Single cell classification using statistical learning on mechanical properties measured by mems tweezers

Bahram Ahmadian, Deborah Mbujamba, Jean Claude Gerbedoen, Momoko Kumemura, Hiroyuki Fujita, Dominique Collard, Sophie Dabo-Niang, Chann Lagadec, Mehmet Tarhan

SINGLE CELL CLASSIFICATION USING STATISTICAL LEARNING ON MECHANICAL PROPERTIES MEASURED BY MEMS TWEEZERS

Bahram Ahmadian[1,2], Deborah Mbujamba[3], Jean-Claude Gerbedoen[2,3], Momoko Kumemura[4], Hiroyuki Fujita[5,6], Dominique Collard[2,3], Sophie Dabo-Niang[7], Chann Lagadec[8] and Mehmet C. Tarhan[1,2,3]

[1]Univ. Lille, CNRS, Centrale Lille, Polytechnique Hauts-de-France, Junia, UMR 8520-IEMN, Villeneuve d'Ascq, FRANCE

[2]CNRS, IIS, COL, Univ. Lille SMMiL-E Project, Lille, FRANCE

[3] LIMMS/CNRS-IIS, UMI 2820, The University of Tokyo, Lille, FRANCE

[4]Kyushu Institute of Technology, Kita-Kyushu, JAPAN

[5]Tokyo City University, Tokyo, JAPAN

[6]Canon Medical Systems Corporation, Tochigi, JAPAN

[7]Univ. Lille, CNRS, UMR 8524 - Laboratoire Paul Painlevé, 59000, Lille, FRANCE

[8]Univ. Lille, CNRS, Inserm, CHU Lille, Centre Oscar Lambret, UMR9020 – UMR-S 1277 - Canther – Cancer Heterogeneity, Plasticity and Resistance to Therapies, F-59000 Lille, FRANCE

ABSTRACT

Cell population is heterogenous and so presents a wide range of properties as metastatic potential. But using rare cells for clinical applications requires precise classification of individual cells. Here, we propose a multi-parameter analysis of single cells to classify them using statistical learning techniques and to predict the sub-population of each cell, although they may have close characteristics. We used MEMS tweezers to analyze mechanical properties (stiffness, viscosity, and size) of single cells from two different breast cancer cell lines in a controlled environment and run supervised learning methods to predict the population they belong to. This label-free method is a significant step forward to distinguish rare cell sub-populations for clinical applications.

KEYWORDS

Single-cell characterization, supervised learning, MEMS tweezers, cancer cell classification

INTRODUCTION

Cancer is a disease exhibiting irregular and uncontrolled growth of cells and being able to spray into various parts of the body. Several tools have been developed as powerful recognition probes (*e.g.*, immunostaining/capture) in the specific detection of cancer cells. However, the high cancer cell heterogeneity and the high complexity of the detection process cause important restrictions for using them for patient applications [1]. Among different cancer cell biomarkers, biophysical properties are excellent candidates as they do not require any staining processes and can be measured instantaneously. Indeed, biophysical properties of the cancer cells, such as size and mechanical properties, are attractive as a potential to understand the general behavior of these cells and tissues. Biophysical studies have demonstrated that cancer progression involves changes in cell morphology and mechanical properties [2,3]. Accordingly, investigation of the biophysical properties of cancer cells can contribute to distinguishing different cancer cell lines.

The single-cell analysis allows measuring biophysical properties of single cancer cells like cell size and mechanical properties. Each cell must be characterized for practical clinical purposes, and its metastatic potential must be predicted individually. This requirement is a great challenge due to the high heterogeneity of cells. Therefore, we target obtaining the mechanical properties of cells with a practical and sensitive MEMS device to apply supervised learning methods for distinguishing individual cells.

METHODS

We perform the classification of cancer cells according to their biophysical property in two steps: single-cell measurement and data analysis using statistical learning. We integrated MEMS tweezers with a microfluidic device for single-cell measurements to characterize and distinguish different cancer cell lines.

Setup

The experiments were performed on the stage of a combined inverted and upright microscopy. MEMS tweezers were placed on the stage and a microfluidic device was positioned with respect to the tweezers using a robotic arm, similar to previous attempts [4]. The inverted microscope provided high-resolution cell visualization (brightfield and fluorescence) with its 60x objective while the upright microscope objective independently monitored tweezers during the cell capturing process (Figure 1b). Experiments were performed in a controlled environment: 37˚C, 5% CO_2 and 80% relative humidity (Figure 1c&d).

MEMS Tweezers

The MEMS Tweezers, fabricated with fundamental micromachining techniques [5], has cell handling tips, a sensing side, and a compression side. Protruding tips capture a cell for mechanical stimulation based on electrostatic actuation provided by interdigitated comb architecture. A cell captured between the tips is pushed with the compressing tip (driven with a function generator and a power amplifier) towards the sensing tip during continuous real-time measurements. Changes in the resonance frequency and the displacement amplitude

measured with the integrated displacement sensor (driven with a lock-in amplifier and two transimpedance amplifiers) provides the information on the mechanical properties of the captured cell.

Microfluidic devices

A microfluidic channel mold structure was fabricated using SU8 lithography. The molded polydimethyl siloxane (PDMS) device had an inlet connected to an outlet with a channel having a side opening (200 µm x 110 µm). The inlet and outlet were created using biopsy punchers with a diameter of 2 mm and 0.5 mm respectively. The outlet was connected to a vacuum pump working in withdraw mode for inducing a flow. After injecting a cell suspension at the inlet, the flow controlled the cell motion in the channel to position them close to the side opening. The side opening was designed to be small enough to avoid any leakage while keeping a large enough area for MEMS tweezers to maneuver during cell capturing. The microfluidic device was positioned on a motorized stage which enabled us to the precise positioning of it during the cell characterization.

Cell solution

We tested two breast cancer cell lines with different characteristics. MCF7 is an epithelial-like cell with luminal, oestrogen receptor and progesterone-receptor positive, while SUM159-PT is a mesenchymal-like cell of triple-negative type. Compared to MCF7's low metastatic properties, SUM159-PT is very aggressive and highly metastatic.

Experimental procedure

Experiments are performed in several steps as follow: First, the MEMS tweezers are inserted into the side opening of the microfluidic device. Secondly, a cell suspension is inserted via the inlet and positioned by the side opening with the flow controlled by the vacuum pump connected to the outlet. Then, we position MEMS tweezers tips around a cell and start the compression assay by closing the gap between tips. The sensing tip detects the cell when the gap between tips is equal to the size of the cell. With this size information, the LabVIEW program automatically applies a potential difference between the electrodes of the compression side actuator to provide 30% compression (Figure 2a). The real-time continuous measurements provides the required information on calculating the analyzed cell to restart the procedure for another cell. mechanical properties of the cell. Finally, we release the analyzed cell, and restart the procedure for another cell.

Working principle

MEMS tweezers are characterized as a mechanical mass-spring system integrating transducers in both compression and sensing parts. Accordingly, a dynamic model was proposed to measure the mechanical properties of a trapped object by monitoring the changes in the frequency response and amplitude [6]. During cell compression, MEMS tweezers detect an increase in the resonance frequency and a decrease in the displacement amplitude value. This increase in the resonance frequency is related to the stiffness of the cell (adding extra spring by cell to the MEMS tweezers mass-spring system) and the decrease in the amplitude is linked to the viscosity of the cell (absorbing energy by damping part of the cell).

In the experiments, the resonance frequency (f_R) and quality factor (Q) of the tweezers with immersed tips are recorded before capturing a cell. Equations (1) and (2) allows calculating the stiffness (K), and viscous losses (η) of the MEMS tweezers (before capturing a cell, K_{cell} and η_{cell} is zero). During a compression assay, the LabVIEW program controls a lock-in amplifier performing phase-lock loops (PLL) to drive the MEMS tweezers at its resonance frequency and records the changes in the resonance frequency and amplitude in real-time. Equations (1) and (2) allows calculation of the stiffness (k_{cell}) and viscous losses (η_{cell}) of the cancer cell [6].

$$Q(t) = \frac{\sqrt{(K+K_{Cell}(t))M}}{\eta+\eta_{cell}(t)} \quad (1)$$

$$f_R(t) = \frac{1}{2\pi}\sqrt{\frac{K+K_{Cell}(t)}{M}} \quad (2)$$

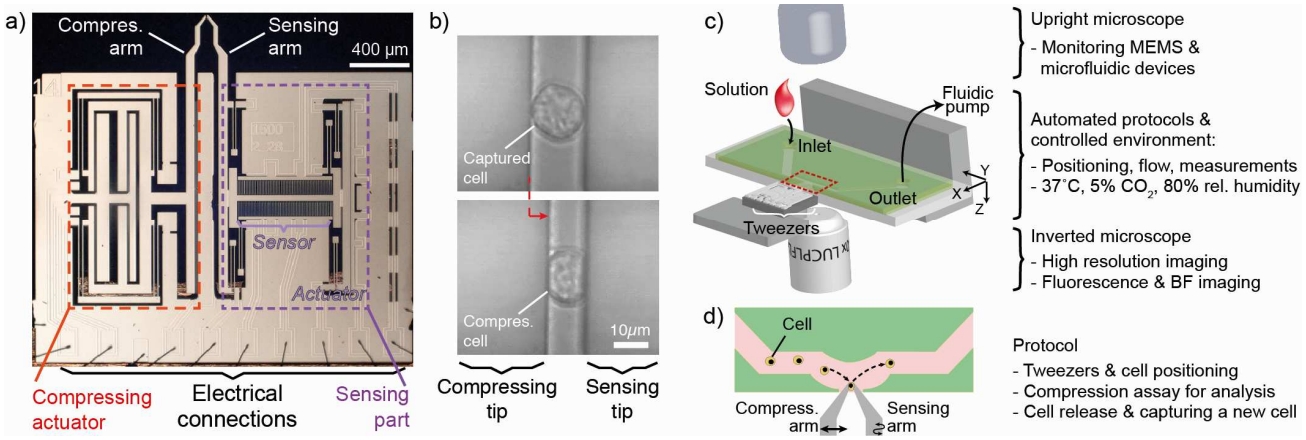

Figure 1: a) MEMS tweezers has a compressing tip connected to an actuator and a sensing tip connected to a sensor/actuator couple. b) A compression assay is performed on captured cells. c) Experiments are performed in a controlled environment on a microscope stage modified to provide both epi- and trans-illumination. d) Tweezers enter a microfluidic channel via a side-opening to perform compression assays. A flow controlled by an outlet pump releases analyzed cells to bring new ones.

Functional Data Analysis

The statistical methods for multivariate data encounter difficulties involving high dimension. This is mainly because handling infinite dimensions of variables of functional forms (curves, shapes, and so on) can be complicated. Functional data analysis (FDA) provides an adequate methodology to study data of functional forms such as the data of interest. It deals with the analysis and theory of data that are in the form of functions, curves, images and shapes, or more general complex mathematical objects, thought of as smooth realizations of a stochastic process.

Classification: supervised learning

Classification aims to assign an individual to a pre-determined group or class based on class-labeled observations. Classical classification or discrimination is about predicting the unknown nature of an object, a discrete quantity for example metastatic or non-metastatic. An object is a collection of numerical measurements such as viscosity and stiffness.

The unknown nature of the object is called a class and is denoted by Y which takes values in a finite set {1,…,M}. In classification, one constructs a function g taking values in {1,…,M} which represents one's guess $g(X)$ of Y given X. The mapping g is called a classifier. We aim to predict the class Y from X for a given cell using a sample of this pair of variables at some other cells (named training sample). The mapping g is then defined on $\mathbb{R}^d$ and takes values in {1,…,M}. We err on Y if $g(X) \neq Y$, and the probability of error for a classifier g is given by:

$$L(g) = P\{g(X) \neq Y\} \quad (3)$$

It is well known that the Bayes classifier, defined by:

$$g = \underset{g:x\to\{1,\dots,M}{argmin} P\{g(X) \neq Y\} \quad (4)$$

Which is the best possible classifier, with respect to quadratic loss. The minimum probability of error is called the Bayes error and is denoted by $L = L(g)$. Note that g depends upon the distribution of (X,Y) which is unknown. An estimator g_n of g is based on the training observations $\{(X_i, Y_i)_{i\in On}\}$; Y is predicted by g_n. The performance of g_n is measured by the conditional probability of error:

$$L_n = L(g_n) = P\{g_n(X; (X_i, Y_i)_{i\in On}) \neq Y\} \geq L \quad (5)$$

The sequence $\{ g_n, n \in N\}$ is the discrimination rule. It has been investigated extensively in the statistical with several parametric and non-parametric methods [7].

RESULTS AND DISCUSSION

We performed the compression assay protocol on each cell and measured changes in the resonance frequency and amplitude values for single-cell biophysical properties. During the compression assay, the LabVIEW software monitored the resonance frequency and amplitude which are related to the mechanical property of the cell (stiffness and viscosity) in addition to the initial measurement on the cell dimension. Due to the calibrated tip displacement, the LabVIEW program provides the mechanical properties as a function of time or the gap between the tips and thus, the compression rate (Figure 2a).

We performed compression assays on two different breast cancer cell lines, *i.e.* SUM159-PT and MCF7 (n=54 and 51, respectively), to obtain their size, stiffness (at 20% deformation), and viscosity (at 20% deformation) (Figure 2b). The triple negative SUM159-PT cell line has higher metastatic potential compared to MCF7 cell line and previous research [8] showed that cancer cells exhibit softer characteristics compared to their benign counterparts. The comparison of the average stiffness values at 20% deformation between MCF7 and SUM159-PT (Figure 2b) indicated that SUM159-PT cell line (having higher metastatic potential) was softer than MCF7 cell line (having lower metastatic potential). Similarly, SUM159-PT was less viscous than the MCF7 (Figure 2b). For all parameters, *i.e.* size, stiffness and viscosity, the SUM159-PT and MCF7 cell lines showed significant differences ($p<0.001$ for all parameters).

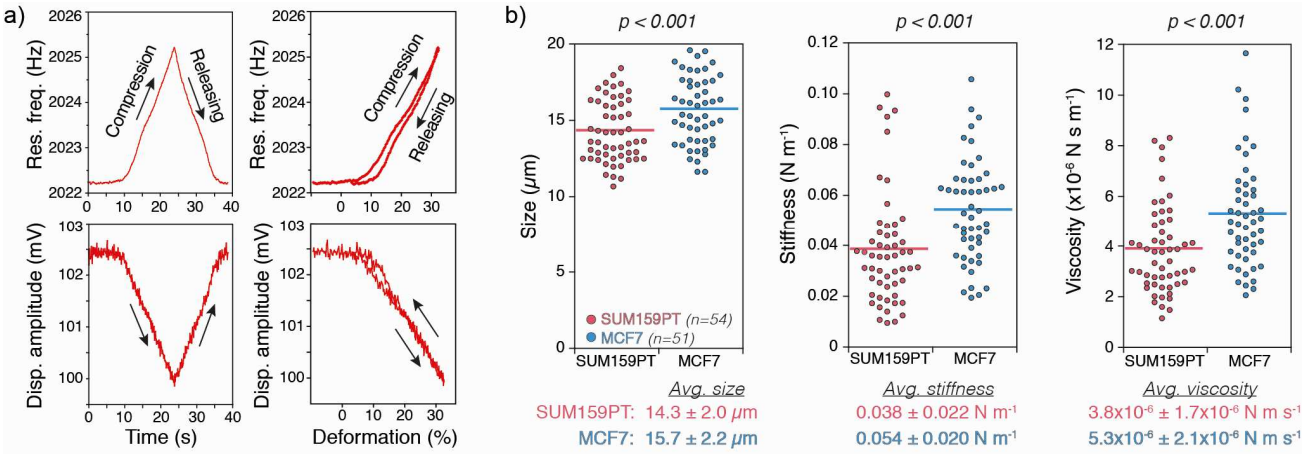

Figure 2: a) Real-time measurements of a compression assay. We measure the resonance frequency and displacement amplitude during a single compression. The results can be shown as a function of time or can they be shown as a function of deformation using the cell size and the gap between the tips. b) Results providing mechanical properties, e.g. size, stiffness and viscosity (at 20% deformation), of highly-metastatic SUM159-PT and less metastatic MCF7 (n=54 and 51, respectively). Although populations have significant differences for each biophysical parameter ($p<0.001$ for all parameters), data ranges overlapping for both cell lines prevent the use for individual cells classification.

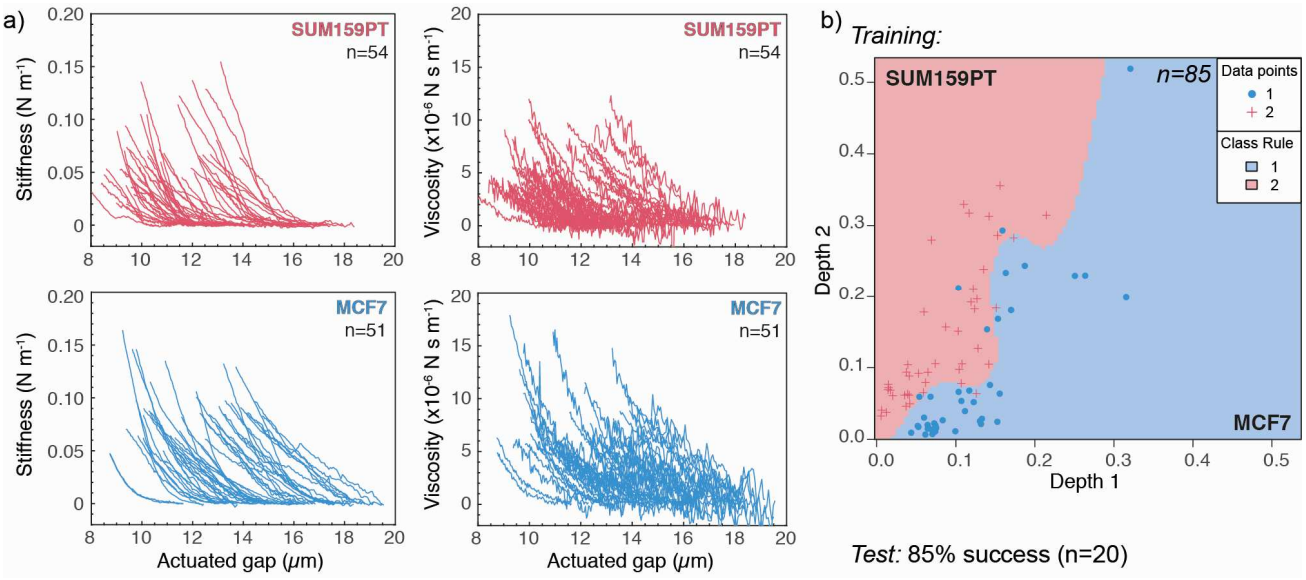

Figure 3: The mechanical property (stiffness and viscosity) with respect to the actuated gap value obtained by compression assays of each cell (a) was used for a training session (generalized additive model analysis) on randomly chosen 85 cells to predict the remaining 20 cells (b). b) Classification of the cancer cell lines according to the mechanical measurement. Blue: MCF7, Red: SUM159PT.

Although populations showed significant differences in mechanical properties, both populations had almost identical data ranges which prohibits using those parameters to distinguish cells individually based on any single parameter. Therefore, we used statistical learning to include all three parameters for a multi-parameter analysis. Moreover, we used functional stiffness and viscosity data with respect to the gap between the tips which included the size information of each cell (Figure 3a). A generalized additive model (GAM) analysis was applied to the splines basis projection of the stiffness and viscosity curves. Among 105 cells, 85 cells were randomly chosen for a training session, and the remaining 20 cells were tested blindly (Figure 3b). The training sample was validated by test data with a correct classification rate of 85% (Figure 3b). Increasing the number of cells for training and including the releasing data can improve the classification and prediction efficiencies.

CONCLUSION

In this study, we proposed using the mechanical property of the single-cell for distinguishing two cell lines using statistical learning. MEMS tweezers performed the single-cell measurement to provide biophysical properties such as stiffness, viscosity, and size of the cancer cells. We tested multiple cells from different cell lines (SUM 159PT and MCF7) for their mechanical characterization. Due to the almost-identical range of mechanical properties in both cell lines, it is not possible to distinguish single cancer cells according to one of their biophysical properties. Therefore, we used generalized additive model analysis for distinguishing these two cell lines. After the classification step, a blind-test predicted 85% of the cells successfully. To sum up, multi-parameter analysis of single cells leads to the practical identification of individual cells by integrating MEMS measurements with statistical learning algorithms.

ACKNOWLEDGEMENTS

This work is in the framework of SMMiL-E activities (a joint project of CNRS, Institute of Industrial Science, Centre Oscar Lambret and University of Lille) and supported by a grant from Contrat de Plan Etat-Région CPER Cancer 2015-2020, and Ruban Rose award.. The authors acknowledge IRCL for hosting SMMiL-E. M. C. Tarhan acknowledge I-SITE ULNE.

REFERENCES

[1] E. M. Hassan and M. C. DeRosa, TrAC - Trends Anal. Chem., vol. 124, p. 115806, 2020, doi: 10.1016/j.trac.2020.115806.

[2] M. Li, D. Dang, L. Liu, N. Xi, and Y. Wang, IEEE Trans. Nanobioscience, vol. 16, no. 6, pp. 523–540, 2017, doi: 10.1109/TNB.2017.2714462.

[3] C. T. Lim, E. H. Zhou, and S. T. Quek, J. Biomech., vol. 39, no. 2, pp. 195–216, 2006, doi: 10.1016/j.jbiomech.2004.12.008.

[4] T. Baetens, G. Perret, et al., IEEE Int. Conf. MEMS (MEMS'17), pp.608-611, 2017.

[5] C. Yamahata et al.J. Microelectromechanical Syst., vol. 17, no. 3, pp. 623–631, 2008, doi: 10.1109/JMEMS.2008.922080.

[6] M. C. Tarhan et al., Sci. Rep., vol. 6, 28001, 2016, doi: 10.1038/srep28001.

[7] James, G., Witten, D., Hastie, T., Tibshirani, R. An introduction to statistical learning (Vol. 112, p. 18). NewYork: springer, 2013.

[8] O. Jonas, C. T. Mierke, and J. A. Käs, Soft Matter, 7, 11488, 2011.

CONTACT

* M.C. Tarhan, cagatay.tarhan@junia.com