cells came from (epithelial or mesenchymal), we used an unsupervised clustering algorithm to group similar cells together, and to find clusters. We then compared the clusters that were found with the actual sampled labels. To measure the accuracy of the fit, we used the homogeneity score. The homogeneity of the clustering measures whether each cluster contains only members of a single class (i.e. phenotype), and its value is between 0 and 1, where 1 means a perfect clustering and classification. Let $A = A_1, A_2, \dots, A_n$ be the true classes of data points that we have (“the ground truth”), and $C = C_1, C_2, \dots, C_l$ the classes obtained after clustering operations. We will set N to be the total number of data points. Let $a_m = \|A_m\|$ be the number of objects (i.e. cells) belonging to the m -th class, $c_k = \|C_k\|$ be the number of objects classified into the k -th cluster by the algorithm, and n_{mk} the number of objects that belong to both A_m and C_k . We can then define the homogeneity measure as:

$$homogeneity = 1 - E(A/C)/E(A) \quad (4)$$

where,

$$E(A/C) = - \sum_m \sum_k \frac{n_{mk}}{N} \log \frac{n_{mk}}{a_k} \quad (5)$$

$$E(A) = - \sum_k \frac{a_k}{N} \log \frac{a_k}{N} \quad (6)$$

Clustering was performed using the k-means algorithm [51]. The principle of this algorithm is to find clusters by minimizing the within-cluster sum of squares (WCSS). At first, k random points, called “means”, are chosen, and for every single point left, the cluster to which it is attributed is the one where the WCSS is minimal. For each cluster formed, the means are calculated, and the attribution process is done again. These steps are repeated until the clusters are stable.

Inducement of EMT in PC-3 cells

This epithelial to mesenchymal transition (EMT) protocol was originally developed by Roca et al [34]. Briefly, a subpopulation of PC-3 cells expressing luciferase and presenting an epithelial morphology were isolated. These cells were then co-cultured with interleukin-4 treated, CD14+ monocytes. These lines were cultured together for four days, which induced strong morphological changes in the PC-3 cell lines (PC-3-EMT). The PC-3-EMT cells were isolated and it was confirmed that concomitant with the morphological changes, the cells experienced a decline in E-cad expression while Vimentin expression had increased—changes consistent with cells having undergone the EMT.

Cell migration assay

MDA-MB-231 cells were loaded into a standard Boyden chamber for cell migration assay [Cultrex Cell Migration Assay by R&D Systems]. After 12 hours, the highly motile cells that went completely through the porous membrane were detached and collected from the bottom part of the chamber. They were immediately loaded into the device and imaged with the help of fluorescence while being rotated.

Supporting information

S1 File. Methods and controls. This document contains detailed description of the dynamic morphology method, both experimental and computational, as well as cell viability controls. (PDF)

Acknowledgments

We thank Dr. Celina Kleer and Dr. Maria Elena Gonzalez of the University of Michigan's Department of Pathology for providing the GFP-expressing MCF-7 and MDA-MB-231 cells.

Author Contributions

Conceptualization: Remy Elbez, Jeff Folz, Kenneth J. Pienta, Shuichi Takayama, Raoul Kopelman.

Data curation: Remy Elbez, Alan McLean.

Formal analysis: Remy Elbez, Alan McLean.

Funding acquisition: Shuichi Takayama, Raoul Kopelman.

Investigation: Remy Elbez, Hernan Roca, Shuichi Takayama, Raoul Kopelman.

Methodology: Remy Elbez, Shuichi Takayama, Raoul Kopelman.

Project administration: Remy Elbez, Jeff Folz, Shuichi Takayama, Raoul Kopelman.

Resources: Hernan Roca, Joseph M. Labuz, Kenneth J. Pienta, Shuichi Takayama, Raoul Kopelman.

Software: Remy Elbez.

Supervision: Shuichi Takayama, Raoul Kopelman.

Visualization: Joseph M. Labuz.

Writing – original draft: Remy Elbez, Jeff Folz.

Writing – review & editing: Jeff Folz, Alan McLean, Raoul Kopelman.

References

1. Institute for Health Metrics and Evaluation (IHME). Findings from the Global Burden of Disease Study 2017. Seattle, WA: IHME, 2018.
2. Tremblay PL, Huot J, Auger FA. Mechanisms by which E-selectin regulates diapedesis of colon cancer cells under flow conditions. *Cancer res.* 2008; 68: 5167–5176. <https://doi.org/10.1158/0008-5472.CAN-08-1229> PMID: 18593916
3. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics. *Ca Cancer J Clin.* 2014; 64: 9–29. <https://doi.org/10.3322/caac.21208> PMID: 24399786
4. Friedl P, Alexander S. Cancer invasion and the microenvironment: plasticity and reciprocity. *Cell.* 2011 Nov; 147 (5): 992–1009. <https://doi.org/10.1016/j.cell.2011.11.016> PMID: 22118458
5. Mitchell MJ, King MR. Computational and experimental models of cancer cell response to fluid shear stress. *Front Oncol.* 2013 Mar; 3 (44). <https://doi.org/10.3389/fonc.2013.00044> PMID: 23467856
6. Turitto VT. Blood viscosity, mass transport, and thrombogenesis. *Prog Hemost Thromb.* 1982; 6: 139–177. PMID: 6762611
7. Fidler IJ. Timeline—the pathogenesis of cancer metastasis: the ‘seed and soil’ hypothesis revisited. *Nat Rev Cancer.* 2003 Jun; 3 (6): 453–458. <https://doi.org/10.1038/nrc1098> PMID: 12778135
8. Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. *J Clin Invest.* 2009 Jun; 119: 1420–1428. <https://doi.org/10.1172/JCI39104> PMID: 19487818
9. Kosla J, Pankova D, Plachy J, Tolde O, Bicanova K, Dvorak M, et al. Metastasis of aggressive amoeboid sarcoma cells is dependent on Rho/ROCK/MLC signaling. *J Cell Commun Signal.* 2013 Jul; 11: 51. <https://doi.org/10.1186/1478-811X-11-51> PMID: 23899007
10. Ren ZX, Yu HB, Li JS, Shen JL, Du WS. Suitable Parameter Choice on Quantitative Morphology of A549 Cell in Epithelial-Mesenchymal Transition. *Biosci Rep.* 2015 Jun; 35 (3): e00202. <https://doi.org/10.1042/BSR20150070> PMID: 26182364
11. Guttilla IK, Phoenix KN, Hong X, Tirnauer JS, Claffey KP, White BA. Prolonged mammosphere culture of MCF-7 cells induces an EMT and repression of the estrogen receptor by microRNAs. *Breast Cancer Res Treat.* 2012; 132: 75–85. <https://doi.org/10.1007/s10549-011-1534-y> PMID: 21553120
12. Pasqualato A, Lei V, Cucina A, Dinicola S, D’Anselmi F, Proietti S, et al. Shape in Migration. *Cell Adh Migr.* 2013; 5: 450–459. <https://doi.org/10.4161/cam.26765> PMID: 24176801
13. Pasqualato A, Palombo A, Cucina A, Mariggio MA, Galli L, Passaro D, et al. Quantitative Shape Analysis of Chemoresistant Colon Cancer Cells: Correlation between Morphotype and Phenotype. *Exp Cell Res.* 2012 Apr; 318 (7): 835–846. <https://doi.org/10.1016/j.yexcr.2012.01.022> PMID: 22342954
14. Alizadeh E, Lyons SM, Castle JM, Prasad A. Measuring Systematic Changes in Invasive Cancer Cell Shape Using Zernike Moments. *Integr Biol.* 2016 Nov; 8 (11): 1183–1193. <https://doi.org/10.1039/C6IB00100A> PMID: 27735002
15. Lyons SM, Mannheimer J, Schuamberg K, Castle J, Schroder B, Turk P, et al. Changes in Cell Shape Are Correlated with Metastatic Potential in Murine and Human Osteosarcomas. *Biol Open.* 2016 Feb; 5 (3): 289–299. <https://doi.org/10.1242/bio.013409> PMID: 26873952
16. Oliver CR, Altemus MA, Westerhof H, Cherivan H, Cheng X, Dzbjubinski M, et al. A Platform for Artificial Intelligence based identification of the extravasation potential of cancer cells into the brain metastatic niche. *Lab Chip.* 2020; 19: 1162–1173. <https://doi.org/10.1039/C8LC01387J>
17. Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. *Nature.* 2013; 501: 328–337. <https://doi.org/10.1038/nature12624> PMID: 24048065
18. Abbott A. Cell culture: Biology’s new dimension. *Nature.* 2003; 424: 870–872. <https://doi.org/10.1038/424870a> PMID: 12931155
19. Cohen AA, Geva-Zatorsky N, Eden E, Frenkel-Morgenstern M, Issaeva I, Sigal A, et al. Dynamic Proteomics of Individual Cancer Cells in Response to a Drug. *Science.* 2008; 322: 1511–1516. <https://doi.org/10.1126/science.1160165> PMID: 19023046
20. Efting D, Schrage YM, Geirnaerd MJA, Le Cessie S, Taminiau AHM, Bovee JVMG, et al. Assessment of interobserver variability and histologic parameters to improve reliability in classification and grading of central cartilaginous tumors. *Am J Surg Pathol.* 2009; 33: 50–57. <https://doi.org/10.1097/PAS.0b013e31817eec2b>
21. Tay S, Hughey JJ, Lee TK, Lipniacki T, Quake SR, Covert MW. Single-cell NF-kappaB dynamics reveal digital activation and analogue information processing. *Nature.* 2010; 466: 267–271. <https://doi.org/10.1038/nature09145> PMID: 20581820
22. Lahav G, Rosenfield N, Sigal A, Geva-Zatorsky N, Levine AJ, Elowitz MB, et al. Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nat Genet.* 2004; 36: 147–150. <https://doi.org/10.1038/ng1293> PMID: 14730303

23. Engers R. Reproducibility and reliability of tumor grading in urological neoplasms. *World J Urol.* 2007; 25 (6): 595–605. <https://doi.org/10.1007/s00345-007-0209-0> PMID: 17828603
24. Wu PH, Gilkes DM, Phillip JM, Narkar A, Cheng TWT, Marchand J, et al. Single-cell morphology encodes metastatic potential. *Sci Adv.* 2020 Jan; 6 (4). <https://doi.org/10.1126/sciadv.aaw6938> PMID: 32010778
25. Wu PH, Phillip JM, Khatau SB, Chen WC, Stirman J, Rosseel S, et al. Evolution of cellular morpho-phenotypes in cancer metastasis. *Sci Rep.* 2015 Jan; 5: 18437. <https://doi.org/10.1038/srep18437> PMID: 26675084
26. Alizadeh E, Castle J, Quirk A, Taylor CMD, Xu W, Prasad A. Cellular morphological features are predictive markers of cancer cell state. *Comput Biol Med.* 2020 Nov; 126: 104044. <https://doi.org/10.1016/j.combiomed.2020.104044> PMID: 33049477
27. Joshi K, Javani A, Park J, Velasco V, Xu B, Razorenova O, et al. Machine Learning-Assisted Nanoparticle-Printed Biochip for Real-Time Single Cancer Cell Analysis. *Adv Biosys.* 2020; 4, 2000160. <https://doi.org/10.1002/adbi.202070115> PMID: 33025770
28. Hasan MR, Hassan N, Khan R, Kim YT, Iqbal SM. Classification of cancer cells using computational analysis of dynamic morphology. *Comput Meth Prog Bio.* 2018; 156, 105–112. <https://doi.org/10.1016/j.cmpb.2017.12.003> PMID: 29428061
29. Elbez R, McNaughton BH, Patel L, Pienta KJ, Kopelman R. Nanoparticle Induced Cell Magneto-Rotation: Monitoring Morphology, Stress and Drug Sensitivity of a Suspended Single Cancer Cell. *PLoS ONE.* 2011; e0028475 <https://doi.org/10.1371/journal.pone.0028475> PMID: 22180784
30. Zhou Z, Alvarez D, Milla C, Zare RN. Proof of Concept for identifying cystic fibrosis from perspiration samples. *PNAS.* 2019; 116 (49): 24408–24412. <https://doi.org/10.1073/pnas.1909630116> PMID: 31740593
31. Elbez R. Nanoparticle Induced Cell Magneto-Rotation for the Multiplexed Monitoring of Morphology, Stress, and Drug Sensitivity of Suspended Single Cancer Cells. PhD Thesis, The University of Michigan. 2015.
32. Folz J. Frontiers of Cancer Diagnostics: From Photoacoustic Chemical Imaging to Cellular Morphodynamics. PhD Thesis, The University of Michigan. 2020.
33. Freund Y, Schapire RE. A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence.* 1999; 14: 771–780.
34. Roca H, Hernandez J, Weidner S, McEachin RC, Fuller D, Sud S, et al. Transcription Factors OVOL1 and OVOL2 Induce the Mesenchymal to Epithelial Transition in Human Cancer. *PLoS One.* 2013 Oct; 8 (10): e76773. <https://doi.org/10.1371/journal.pone.0076773> PMID: 24124593
35. Steinley D. K-means clustering: A half-century synthesis. *Br J Math Stat Psychol.* 2010; 59: 1–34. <https://doi.org/10.1348/000711005X48266>
36. Chen HC. Boyden Chamber Assay. *Methods Mol. Biol.* 2005; 294: 15–22. PMID: 15576901
37. Park MH, Reategui E, Li W, Tessier SN, Wong KHK, Jensen AE, et al. Enhanced Isolation and Release of Circulating Tumor Cells Using Nanoparticle Binding and Ligand Exchange in a Microfluidic Chip. *J Am Chem Soc.* 2017; 139, 2741–2749. <https://doi.org/10.1021/jacs.6b12236> PMID: 28133963
38. Andree KC, van Dalum G, Testappen LWMM. Challenges in circulating tumor cell detection by the Cell-Search system. *Mol Oncol.* 2016 Mar; 10 (3), 295–407 <https://doi.org/10.1016/j.molonc.2015.12.002>
39. Giri A, Baipal S, Trenton N, Jayatilaka H, Longmore GD, Wirtz D. The Arp2/3 complex mediates multi-generation dendritic protrusions for efficient 3-dimensional cancer cell migration. *FASEB J.* 2013; 27: 4089–4099. <https://doi.org/10.1096/fj.12-224352> PMID: 23796785
40. Wolf K, Mazo I, Leung H, Engelke K, von Andrian UH, Deryugina EI, et al. Compensation mechanism in tumor cell migration: mesenchymal-amoeboid transition after blocking of pericellular proteolysis. *J Cell Biol.* 2003; 160: 267–277. <https://doi.org/10.1083/jcb.200209006> PMID: 12527751
41. Asai A, Konno M, Ozaki M, Kawamoto K, Chijimatsu R, Kondo N, et al. Scent test using *Caenorhabditis elegans* to screen for early-stage pancreatic cancer *Oncotarget.* 2021; 12 (17): 1687–1696. <https://doi.org/10.18632/oncotarget.28035> PMID: 34434497
42. Das A, Haque I, Ray P, Ghosh A, Dutta D, Quadir M, et al. CCN5 activation by free or encapsulated EGCG is required to render triple-negative breast cancer cell viability and tumor progression. *Pharmacol Res Perspect.* 2012; 9: e00753.
43. Abdullah CS, Ray P, Alam S, Kale N, Aishwarya R, Morshed M. Chemical Architecture of Block Copolymers Differentially Abrogate Cardiotoxicity and Maintain the Anticancer Efficacy of Doxorubicin. *Mol Pharmaceutics.* 2020; 17: 4676–4690. <https://doi.org/10.1021/acs.molpharmaceut.0c00963> PMID: 33151075
44. Campbell PJ, Yachida S, Mudie LJ, Stephens PJ, Pleasance ED, Stebbings LA, et al. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature.* 2010; 467: 1109–1113. <https://doi.org/10.1038/nature09460> PMID: 20981101

45. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*. 2013; 501: 338–345. <https://doi.org/10.1038/nature12625> PMID: [24048066](https://pubmed.ncbi.nlm.nih.gov/24048066/)
46. Park JY, Morgan M, Sachs AN, Samorezov J, Teller R, Shen Y, et al. Single cell trapping in larger micro-wells capable of supporting cell spreading and proliferation. *Microfluid Nanofluidics*. 2009; 8: 263–268. <https://doi.org/10.1007/s10404-009-0503-9>
47. Stuurman N, Edelstein AD, Amodai N, Hoover KH, Vale RD. Computer Control of Microscopes using μ Manager. *Curr Protoc Mol Biol*. 2010 Oct. <https://doi.org/10.1002/0471142727.mb1420s92> PMID: [20890901](https://pubmed.ncbi.nlm.nih.gov/20890901/)
48. Carpenter A, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol*. 2006; 7. <https://doi.org/10.1186/gb-2006-7-10-r100> PMID: [17076895](https://pubmed.ncbi.nlm.nih.gov/17076895/)
49. Cawley GC, Talbot NLC. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *J Mach Learn Res*. 2010; 11: 2079–2107.
50. Gneiting T, Raftery AE. Strictly Proper Scoring Rules, Prediction, and Estimation. *J Am Stat Assoc*. 2007; 102: 359–378. <https://doi.org/10.1198/016214506000001437>
51. MacQueen J. Some methods for classification and analysis of multivariate observations. *Proc. 5th Berkeley Symp. On Math. Statist. And Prob.* 1967; 1: 281–297.