Automated bio-AFM generation of large mechanome data set and their analysis by machine learning to classify cancerous cell lines

Ophélie Thomas- -Chemin, Childérick Séverac, Abderazzak Moumen, Adrian Martinez Rivas, Christophe Vieu, Marie-Véronique Le Lann, Emmanuelle Trévisiol, Etienne Dague

To cite this version:

Ophélie Thomas- -Chemin, Childérick Séverac, Abderazzak Moumen, Adrian Martinez Rivas, Christophe Vieu, et al.. Automated bio-AFM generation of large mechanome data set and their analysis by machine learning to classify cancerous cell lines. ACS Applied Materials & Interfaces, 2024, 16 (34), pp.44504-44517. 10.1021/acsami.4c09218 . hal-04674006

HAL Id: hal-04674006
https://hal.science/hal-04674006v1
Submitted on 20 Aug 2024

Automated bio-AFM generation of large mechanome data set and their analysis by machine learning to classify cancerous cell lines

Ophélie Thomas - - Chemin,[1] Childérick Séverac,[1,2] Abderazzak Moumen,[1] Adrian Martinez-Rivas,[3] Christophe Vieu,[1] Marie-Véronique Le Lann,[1] Emmanuelle Trevisiol[1,4] and Etienne Dague[1,]*

[1]LAAS-CNRS, Université de Toulouse, CNRS, 31031, Toulouse France

[2]RESTORE Research Center, Université de Toulouse, INSERM, CNRS, EFS, ENVT, Université P. Sabatier, 31100, Toulouse, France

[3]CIC, Instituto Politécnico Nacional (IPN), 07738, Mexico City, Mexico

[4]TBI, Université de Toulouse, CNRS, INRAE, INSA, 31400, Toulouse, France

*edague@laas.fr

Abstract

Mechanobiological measurements have the potential to discriminate healthy from pathological cells. However, a technology frequently used to measure these properties *i.e.* atomic force microscopy (AFM), suffers from its low output and lack of standardization. In this work, we have optimized AFM mechanical measurement on cell populations and developed a technology combining cell patterning and AFM automation that has the potential to record data on hundreds of cells (956 cells measured for the publication). On each cell, 16 force curves (FCs) and seven features/FC, constituting the mechanome, were calculated. All the FCs were then classified using machine learning tools with a statistical approach based on a fuzzy logic algorithm, trained to discriminate between non-malignant and cancerous cells (training base, up to 120 cells/cell line). The proof of concept was first made on prostate non-malignant (RWPE-1) and cancerous cell lines (PC3-GFP), then on non-malignant (Hs 895.Sk) and cancerous (Hs 895.T) skin fibroblast cell lines and demonstrated the ability of our method to classify correctly 73 % of the cells (194 cells in the data base/cell line) despite the very high degree of similarity of the whole set of measurements (79 to 100 % similarity).

Keywords: AFM, mechanome, cell classification, machine learning, automation

Introduction

Atomic force microscopy (AFM), since its inception in life sciences, has allowed the description of new properties of cell interfaces[1–4]. Nano-mechanics[5], physico-chemical interactions[6] and single molecule interactions[7,8] are some of the fields in which major advances have been made thanks to this technique. From a mechanical point of view, for example, it has been shown that a cancerous cell has different properties than a healthy cell[9], that a normal cartilage tissue is harder than a tissue evolving into osteoarthritis[10], that a cardiomyocyte from a healthy heart is softer than a cardiomyocyte from a heart failure context[11,12]. More recent work by Plodinec *et al.*[13,14] on cancerous biopsies have shown that the nanomechanical signatures of healthy and cancerous tissues are also different.

In this work, we therefore wanted to find out whether it was possible to classify these nanomechanical signatures. Our objective is to be able, on the basis of one measurement (or a small series of measurements), to classify a cell as belonging to one class or another. In this context, machine learning can help[15], but requires a big amount of data, to be trained, and thus we need a method to acquire those big mechanobiological data. This can be achieved by automatizing AFM biomechanical measurements[16,17]. We chose to work on a prostate epithelium cell line, RWPE-1 (isolated prostate epithelial cell from a 54-year-old white patient) and on a prostate cancer cell line, modified to express the GFP fluorescent protein, PC3-GFP (cell line initiated from bone metastasis of grade IV prostatic adenocarcinoma in a 62-year-old white male).

To reach our goal, we trained a machine learning algorithm. ML is a technology that automatically learns from the data it is given. In our case, we based the ML on fuzzy logic, called LAMDA[18–20] to recognize the two cell types. To be performant, this training requires a large amount of data, which is poorly available in conventional bio-AFM. To overcome this difficulty, we have developed a process for automating bio-AFM measurements that combines a strategy of living cells immobilization, on extracellular matrix protein patterns, and a jython script that automatically controls the movement of the AFM stage from cell to cell together with the mechanical AFM measurements (approach and retract force-distance curves: FCs). The measurements have been optimized in terms of cell patterning, indenter geometry (pyramid *versus* sphere), indenter velocity and number of measurements per cell.

By posing the problem of classification, bio-AFM can be used as a real-world technique for diagnostic purposes, for example. Indeed, a diagnosis starts with a population of unknown cells and it is then necessary to find among this population, the cells exhibiting a pathological signature on the basis of their biomechanical features. In this context, the statistical comparison is not valid, because even if the two conditions are statistically different, it is often impossible, on the basis of a single measurement or a series of measurements, to attribute a cell to one or the other condition. Indeed, the fact that the differences are statistically significant does not mean, actually, that one can attribute a measurement to one class or another.

This is where our work differs from most studies involving the measurement of mechanobiological properties by AFM. Fundamental studies are based on the statistical comparison of two populations of cells, supposedly homogeneous. The work consists in determining a biomechanical feature and evaluating its relevance in the context of the discrimination of two cell types. The scientific problem is always formulated in this way. The two cell populations are known, and the researcher tries to show how they are biomechanically different. The comparison is then carried out by statistical tests that show to what degree of certainty it is possible to affirm that condition n°1 is different from condition n°2 on a given set of specific features.

Results

Selection of the parameters for mechanobiological data acquisition

The optimization of the measurements was conducted to meet two objectives: eliminate the heterogeneity caused by the measurement and carry out as many measurements as possible. To eliminate measurement heterogeneity, we determined the influence of immobilizing cells on patterns and indenter geometry on the mechanical signature of the cells (Fig. 1A-L). To achieve a maximum number of measurements, we determined the impact of the indenter velocity and the number of measurements per cell on the mechanical signature of the cells (Fig. 1M-Y). We have observed that cells that had adhered to a 40 x 40 μm^2 square of fibronectin, and had therefore been constrained to adopt the shape of this square, had a lower rigidity than unconstrained cells (from 3.0

± 0.4 nN/µm for cells randomly seeded on fibronectin to 2.3 ± 0.2 nN/µm for square constrained cells) but more importantly have a smaller standard deviation on the measurements performed on several cells (10) from 0.88 ± 0.05 nN/µm to 0.47 ± 0.01 nN/µm.

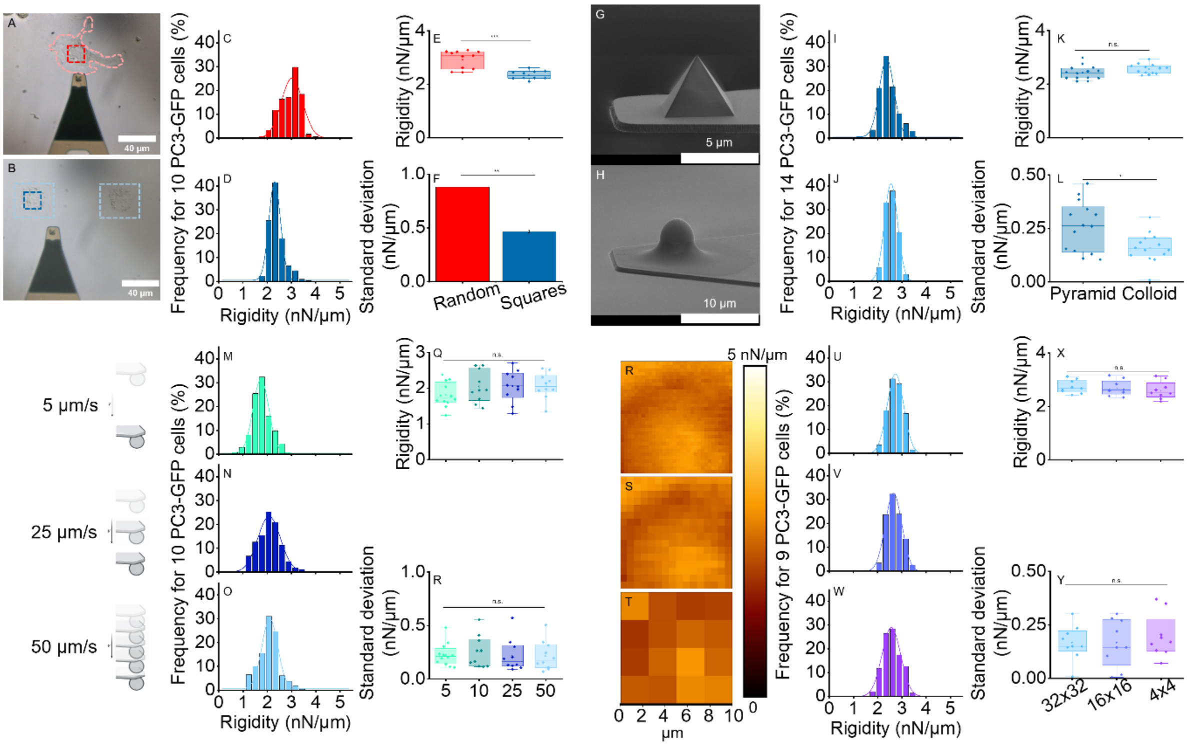


Figure 1|. Influence of 4 parameters (cell patterning, indenter geometry, indenter velocity and number of measurements (FCs)/cell) on the histograms of the stiffness of PC3 cells obtained by force-distance measurements. (**A-F**) Influence of cell patterning on the rigidity (k_{ext}) of PC3-GFP cells. Optical image of a PC3-GFP cell randomly seeded on fibronectin (with no pattern) (**A**) and a PC3-GFP cell in adhesion on a 40 µm x 40 µm fibronectin square pattern (**B**). k_{ext} distributions measured with a pyramidal indenter (**C, D**). Measurements are carried out on 10 cells seeded on a fibronectin coated coverslip (**C**) and on 10 cells arranged on 40 µm-square patterns of fibronectin (**D**). Values of the mean of the k_{ext} (**E**) and the distribution (**F**) are compared (red: cells randomly seeded on fibronectin (with no pattern), blue: cells seeded on-squares patterns of fibronectin). Number of FCs: n_{FCs} = 10 240. (**G-L**) Influence of the indenter geometry on the cells k_{ext}. SEM images of a pyramidal probe (**G**) and of a spherical probe (**H**). Distributions of the k_{ext} measured on 14 PC3-GFP cells arranged on 40 µm-square patterns measured with a pyramidal probe (**I**) or a spherical probe (**J**). Comparison of the k_{ext} mean values (**K**) and of the k_{ext} standard deviations (**L**) (dark blue: pyramid, light blue: sphere) n_{FCs} = 14 336. (**M-Q**) Influence of the indenter velocity on cells k_{ext}. k_{ext} distributions of PC3-GFP cell arranged on 40 µm-square patterns (**M-O**). 10 PC3-GFP cells per panel measured with a spherical indenter at an indentation velocity of respectively 5 (**M**), 25 (**N**) and 50 µm/s (**O**). Comparison of the k_{ext} mean values (**P**) and of the standard deviation (**Q**) at indentation velocities of 5 (light green), 10 (dark green), 25 (dark blue) and 50 µm/s (light blue). Distribution histogram not represented for 10 µm/s. n_{FCs} = 10 240. (**R-Y**) Influence of the number of measurements per cell on k_{ext} of cells. k_{ext} maps of 1 024 (**R**), 256 (**S**) and 16 points (**T**) of the cell surface. Distributions of k_{ext} (PC3-GFP cell populations arranged on 40 µm-square patterns, probed with a sphere at 50 µm/s) of 9 cells per panel and 9 216 (**U**), 2 304 (**V**) and 144 indentations (**W**). Comparison of the k_{ext} (**X**) and standard deviation values (**Y**) for 1 024 (light blue), 256 (dark blue) and 16 indentations (purple). The error bars represent the standard deviation of the measurements. Statistical analysis was performed using a t-test: *** = p-value ≤ 0.001; ** = p-value ≤ 0.005; * = p-value ≤ 0.05; n.s. = not significant.

AFM measurements performed over a larger contact area on the cells give a global mechanical signature and should be not dependent on the exact position of the indenter over the surface. The geometry of the indenter is thus expected to play a role in the dispersion of the results acquired over a population of cells. To evaluate the impact of the indenter geometry and size we compared the rigidity of patterned cells, at the velocity of 50 µm/s, using either pyramidal indenters (radius = 20 nm, half open-angle = 35 °, nominal spring constant = 0.01 N/m, MLCT-BIO-DC, Bruker AFM Probes) or spherical levers (microsphere: 5 µm diameter). The mean cell rigidity (measured as the mean of the slopes of the cell approach curves and named k_{ext}) did not exhibit any significant difference (k_{ext} = 2.4 ± 0.3 nN/µm for a pyramid indenter and k_{ext} = 2.6 ± 0.3 nN/µm for a spherical indenter) with either indenter. However, as expected, the results are significantly less dispersed in the case of the spherical probe (mean standard deviation = 0.16 ± 0.07 nN/µm), compared to the pyramidal probe (mean standard deviation = 0.26 ± 0.12 nN/µm). We have therefore retained the spherical probe for the rest of the study.

In the conditions we tested, the fact of carrying out measurements at higher or lower velocity (from 5 to 50 µm/s) did not change the values of the apparent cell rigidity, and moreover did not affect the dispersion of the results at 5 µm/s (k_{ext} = 1.7 ± 0.3 nN/µm), 10 µm/s (k_{ext} = 2.0 ± 0.5 nN/µm), 25 µm/s (k_{ext} = 2.1 ± 0.5 nN/µm) and 50 µm/s (k_{ext} = 2.1 ± 0.3 nN/µm). Our goal being to measure a large number of cells in a limited time, we have selected a velocity of 50 µm/s for the rest of the study.

Finally, to reduce the measurement time, we evaluated the impact of the number of measurements performed on each patterned cell, at 50 µm/s with a spherical indenter. We compared data from 1 024, 256, 16 (Fig. 1R-Y) and 4 measurements (Supplementary Information 1) per cell. It appears that under these conditions, neither the value of the rigidity (k_{ext} = 2.7 ± 0.3 nN/µm for 1 024, k_{ext} = 2.7 ± 0.3 nN/µm for 256 and k_{ext} = 2.6 ± 0.4 nN/µm for 16 measurements) nor the dispersion of the result (mean standard deviation = 0.15 ± 0.08 nN/µm for 1 024, mean standard deviation = 0.15 ± 0.11 nN/µm for 256 and mean standard deviation = 0.17 ± 0.10 nN/µm for 16 measurements) change between the three sets of measurements with different numbers thereof. To increase the number of cells measured in a given time, we have chosen to limit our experiments to 16 force-distance curves (FCs) per cell.

Automation and analysis of AFM mechanobiological measurements

The next step was to automate the AFM measurements on live cell lines to generate sufficient data for machine learning analysis. We started from the concept developed by Proa-Coronado *et al.* [16,17,21]. In these pioneering works, a smart strategy combining directed immobilisation of microorganisms (in a micro-structured PDMS stamp) and AFM automation (script, compatible with the JPK Nanowizard III AFM[17]) had been developed to acquire data FCs automatically, from cell to cell. The mammalian cells under investigation in our current work being much larger (30-40 µm against only 5 µm for the microorganisms in these articles) we had to use another method to immobilize the cells. We decided to prepare cells arrays (Fig. 2A) following the protocol described in Lagraulet *et al.*[22]. This method is based on microcontact-printing of extracellular matrix proteins (here fibronectin). After passivation by PLL-g-PEG of the area not covered by fibronectin (Fig. 2Af), cells immobilize themselves in a preferential way on the fibronectin squares patterns (40 x 40 µm²), creating cells arrays (Fig. 2B). As far as automation is concerned, we adapted the Proa-Coronado's script[16] to consider larger displacement between cells. Overall the script (Supplementary Information 2 and in a Github open repository: https://github.com/OphelieThomas/Automated-BioAFM.git) automatically performs the following steps (Fig. 2C): 1) Indenter positioning on the centre of the first cell; 2) Indenter positioning on the centre of the last cell; 3) Calculation of the position of each cell on the array; 4) Indenter positioning and FCs acquisition on the first cell; 5) Movement to the next cell and data acquisition and so on until the end of the array (Supplementary Information 3 and Methods).

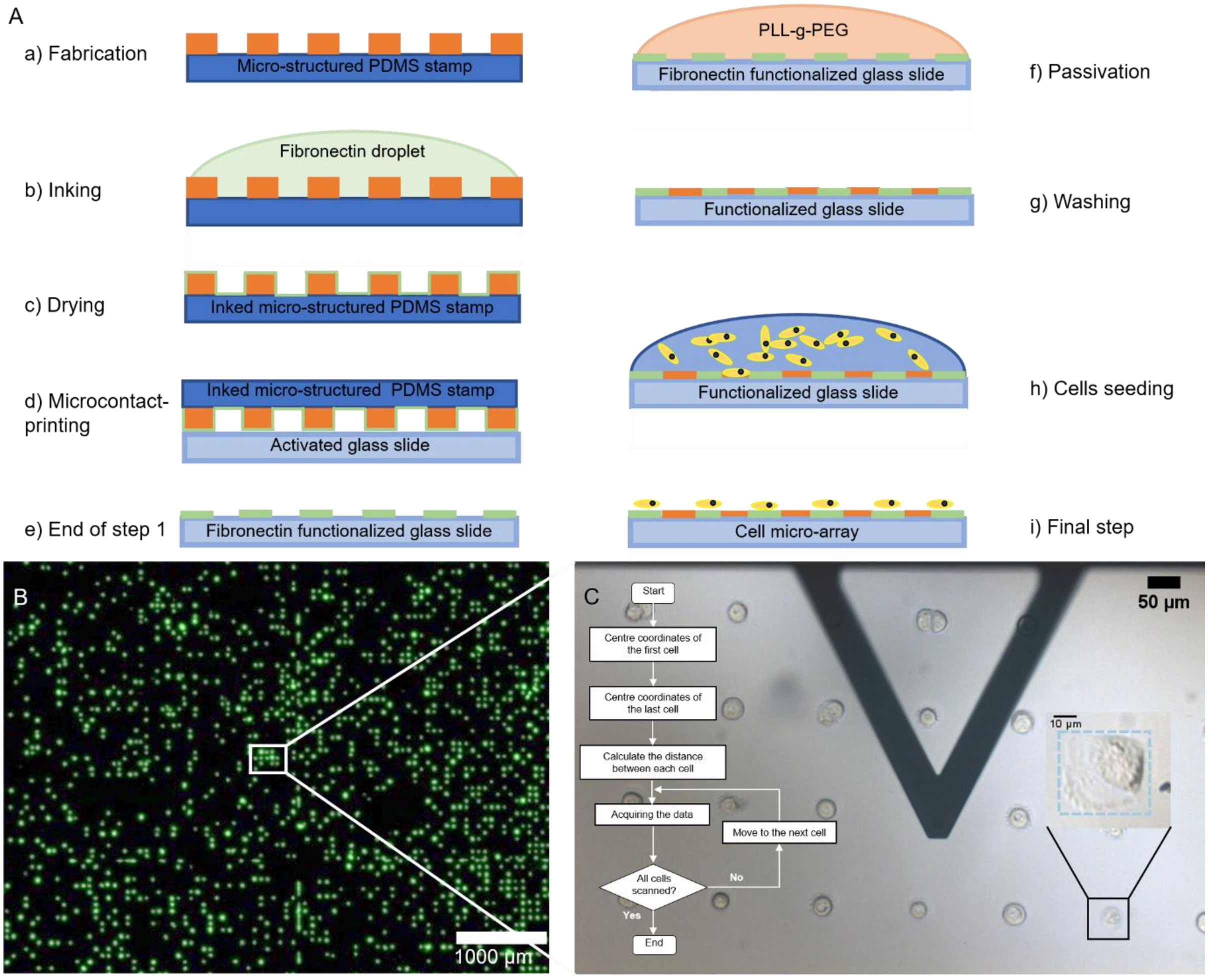


Figure 2|. Preparation of cells array. (**A**) Microcontact-printing steps. The magnetic, micro-structured PDMS stamp (**a**) is inked with extracellular matrix proteins (fibronectin) for 1 min (**b**), then dried (**c**). It is then brought into contact with a plasma-activated glass slide for the microcontact-printing step (**d**), which makes it possible to obtain a slide functionalized with fibronectin (**e**). Subsequently, the non-adhesive molecules of PLL-g-PEG are added to cover the unpatterned areas (**f**), then washed (**g**). Finally, the cells are seeded on the substrate (**h**). After 24 h, the cells are specifically attached to the fibronectin patterns and a cell micro-array is obtained (**i, B**). (**B**) Fluorescence image of PC3-GFP cells having adhered to the fibronectin patterns. (**C**) Steps for automating AFM measurements on an array of cells.

This automatic data acquisition protocol turned out to exhibit a limitation due to the fabrication of cell arrays. During fabrication, as we work with living cells, it can happen that after adhering to the fibronectin patterns, a cell divides, resulting in 2 cells on a micropattern. As a photo is taken before a position is indented, it is possible, at the time of analysis, to eliminate positions occupied by 2 cells. Another limitation was the precision of the automatic sample displacement stage. This implied that the operator needed to regularly reposition the indenter manually on the cells. A precise stage would allow us to overcome this limitation and to accelerate the data acquisition process. Nevertheless, thanks to the regularly positioned cell array, it was quicker to reposition the AFM stage on the cells. In this way, we were able to record data on 244 cells of RWPE-1 and 244 cells of PC3-GFP. On each cell, 16 FCs were recorded and 7 features were measured or calculated (Fig. 3 and Methods). These features are the rigidity measured on the approach curve (k_{ext}), the rigidity measured on the retract curve (k_{ret}), the indenter/cell distance in nanometres of the peak of adhesion relatively to the point of contact (x_{min}), the value in nanonewtons of the peak of adhesion in relation to the baseline of the FC

(f_{adh}), the work of adhesion in femtojoule (W_{adh}), the Young's modulus calculated by the Hertz model adapted to a sphere of 5 µm in diameter (E) and the RMS, *i.e.* the square root in piconewton of the sum of square errors between the model and the indentation curve (see Methods section for details).

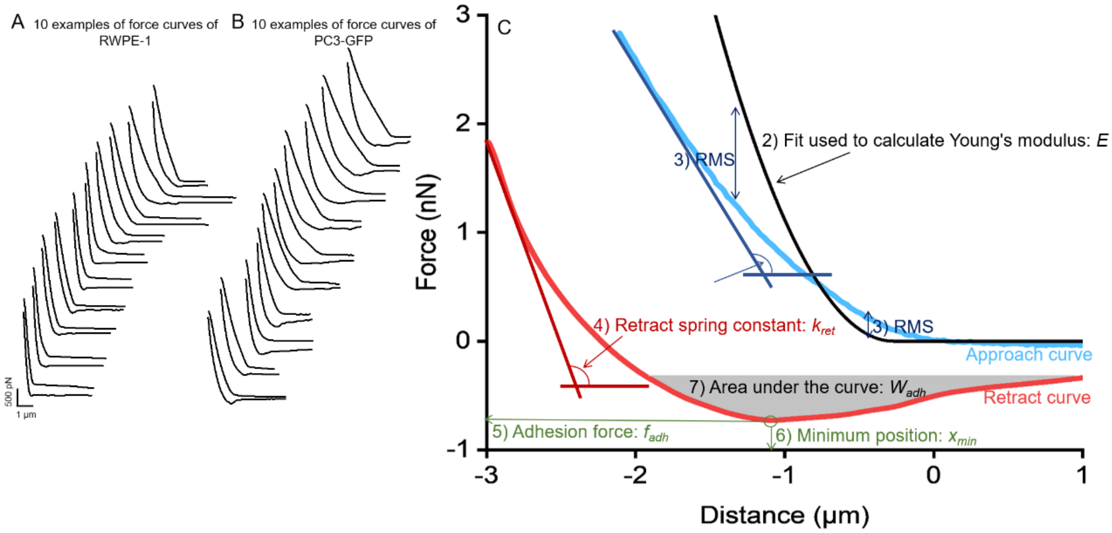


Figure 3|. Cell biomechanical features extracted from the measured FCs and used as inputs for the machine learning algorithm. (**A, B**) Typical FCs (approach and retract) on two population of prostatic cells RWPE-1 (**A**) and PC3-GFP (**B**). (**C**) is a pedagogical force curve representing the 7 features used in the machine learning algorithm. The Young's modulus fit is deliberately very poor to illustrate the RMS feature.

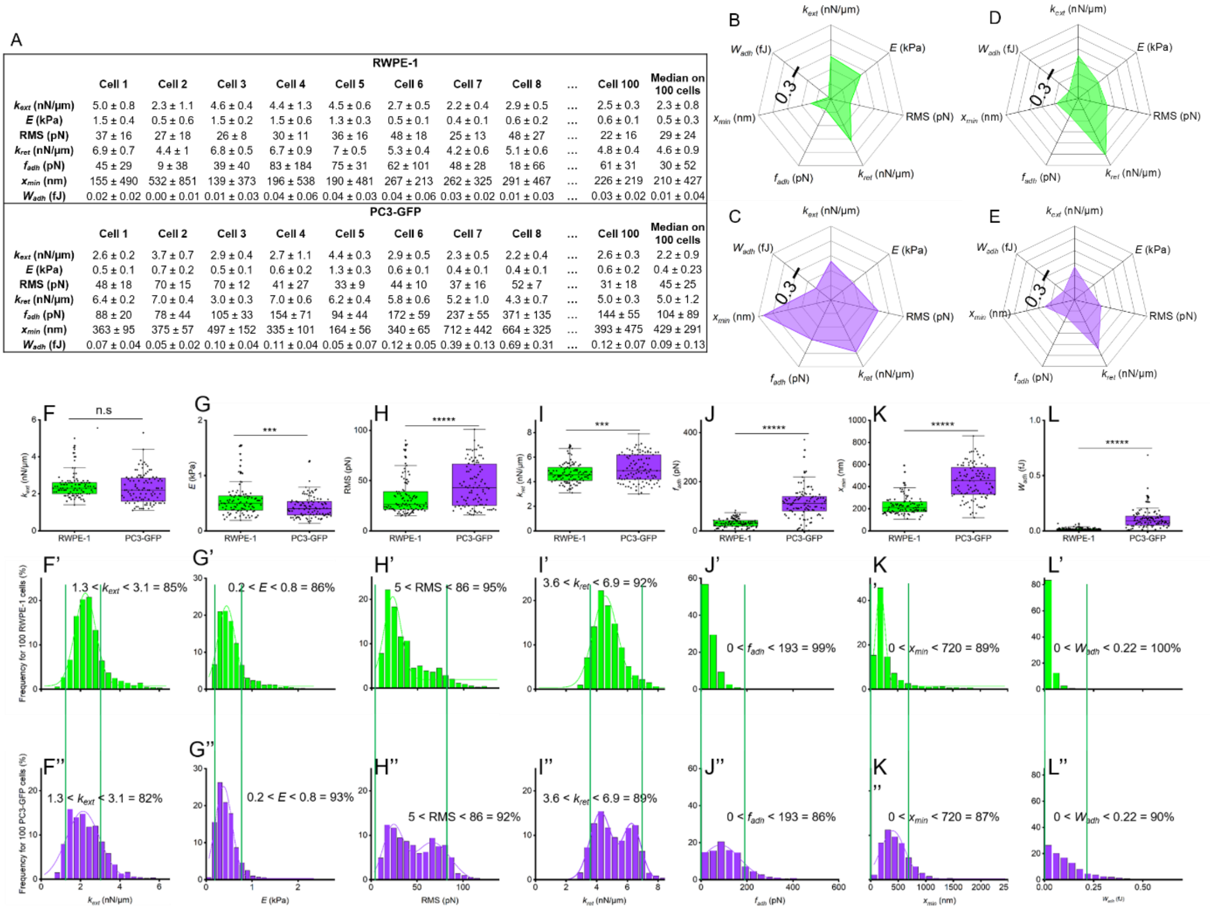
These 7 features were measured or calculated for all the FCs (see Methods), *i.e.* 54 656 values in total, which represents a quantity of data compatible with an analysis by machine learning. All the data, cell by cell are right now available in Supplementary Information 4 and raw data are available in a Zenodo open repository (10.5281/zenodo.12652760).

Mechanome data and its analysis using a dedicated machine learning algorithm based on fuzzy logic

The results obtained on 9 typical cells, the mean of 100 cells and a graphical representation in the form of a spider chart (the values are normalized between 0 and 1 for each feature) of the results obtained on 100 RWPE-1, 100 PC3-GFP, 100 Hs 895.Sk and 100 Hs 895.T are presented in Fig. 4. The conventional comparison of the two cell lines is also presented in Fig. 4, feature by feature. Indeed, the two cell lines are statistically different (except for the rigidity of the approach curve, first column) but it also appears that for each feature, a majority of the data are included, for the two cell lines, between two bounds represented by green vertical lines (Fig. 4F'-L''). The lower bound is the smaller of the two means (RWPE-1 or PC3-GFP) minus standard deviation and the upper bound is the larger of the two means (RWPE-1 or PC3-GFP) plus standard deviation. The lowest percentage of data within these bounds was 82 % for the rigidity on the approach curve, and the highest was 100 % for the area under the curve (see Supplementary Information 5). In Fig. 4, comparisons are made on 40 µm patterns, but we also compared 30 and 20 µm patterns (see Supplementary Information 6). For these smaller patterns, the mean values of k_{ext} and E do not overlap, unlike for the 40 µm patterns. To demonstrate the power of the fuzzy logic algorithm used in this study, we have chosen to work in the most difficult situation for ourselves *i.e.* when the 2 populations overlap (on 40 µm patterns). In this situation, classifying an unknown cell into one group or the other, on the basis of direct comparison, is not possible. The two cell lines are statistically different, but this is not sufficient to establish a diagnosis and classify an unknown cell in one or the other category. To further demonstrate the validity of our method, we tested another pair of cell lines derived from cancerous and non-cancerous skin fibroblasts (Hs 895.T and Hs 895.Sk, respectively). The conventional comparison of these 2 cell lines is presented in Supplementary Information 7, feature by feature. In the same way as the PC3-GFP/RWPE-1 pair, these 2 cell lines are statistically different but it also appears that for each feature, the majority of the data overlap.

For this reason, we decided to train a supervised machine learning algorithm[18–20], named LAMDA and based on fuzzy logic, to discriminate between the two cell lines. We first determined the relevant parameters to perform the classification using another algorithm called MEMBAS[18-19], testing combination of over 20 parameters from the force curves. The results are shown in Supplementary Information 8. This analysis showed that the 7 features presented in Fig. 3 are relevant and enable the best classification. The first step is then, to generate a training data set of sufficient size to allow the algorithm to learn how to discriminate the FCs in an optimum way (80 % of the data set) and to test its performance (20 % of the data set: predicted performance). Each FC (including all 7 features) is correctly labelled as either PC3-GFP or RWPE-1. Supplementary Information 9 presents the predicted performance of the algorithm (grey line) and the actual performance (obtained after training on test data sets of 144 cells) for classification of RWPE-1 (green line) and PC3-GFP (purple line) as a function of the number of cells included in the training. The predicted performance varies little with the number of cells in the training base. However, the actual performance exhibited a different tendency. The performance for PC3-GFP classification is low for a small training base (17 cells: 59 %) and increases for 39 cells (91 %) and then slowly deteriorates for larger training bases (47: 90 %; 69: 86 %; 100: 72 %; 120: 82 %). Meanwhile, for RWPE-1 cells, the actual performance for a 17-cell training base is 89 %, then decreases to 67 % for 39 cells and then increases again with larger bases (47: 70 %; 69: 75 %; 100: 81 %; 120: 78 %). It appears from this analysis that an optimum is reached for a training base of 100 cells, the performances for PC3-GFP and RWPE-1 being close (72 % and 81 % respectively). At 120 cells, the same observation can be made, and it is therefore unnecessary to increase the learning data set beyond 100 cells.

These 100 cells presented Fig. 4, randomly selected from the different days of experimentation (4 different days for RWPE-1 and 9 different days for PC3-GFP) constitute our training database for our machine learning algorithm. We tested several learning algorithms, the fuzzy binomial, the fuzzy centred binomial and the normal gaussian associated with probabilistic or MinMax connectives[18]. The performance of each of these algorithms is summarized in Supplementary Information 10. The normal gaussian, whether with probabilistic connectives or MinMax, had the best predicted classification rate. However, when it was actually tested, on a test data set (containing 144 RWPE-1 cells and 144 PC3-GFP cells, unknown to the algorithm), it associated all the FCs with a single cell line (the PC3-GFP line) and none was classified as RWPE-1, which made it a poor algorithm in this situation. The fuzzy (centred) binomial functions presented interesting performances on our test data set (between 54 % and 95 % depending on the function, connective and cell line considered), especially when they were associated with the probabilistic connective. The fuzzy centred binomial classified the RWPE-1 "unknown" test set better than the fuzzy binomial (95 % and 89 %, respectively), but was not as performant with PC3-GFP (54 % and 67 %, respectively). We therefore chose the binomial fuzzy function coupled with the probabilistic connective which had the advantage of having the best classification performance on the test data set while minimizing the classification difference between the RWPE-1 and PC3-GFP cell lines (89 % and 67 %, respectively). The predicted global performance is 76 % taking into consideration a combination of seven biomechanical features measured or calculated from the AFM FCs.



A

RWPE-1	Cell 1	Cell 2	Cell 3	Cell 4	Cell 5	Cell 6	Cell 7	Cell 8	...	Cell 100	Median on 100 cells
k_{ext} (nN/µm)	5.0 ± 0.8	2.3 ± 1.1	4.6 ± 0.4	4.4 ± 1.3	4.5 ± 0.6	2.7 ± 0.5	2.2 ± 0.4	2.9 ± 0.5	...	2.5 ± 0.3	2.3 ± 0.8
E (kPa)	1.5 ± 0.4	0.5 ± 0.6	1.5 ± 0.2	1.5 ± 0.6	1.3 ± 0.3	0.5 ± 0.1	0.4 ± 0.1	0.6 ± 0.2	...	0.6 ± 0.1	0.5 ± 0.3
RMS (pN)	37 ± 16	27 ± 18	26 ± 8	30 ± 11	36 ± 16	48 ± 18	25 ± 13	48 ± 27	...	22 ± 16	29 ± 24
k_{ret} (nN/µm)	6.9 ± 0.7	4.4 ± 1	6.8 ± 0.5	6.7 ± 0.9	7 ± 0.5	5.3 ± 0.4	4.2 ± 0.6	5.1 ± 0.6	...	4.8 ± 0.4	4.6 ± 0.9
f_{adh} (pN)	45 ± 29	9 ± 38	39 ± 40	83 ± 184	75 ± 31	62 ± 101	48 ± 28	18 ± 66	...	61 ± 31	30 ± 52
x_{min} (nm)	155 ± 490	532 ± 851	139 ± 373	196 ± 538	190 ± 481	267 ± 213	262 ± 325	291 ± 467	...	226 ± 219	210 ± 427
W_{adh} (fJ)	0.02 ± 0.02	0.00 ± 0.01	0.01 ± 0.03	0.04 ± 0.06	0.04 ± 0.03	0.04 ± 0.06	0.03 ± 0.02	0.01 ± 0.03	...	0.03 ± 0.02	0.01 ± 0.04

PC3-GFP	Cell 1	Cell 2	Cell 3	Cell 4	Cell 5	Cell 6	Cell 7	Cell 8	...	Cell 100	Median on 100 cells
k_{ext} (nN/µm)	2.6 ± 0.2	3.7 ± 0.7	2.9 ± 0.4	2.7 ± 1.1	4.4 ± 0.3	2.9 ± 0.5	2.3 ± 0.5	2.2 ± 0.4	...	2.6 ± 0.3	2.2 ± 0.9
E (kPa)	0.5 ± 0.1	0.7 ± 0.2	0.5 ± 0.1	0.6 ± 0.2	1.3 ± 0.3	0.6 ± 0.1	0.4 ± 0.1	0.4 ± 0.1	...	0.6 ± 0.2	0.4 ± 0.23
RMS (pN)	48 ± 18	70 ± 15	70 ± 12	41 ± 27	33 ± 9	44 ± 10	37 ± 16	52 ± 7	...	31 ± 18	45 ± 25
k_{ret} (nN/µm)	6.4 ± 0.2	7.0 ± 0.4	3.0 ± 0.3	7.0 ± 0.6	6.2 ± 0.4	5.8 ± 0.6	5.2 ± 1.0	4.3 ± 0.7	...	5.0 ± 0.3	5.0 ± 1.2
f_{adh} (pN)	88 ± 20	78 ± 44	105 ± 33	154 ± 71	94 ± 44	172 ± 59	237 ± 55	371 ± 135	...	144 ± 55	104 ± 89
x_{min} (nm)	363 ± 95	375 ± 57	497 ± 152	335 ± 101	164 ± 56	340 ± 65	712 ± 442	664 ± 325	...	393 ± 475	429 ± 291
W_{adh} (fJ)	0.07 ± 0.04	0.05 ± 0.02	0.10 ± 0.04	0.11 ± 0.04	0.05 ± 0.07	0.12 ± 0.05	0.39 ± 0.13	0.69 ± 0.31	...	0.12 ± 0.07	0.09 ± 0.13

Figure 4|. Conventional comparison of the 7 biomechanical features measured for the non-malignant (RWPE-1 or Hs 895.Sk) and the malignant (PC3-GFP or Hs 895.T) cell lines. (**A**) Table presenting the median values of the 7 biomechanical features extracted from the AFM measurements (rigidity on approach, Young's modulus, RMS, rigidity on retract, adhesion, minimum position and area under the curve) for the cell line RWPE-1 (top) and PC3-GFP (bottom). Values from 'Cell 1' to 'Cell 100' correspond to single cell values. The column is the median values of the 100 cells database. The individual values of all cells used for the training base (n_{cells} = 100) are in Supplementary Information 4. (**B-E**) Spider charts of the 7 normalized features of the median of 100 RWPE-1 cells (**B**), 100 PC3-GFP cells (**C**), 100 Hs 895.Sk cells (**D**) and 100 Hs 895.T cells (**E**). (**F-L**) Comparison of the mean values of the 7 features measured between a population of RWPE-1 cells (green) and a population of PC3-GFP cells (purple) arranged on 40 µm-square patterns of fibronectin. The error bars represent the standard deviation of the various measurements made. Statistical analysis was performed using a t-test: *** = p-value ≤ 0.005; ** = p-value ≤ 0.01; n.s. = not significant. (**F'-L''**) Distributions of the 7 features measured with a spherical indenter on populations of RWPE-1 (green) and PC3-GFP (purple) cells. n_{cells} = 100; n_{FCs} = 1 600.

This classification was done for each measured FC but, as explained before, we made several measurements per cell (16). We needed to keep in mind that what must be classified *in fine* are cells, *i.e.* batches of 16 FCs. To classify a cell and not just a single FC, we had to choose, arbitrarily or according to an objective, a threshold number above which a set of 16 curves was classified in the RWPE-1 category or in the PC3-GFP category. Overall classification performance for PC3-GFP (Fig. 5A) and RWPE-1 (Fig. 5B) cells was calculated based on the threshold that determined classification. It is important here to discuss what criteria might be objective in determining this threshold. In the context of cancer diagnosis or treatment, it could be very serious to miss a cancerous cell (false negative) while misclassifying a healthy cell into the diseased class would be equally serious (false positive). This means that we want to obtain a test with the best possible sensitivity. In other words, to objectify the threshold, it is therefore necessary that the clinician makes a benefit/risk balance between false negatives and false positives. In a different context, *e.g.* therapeutic follow-up, false positive classifications could be an issue because it would wrongly call into question the effectiveness of a treatment. In this case, the aim is to obtain a test with the best possible specificity. The use of AI algorithms in medicine has raised ethical issues that are discussed in literature[23–26]. In this paper, it is not our intend to address these questions, but it seems legitimate to point out these issues. As an example, we chose arbitrarily to place ourselves in the case where we did not want to miss any cancerous cells, at the risk of misclassifying non-malignant cells as cancerous. To achieve this, we defined a very conservative threshold for cancerous cells at four FCs. When four out of the 16 force-distance measurements per cell are classified as PC3-GFP, we decide that the cell is classified as a PC3-GFP, *i.e.* a cancerous cell. With this threshold, the sensitivity of the test (*i.e.* the rate of correct classification of true positives, light purple, Fig. 5A) was 94 %. The false negative rate, *i.e.* PC3-GFP cells (cancerous) misclassified as RWPE-1 (non-malignant), was only 6 % (dark green, Fig. 5A). This rate increased very rapidly (slope of 1.9 for a threshold number between 0 and 6), while the false positive rate (Fig. 5B, dark purple curve) increased more slowly (slope of 1.4 for a threshold number between 0 and 6). A threshold of only 4 FCs to assign a PC3-GFP cell, leaded in “mirror” to a threshold number of 13 FCs to assign a RWPE-1 cell. This is possible thanks to the slow increases of false positives as a function of the threshold. In doing so, we obtained a specificity of 71 % (this is the rate of correct classification of true negatives, light green, Fig. 5B) and 29 % of RWPE-1 cells were incorrectly assigned to the PC3-GFP category (cancerous), *i.e.* false positives (dark purple, Fig. 5B).

This threshold produced a specific benefit/risk ratio where cancerous cells were recognised at 94 % but conversely the number of false positives was quite high. The same exercise could be carried out with different thresholds, the only constraint being that the sum of the 2 thresholds must be equal to 17. In fact, we found that if this sum is less than 17, some cells may correspond to the 2 categories and an ”indeterminate” class will have to be created. With this in mind, we also classified our pair of Hs 895.Sk/Hs 895.T cell lines. To have a similar number of true positives and true negatives, we decided to set the threshold at 8 FCs per cell classified as Hs. 895.T to classify the entire cell as Hs 895.T. In this way, we were able to correctly classify 63 % of Hs 895.T cells and 69 % of Hs 895.Sk cells (Fig. 5C-D). Naturally, this threshold is arbitrarily assigned to 8 and can be moved to increase the rate of true positives or true negatives. In total, the performance achieved by the algorithm selected with a numerical threshold of 4 or 8, to be correctly classified, must be compared with the similarity rates (between 79 and 100 % depending on the feature concerned) between the different cell lines (see Supplementary Information 5).

Another way to classify cells would be to use the mean of the 16 FCs in each cell. However, using means takes away some freedom in finding solutions to the classification (no adjustable threshold), *i.e.* we can no longer tune the classification as a function of the clinical objectives and pathological context where sometimes false positive can be tolerated or not.

Finally, we observed that there was a difference in the profile between the correct classification of PC3-GFP (Fig. 5A, light purple line) and RWPE-1 (Fig. 5B, light green line). The rate of correct classification drops more rapidly for PC3-GFP than for RWPE-1 as a function of the threshold number of FCs used for the cell classification. This indicated that the biomechanical measurements were probably more homogeneous with RWPE-1 than with PC3-GFP. Indeed, for RWPE-1 cells, the probability of correct recognition was very close for a threshold number between 1 curve/16 (99 %)

and 13 curves/16 (71 %) (Fig. 5B). In contrast, for PC3-GFP the correct recognition for a threshold number of 1 curve/16 is 99 % but only 49 % for a threshold number of 13 curves/16.

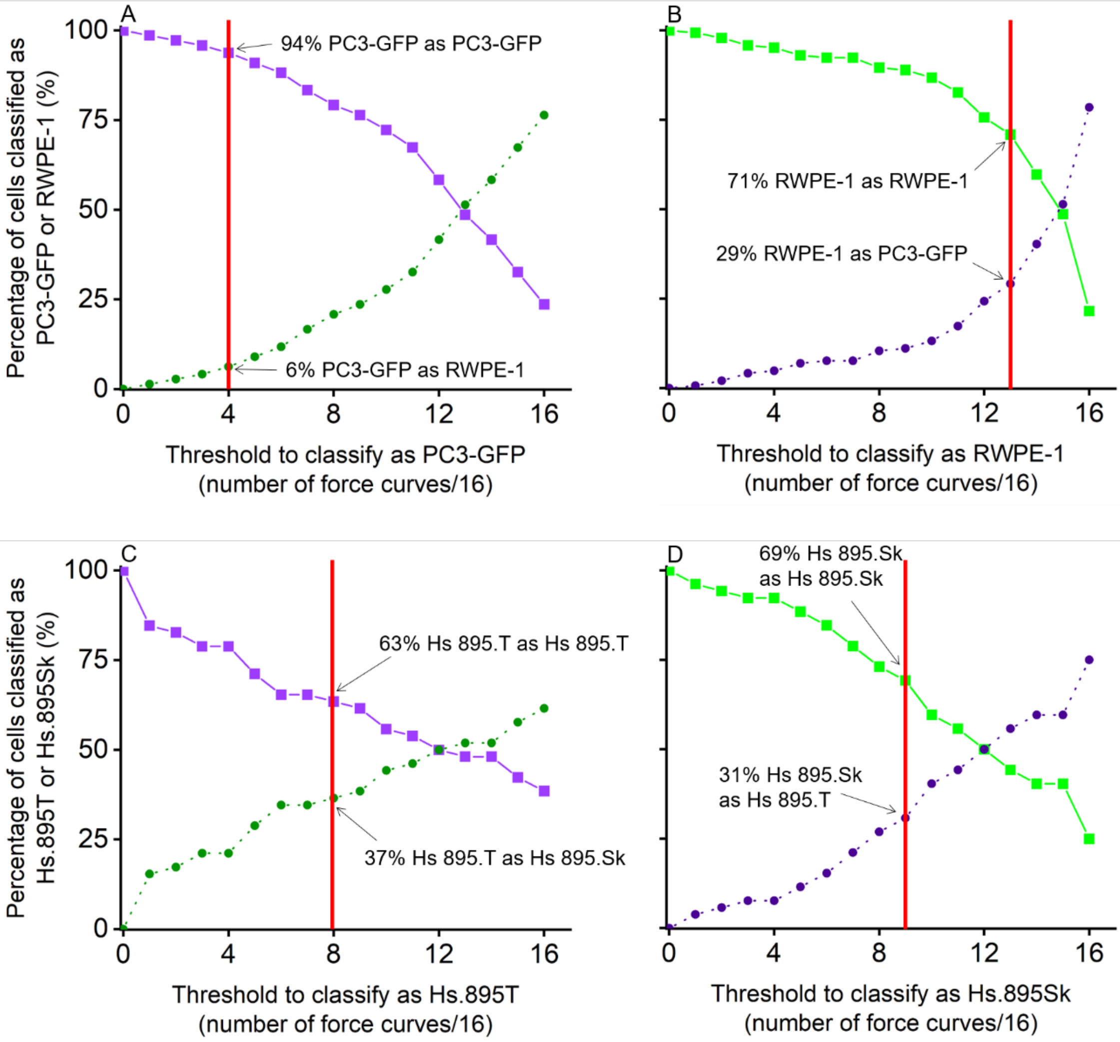
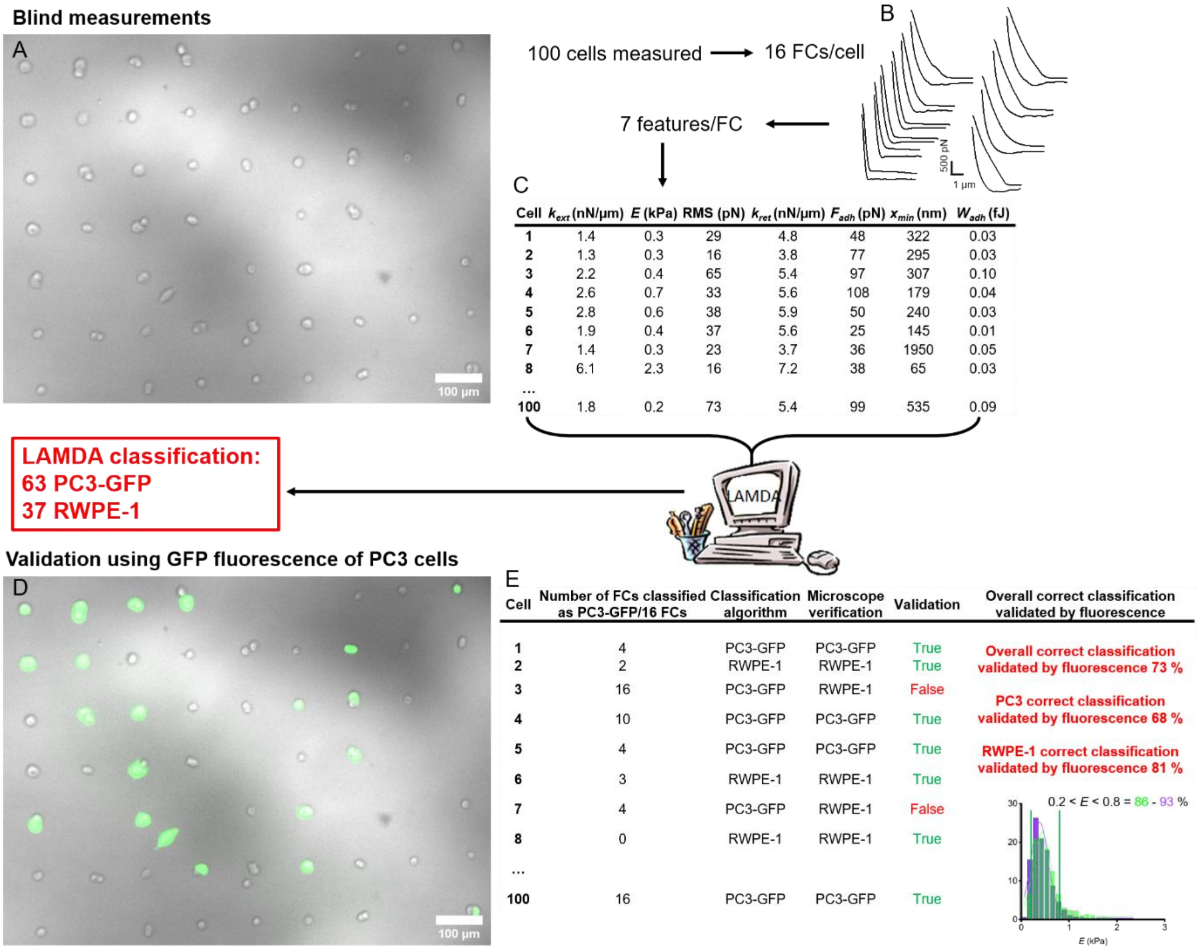


Figure 5|. Performance of LAMDA for the classification of PC3-GFP and RWPE-1 cells (A, B) and of Hs 895.T and Hs 895.Sk cells (C, D) as a function of the threshold number of FCs/cell assigned to one or the other category. The red bar represents a possible assignment of the threshold in order to avoid misclassification of cancerous cells (PC3-GFP and Hs 895.T cells). (**A, B**) Percentages of classification of cells as PC3-GFP or RWPE-1 based on the minimum threshold to classify as PC3-GFP (**A**) or RWPE-1 (**B**). (**C, D**) Percentages of classification of cells as Hs 895.T or Hs 895.Sk based on the minimum threshold to classify as Hs 895.T (**C**) or Hs 895.Sk (**D**). The light purple curve represents true positives and the light green curve represents true negatives. The dark green curve represents false negatives and the dark purple curve represents false positives.

To validate the concept, we then designed an experiment allowing us to confirm the classification performed by the algorithm. In this experiment, we performed a co-incubation of the two cell lines by seeding them on the same fibronectin array (Fig. 6A). AFM measurements were then performed on 100 cells (Fig. 6B), analysed for the seven features considered (Fig. 6C), and classified by our previously developed algorithm. The classification produced by the algorithm predict that 63 cells are PC3-GFP cells and that 37 cells are RWPE-1 cells. Because PC3-GFP express GFP (Fig. 6D), we were able to compare the classification of the algorithm with the fluorescence images. Considering all the 100 cells, the overall concordant classification rate between the algorithm and the fluorescent validation was 73 %; among which 68 % of PC3 cells and 81 % of RWPE-1 cells were correctly identified (Fig. 6E). Once again, this result must be put in perspective with the overlap between the measured biomechanical feature values of the two cell lines (Fig. 4). This experiment allowed us to quantitatively determine the performance of our algorithm and that it was able to discriminate between cancerous cells (PC3-GFP) and non-malignant cells (RWPE-1). The sensitivity which refers to the ability of a method to correctly identify true positives from false negatives (= true positives / (true positives + false negatives)) where true positives are PC3-GFP cells classified as PC3-GFP cells, and false negatives are PC3-GFP cells classified as RWPE-1 cells is 0.78; In contrast, the specificity which is the ability to detect true negatives from false positives (= true negatives / (true negatives + false positives)) where true negatives are RWPE-1 cells classified as RWPE-1 cells and false positives are RWPE-1 cells classified as PC3-GFP cells is 0.72. The accuracy, which corresponds to all the correct assignments in relation to the total (= (true-positives + true-negatives) / (true-positives + true-negatives + false-positives + false-negatives)) is 0.75. These results should be compared with the percentages of overlap between the values of the 7 features of the 2 cell lines, which ranged from 82 % to 100 %. At this stage, a sensitivity of 78 % does not provide a marketable diagnostic tool for cancer detection. The aim of this study is rather to demonstrate the ability of the LAMDA algorithm to discriminate between cell lines under co-incubation conditions.



Cell	k_{ext} (nN/µm)	E (kPa)	RMS (pN)	k_{ret} (nN/µm)	F_{adh} (pN)	x_{min} (nm)	W_{adh} (fJ)
1	1.4	0.3	29	4.8	48	322	0.03
2	1.3	0.3	16	3.8	77	295	0.03
3	2.2	0.4	65	5.4	97	307	0.10
4	2.6	0.7	33	5.6	108	179	0.04
5	2.8	0.6	38	5.9	50	240	0.03
6	1.9	0.4	37	5.6	25	145	0.01
7	1.4	0.3	23	3.7	36	1950	0.05
8	6.1	2.3	16	7.2	38	65	0.03
...							
100	1.8	0.2	73	5.4	99	535	0.09

Cell	Number of FCs classified as PC3-GFP/16 FCs	Classification algorithm	Microscope verification	Validation	Overall correct classification validated by fluorescence
1	4	PC3-GFP	PC3-GFP	True	**Overall correct classification validated by fluorescence 73 %**
2	2	RWPE-1	RWPE-1	True	
3	16	PC3-GFP	RWPE-1	False	**PC3 correct classification validated by fluorescence 68 %**
4	10	PC3-GFP	PC3-GFP	True	
5	4	PC3-GFP	PC3-GFP	True	**RWPE-1 correct classification validated by fluorescence 81 %**
6	3	RWPE-1	RWPE-1	True	
7	4	PC3-GFP	RWPE-1	False	
8	0	RWPE-1	RWPE-1	True	
...					
100	16	PC3-GFP	PC3-GFP	True	

Figure 6|. Performance validation of the LAMDA algorithm on RWPE-1 and PC3-GFP cell lines co-incubated on the same fibronectin array. (**A**) Blind measurements. RWPE-1 and PC3-GFP cell lines were co-incubated on the same fibronectin array. (**B**) Different FCs measured on the cell array in (**A**). (**C**) Table showing the mean values over the 16 FCs/cell of the seven features used (approach rigidity, Young's modulus, RMS, retract rigidity, adhesion, minimum position and area under the curve) for the different indented cells in (**A**). (**D**) Validation using GFP fluorescence of PC3 cells on the same cell array as in (**A**). Green cells correspond to PC3-GFP while cells in bright field correspond to RWPE-1. (**E**) Table showing the class assigned by the algorithm for each cell measured in (**A**) and its validation by fluorescence microscopy. The correct classification corresponds to the matches between the classification by algorithm and the verification by fluorescence.

Discussion

Atomic force microscopy is a popular technology for obtaining mechanobiology data on living cells. Among the other available methods used to address the measurement of mechanical features on cells, one can cite micropipette aspiration (MPA), laser particle tracking micro-rheology (PTM), optical stretching (OS), parallel plate rheometry (PPR), and real-time deformability/magnetic twisting cytometry (MTC) [27,28]. Unlike AFM, which probes the cells locally, these techniques (except MTC) probe the cells globally. Therefore, they cannot be used to obtain information about specific regions of the cell such as the cytoplasm and nucleus. In addition, OS and PTM probe non-adherent cells which is therefore suitable for studying blood cells or circulating tumour cells but is not relevant for exploring adherent epithelial cells. One advantage of AFM over other techniques is that it allows multiparametric measurements[5]. Smolyakov *et al.* demonstrated a few years ago that adhesion, elasticity and tethers extrusion were interesting features that can be interpreted from FCs. Very recently Wang *et al.*[15] called the combination of several mechanical features mechanome and demonstrated its interest in deciphering between non-malignant and malignant cells. In the present work, we developed this concept a step further, by extracting and combining seven features from approach and retract couple of force-distance curves (FCs). Another advantage of AFM compared to some other techniques lies in the possibility of combining it with other techniques, especially optical ones, which makes it a multimethodological approach[29]. Combining AFM with optical and/or fluorescence microscopy allowing, on the one hand, to visualize the cell morphology globally, and on the other hand to target, with the indenter, specific elements related to cell mechanics (vinculin, actin, myosin, *etc.*), labelled with fluorescence[30–32].

Although AFM is a technique of choice compared to other methods used to obtain mechanobiology data, it presents a limitation in its development, related to the great dispersity of the experimental procedures (cell immobilization, indenter velocity, indenter geometry), as well as the data processing methods (model used to fit the indentation curves, determination of the contact point, indentation depth fitted, Poisson ratio choice, *etc.*). This is concretized by the fact that, concerning Young's modulus, the literature is full of contradictory results, most probably because they were not obtained under comparable conditions[33]. For example, on non-cancerous cell lines, such as HCV29 from the bladder, the results can vary between 7.5 and 33.0 kPa[34–36], and this is also the case for the prostate cell line RWPE-1 with results ranging from 6.1 to 87.0 kPa[37–39] and in our case, a Young's modulus of 0.5 ± 0.2 kPa. But the same is true for cancerous cell lines, such as the PC3 cell line from the prostate, with results ranging from 1.4 to 2.5 kPa[40–42] and even less in our work: 0.3 ± 0.1 kPa or the MCF-7 cell line from the breast, with results ranging between 0.4 and 87.0 kPa[41,43,44]. It should be noted that in the present work we use a feature measuring the quality of the fit to the model used to calculate the cell Young's modulus making the fit quality one of the discriminating features used in the multiparametric clustering algorithm.

To further minimize these problems, we questioned the experimental parameters used in our study. This step is important to standardize and produce comparable results. We evaluated the impact of 4 parameters (cell patterning, indenter geometry, indenter velocity, number of measurements per cell) and propose a set of parameters that minimizes the data acquisition time while limiting the data dispersion due to measurement. This step (Fig. 1) allowed us to choose the parameters to indent the different cell lines for this study. It was thus concluded that the cells would be localized on 40 µm x 40 µm square fibronectin patterns on which they adhere and that 16 indentations per cell with a spherical indenter at an indentation velocity of 50 µm/s would be performed.

Another limitation of mechanobiological measurements by AFM comes from its low throughput. To our knowledge, there are no bio-AFM publications based on the measurement of many cells, the average number being around 30 cells. One way to overcome this limitation is to automate the measurements to increase the data throughput and obtain measurements on a statistically significant number of cells. Different automation approaches have been tried. For example, Favre *et al.*[45] multiplexed the number of cantilevers using two-dimensional probe arrays to run up to 35 AFM measurements in parallel. Sadeghian *et al.*[46] designed 4 miniaturized AFMs in parallel, making possible 4 measurements at the same time and at different locations over a large surface sample. Dujardin *et al.*[47] developed a script allowing the AFM to automatically switch from one sample to another; thus, they were able to analyse without user intervention 485 bacteria in about 8 h. These works are dedicated to imaging[45,47] or performed on non-living materials[46]. On the contrary, even if the precision of the sample displacement stage poses a limit (Fig. 2, Supplementary Information 2-3), we were able to measure a record number of living mammalian cells (956 cells) and to produce a large number of data (107 072 features in total) compatible with artificial intelligence analysis. In fact, we were able to test and use training bases containing up to 240 cells (120 non-cancerous and 120 cancerous cells: Supplementary Information 9A) and decided to work with a training dataset consisting of 100 cells. For the RWPE-1/PC3-GFP cell line pair, it appears that for a training base containing less than 50 cells, only one category of cells (*e.g.* cancerous) is well classified while the second category (non-cancerous) has a much lower rate of correct classification. This may be due to an underfitting linked to the low amount of data in the training database. Conversely, for the cell lines pair Hs 895.Sk/Hs 895.T, it appears that for a training database containing more than 100 cells, only one category of cells (*e.g.* cancerous) is well classified whereas the second category (non-cancerous) has a much lower rate of correct classification (Supplementary Information 9B). This may be due to an overfitting linked to a too large number of data in the training database. Finally, thanks to cell patterning and the automation of AFM, this work opens the door to data analysis by machine learning algorithms.

The majority of the few studies combining AFM and artificial intelligence are concerning AFM image analysis[48–50]. However, machine learning approaches to sort force-distance curves are emerging, for example Müller *et al.*[51] trained an algorithm, called *nanite*, to estimate the quality of such force curves, based on subjective ratings extracted from predefined features. The idea of classifying cell lines, using a machine learning algorithm trained on biomechanical features derived from AFM measurements was introduced by Wang *et al.*[15] In their work, the team combined two mechanobiological features extracted from each approach FC (prestress and elastic modulus) which they named mechanome. From this mechanome, they were able to create a cell malignancy classifier based on SVM (support vector machine[52]). This can be considered as a breakthrough in AFM mechanobiology, however, only two features extracted from a fit of the approach force curves were considered and hundreds of FCs per cell were recorded. During the optimization of the AFM measurements, we were able to show that making 256 or 16 measurements per cell had no impact on the value of the rigidity nor on the dispersion of the results (Fig. 1R-Y), which implied that the comparable values were measured several times in both cases. Therefore, by performing hundreds of FCs per cell, the authors only measured between 30 and 57 cells per cell line on 4 pairs of cancerous/non-cancerous cell lines (Hela/Cervical, MCF7/MCF10A, A549/Airway and PL16T/PL16B). As a consequence, their training data set contained 49 cells which, in our case has been demonstrated to be too small to get correct classification rates (Supplementary Information 9). In addition, this 49-cell training base contained only the Hela/Cervical cell lines pair and was used to classify all 4 cell lines pairs. By doing so, the authors surprisingly obtained correct classification rates between 71 and 100 %. However, it is questionable whether this classifier will still be effective for other cell lines.

In our study, we extended the mechanome concept by combining 7 features in a machine learning algorithm capable of distinguishing between a cancerous cell line (PC3-GFP) and its non-cancerous counterpart (RWPE-1). We tested it under co-incubation conditions and demonstrated its ability to discriminate between cell lines. In addition, we extended our study with a pair of other cell lines (pairs of cancerous and non-cancerous cells, Hs 895.T/Hs 895.Sk from the same organ). It is also possible to include this classification in clinical studies based on cancerous or non-cancerous biopsies for diagnostic purposes. In the study by Sokolov *et al.*[53], the authors collect samples from more than 60 patients with or without bladder cancer. This constitutes their database, which they then divide between a training database and a test database. They then analyse five urine cells per patient. Each cell is classified as cancerous or non-cancerous. The authors then decide to set a threshold of two out of five cells classified as cancerous to assign a cancerous diagnosis to the patient. This is the first application of machine learning classification in clinical practice. However, unlike our study, the authors based their analysis on AFM images from which they extracted parameters, but did not look directly at the mechanical properties of their samples. In addition, given that cancer cells tend to aggregate to form a spherical structure, it could also be interesting to classify spheroids. Recent studies[54,55] have focused on the mechanical properties of spheroids as measured by AFM. The mechanical properties of individual cells within the spheroid change according to the position of the cell within the structure, with stiffer cells at the periphery and softer cells in the centre. So far, these studies have measured up to a hundred spheroids. We certainly need to start thinking about how to incorporate these new findings into machine learning classification.

In future works, still with the aim of better discriminating different cell types, it could be possible to complete the mechanome with additional features such as visco-elasticity[56,57] which seems to be a promising feature, specific adhesion (indenter functionalized with an antibody[58,59]), membrane tethers (force, number, distance[5,60,61]), optical images of the cells, *etc.*

Conclusion

To conclude, our work tackles head-on several issues in the field of AFM mechanobiology. We propose a set of optimized and justified AFM data acquisition parameters. We also present a method for automatic AFM data acquisition. Once the script is set, the cells organized according to a defined pattern are automatically moved under the AFM indenter to be measured. Thanks to this process, we were able to perform measurements on hundreds of cells. The large amount of data generated was then used to train a machine learning algorithm to discriminate non-malignant and cancerous cells based on their nanomechanical properties. This is a major advance when compared to conventional analysis (Fig. 4), which shows that 82-100 % of the features extracted from the FCs of the two cell lines fall in the same range. This means that if we seek to classify a measurement, or a small number of measurements performed on a cell, we cannot conventionally conclude to belong to one category or the other, whereas this is made possible with machine learning. This quite versatile method is undoubtedly generalizable to other cell types and to other problems of cell classification based on their mechanome under the conditions that learning data sets are conveniently defined with respect to the classification objective and clinical context of the related application.

Methods

Cell culture. The RWPE-1 cell line (ATCC CRL-11609) was cultured in K-SFM medium with 0.05 mg/mL Bovine Pituitary Extract (BPE), 5 ng/mL Epidermal Growth Factor (EGF) and 1 % penicillin-streptomycin (Gibco™, Thermo Fisher Scientific Inc.). The PC3-GFP cell line (kindly provided by the Cuvillier laboratory at IPBS, Toulouse, France) is a cell line transformed to constitutively express the green fluorescent protein (GFP) from the PC3 cell line. The PC3-GFP cell line was cultured in RPMI medium, containing L-glutamine, HEPES buffer and phenol red (Gibco™, Thermo Fisher Scientific Inc.) with 10 % foetal bovine serum (FBS, Gibco™, Thermo Fisher Scientific Inc.), 1 % penicillin-streptomycin (Gibco™, Thermo Fisher Scientific Inc.) and 1 % geneticin (G418, Gibco™, Thermo Fisher Scientific Inc.). The Hs 895.Sk cell line (ATCC CRL-7636) was cultured in DMEM medium containing sodium bicarbonate, glucose and phenol red (Gibco™, Thermo Fisher Scientific Inc.) with 4 mM L-glutamine (Gibco™, Thermo Fisher Scientific Inc.), 10 % foetal bovine serum (FBS, Gibco™, Thermo Fisher Scientific Inc.). and 1 % penicillin-streptomycin (Gibco™, Thermo Fisher Scientific Inc.). The Hs 895.T cell line (ATCC CRL-7637) was cultured in DMEM medium containing sodium bicarbonate, glucose and phenol red (Gibco™, Thermo Fisher Scientific Inc.) with 10 % foetal bovine serum (FBS, Gibco™, Thermo Fisher Scientific Inc.). and 1 % penicillin-streptomycin (Gibco™, Thermo Fisher Scientific Inc.). The 4 cell lines were grown in an incubator at 37 °C and 5 % CO_2. The cells were then seeded on 20 mm side glass coverslips at a concentration of 30 000 or 50 000 cells per cm² and placed in 40 mm diameter Petri dishes overnight.

Production of cell arrays. The cell arrays production follows a two steps process. i) microcontact-printing (µCP) of fibronectin patterns using the InnoStamp 40™ device[22]. ii) cell seeding.

Microcontact-printing. Flat or micro-structured (40 x 40 µm^2 squares) PDMS stamps, manufactured according to Fredonnet *et al.*[62], were deposited at the loading point of the InnoStamp 40™. 101 µL (for flat stamps) or 71 µL (for stamps with square patterns) of bovine plasma fibronectin (Sigma Aldrich, Merck KGaA.) at a concentration of 100 µg/mL (in 1X PBS, pH 7.4: NaCl: 137 mM; KCl: 2.7 mM; Na_2HPO_4: 10 mM; KH_2PO_4: 1.8 mM) were deposited on a glass slide at the inking location of the Innostamp 40™. The glass slide serving as the substrate was cleaned for 10 s with acetone, 10 s with deionized water and 10 s with isopropanol before undergoing an oxygen plasma treatment (Diener Pico, 50 W, 0.3 mbar, 1 min 30) (plasma activation). It was then placed on the InnoStamp 40™ drop zone. The automated µCP protocol was then initiated and consisted of inking the surface of the PDMS stamps with fibronectin, followed by drying and finally µCP of the fibronectin onto a glass coverslip. The functionalized glass slide was removed from the InnoStamp 40™ and placed in a 40 mm diameter Petri dish. 200 µL of PLL-g-PEG (PLL (20)-g [3.5]-PEG (2), SuSoS AG.) at a concentration of 100 µg/mL (in 1X PBS, pH 7.4) were poured on the glass slides and after a 30 min incubation, the coverslips were rinsed 3 times with 1X PBS, pH 7.4.

Cell seeding. RWPE-1, PC3-GFP, Hs 895.Sk and Hs 895.T cells were concentrated to 30,000 or 50,000 cells/cm² in K-SFM medium with 0.05 mg/mL Bovine Pituitary Extract (BPE), 5 ng/mL Epidermal Growth Factor (EGF) and 1 % penicillin-streptomycin (Gibco™, Thermo Fisher Scientific Inc.), in RPMI medium, containing L-glutamine, HEPES buffer and phenol red (Gibco™, Thermo Fisher Scientific Inc.) with 10 % foetal bovine serum (FBS, Gibco™, Thermo Fisher Scientific Inc.), 1 % penicillin-streptomycin (Gibco™, Thermo Fisher Scientific Inc.) and 1 % geneticin (G418, Gibco™, Thermo Fisher Scientific Inc.), in DMEM medium containing sodium bicarbonate, glucose and phenol red (Gibco™, Thermo Fisher Scientific Inc.) with 4 mM L-glutamine (Gibco™, Thermo Fisher Scientific Inc.), 10 % foetal bovine serum (FBS, Gibco™, Thermo Fisher Scientific Inc.) and 1 % penicillin-streptomycin (Gibco™, Thermo Fisher Scientific Inc.) and in DMEM medium containing sodium bicarbonate, glucose and phenol red (Gibco™, Thermo Fisher Scientific Inc.) with 10 % foetal bovine serum (FBS, Gibco™, Thermo Fisher Scientific Inc.). and 1 % penicillin-streptomycin (Gibco™, Thermo Fisher Scientific Inc.) respectively, then seeded onto glass coverslips placed in 40 mm diameter Petri dishes supplemented with 2 mL of culture medium in an incubator at 37 °C and 5 % CO_2. 1.5 h after seeding, the culture medium was changed to eliminate non-adherent cells. Then, cells were incubated overnight before AFM measurements.

Making spherical probes. To prepare the spherical probe cantilevers, NP-O10 tipless AFM probes were used. The cantilevers were cleaned with oxygen plasma (Diener Pico, 50 W, 3 min, 0.5 mbar, 100 %). Thanks to the AFM Nanowizard® III device (JPK Instruments, Bruker Nano GmbH.), the AFM probes were approached to glue (Norland optical adhesive 63) deposited on a glass slide with a setpoint of 2 V, then on silica beads of 5.04 µm in diameter (Microsil Microspheres) for 30s with a setpoint of 10 V. Finally, the cantilevers were placed under a UV lamp for 10 min.

Calibration of the cantilevers: sensitivity and spring constant. All AFM experiments were performed on a NanoWizard® III AFM device (JPK Instruments, Bruker Nano GmbH.) in contact mode, force mapping. Before any experiment, the sensitivity of the AFM photodiode was calibrated. In this study, MLCT-BIO-DC cantilevers of triangular geometry (radius 20 nm; semi-open angle 35 °) with a sensitivity of 38.22 nm/V to 52.52 nm/V were used as well as spherical-shaped cantilevers with a sensitivity of 33.49 nm/V to 46.64 nm/V. The spring constant was then calibrated by studying the thermal fluctuation of the cantilever at 37 °C. The peak of the thermal spectrum was fitted to extract the resonant frequency. In this study, the MLCT-BIO-DC cantilevers had spring constants ranging from 0.007 N/m to 0.011 N/m and the spherical cantilevers between 0.016 N/m and 0.031 N/m. Force mapping measurements were performed in culture medium buffered with 5 % CO_2 and maintained at 37 °C using a PetriDishHeater (Bruker). Mapping was done with a relative setpoint of 3 nN, Z length of 5 µm, extension velocity of 5, 10, 25, or 50 µm/s, and in 32 x 32; 16 x 16; or 4 x 4 pixels over 10 x 10 µm² areas.

Force spectroscopy: measurement of the seven features. seven features were extracted from the different FCs. These are the approach rigidity, the retract rigidity, the value of the adhesion force relative to the baseline, the minimum position of the adhesion force relative to the contact point, the adhesion work, the Young's modulus and the RMS.

- The approach and retract rigidities are calculated from the indentation slopes of the approach and retract curves, respectively. These slopes fit 300 nm of indentation before and after the setpoint, respectively.

- The adhesion force corresponds to the minimum value of the force of the retract curve and to this position of indentation, the minimum position is assigned.

- The adhesion work is the area between the retract curve and the 1, describing the work required to detach the AFM indenter from the sample surface.

- The Young's modulus (or elastic modulus) was calculated with the Hertz model for a sphere and according to the formula:

$$F = \frac{4}{3}\frac{(E)}{1-v^2}\sqrt{R}\,\delta^{3/2} \quad (1)$$

where *F* is the force applied to the sample, *E* the Young's modulus, *v* is the Poisson's ratio, which is considered equal to 0.5 for soft biological samples, *δ* the indentation and *R* the radius of the sphere, *i.e.* 2.52 µm for Microspheres beads.

- The RMS is the square root in piconewton of the sum of square errors between the model and the indentation curve and describes the quality of the fit.

All data analyses were performed with JPK Data Processing software (version 6.1.186).

For spider chart representation, the median values (median over the 16 FCs) of all cells in the training base (100 PC3-GFP + 100 RWPE-1) were normalized between 0 and 1 for each feature. This gave 200 values for each feature, with a single minimum corresponding to a normalized value of 0 and a single maximum corresponding to a normalized value of 1. We used the following formula:

$$normalized\ feature = \frac{feature\ value\ -\ minimum\ of\ this\ feature}{maximumof\ this\ feature -\ minimum\ of\ this\ feature}$$

Finally, the median of 100 normalized values per feature and per cell line was calculated, and it is these normalized values that are represented on the spider charts (figure 4B and C).

Automation of mechanobiological measurements. The measurements by automation could be carried out on 376 positions in around 1.5 h. AFM automation was performed using a motorized precision stage MotStage Zeiss AxioObserver (S/N SM-01-0027) mounted on a Zeiss Axiovert 200M inverted optical microscope. To run the automation script, the Experiment Planner Module, included in the JPK SPM software control (SPM version 6.1.146) (see Supplementary Information 2), was used. This module works in Jython programming language and is compatible with Ubuntu 10.04 LTS (Lucid Lynx). Before running the script, we must first create a folder where the data will be saved. When running the script, the first step is to indicate the name of this folder as Parent_dir (line 61). Then, the second step is to indicate the variables of the ForceMaps, namely the size of the scan in ScanSize and the number of measurements in ScanPixels (lines 132-133). Then, we must also indicate the coordinates of the centre of the first cell (cell which is at the top left of the cells array) in P1 (line 111) and those of the centre of the last cell of the line (cell which is at the top right of the cells array) in P2 (line 112) as well as the number of cells on this line (cells P1 and P2 included) in nFP (line 118). With these three data, the script is able to calculate all other coordinates of the areas to be mapped within the cells array. Next, the AFM stage is moved in order to position the cantilever in the centre of the first cell. First, it takes an image of the cell. On that cell, several indentations are performed by moving the cantilever, ensuring measurements on different regions of the cell. Once the cell has been measured, the piezo is retracted and the AFM stage moves the cells array to the next cell, and so on.

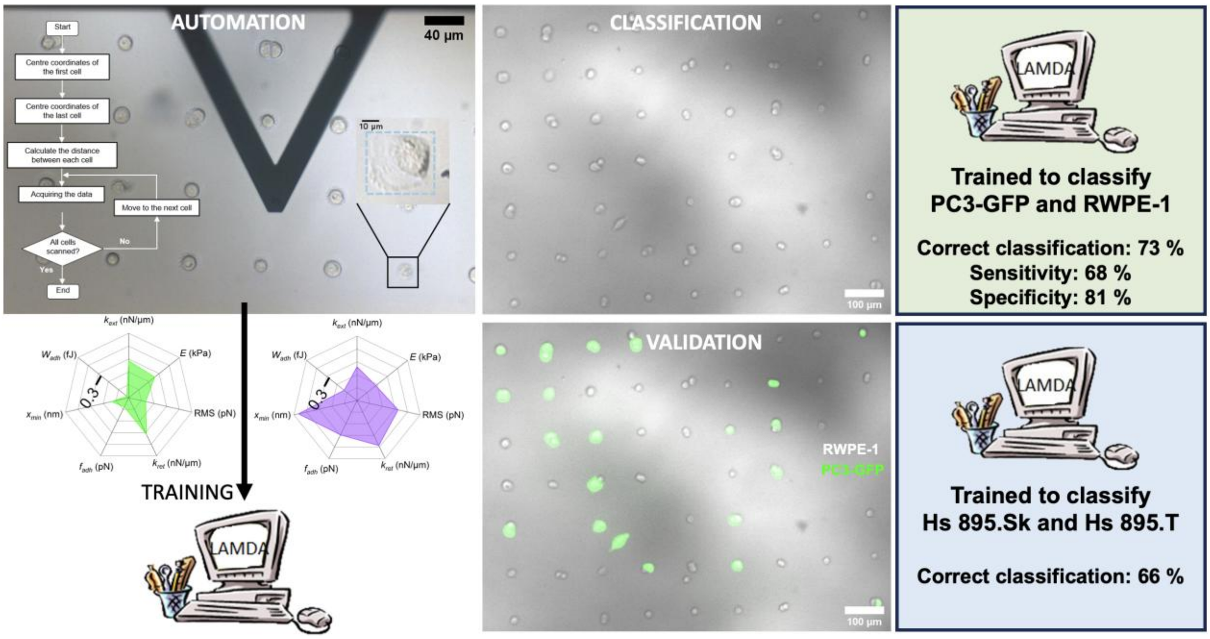
Fuzzy feature partition. The fuzzy function is a logic used to describe fuzziness. If the event to be classified changes continuously, it cannot always be defined in one class rather than another. The event is then defined in a fuzzy way. In general, the fuzzy function deals with vague information. In the fuzzy function, the values assigned are between 0 and 1 and the object to be classified can simultaneously belong to different classes with a certain degree of membership (between 0 and 1).

Class prediction – Fuzzy classification algorithm. The learning algorithm for multivariate data analysis (LAMDA) is a fuzzy logic-based method for grouping features of an object together into classes. This method is based on the computation of the adequacy degree, by means of a membership function (the fuzzy binomial function) considering all the contributions of each input feature of an object and the previously established classes. Several marginal adequacy degrees (MAD) are associated to an object. Thus, *n* MADs are created for each object with each class.

Subsequently, these *n* MADs of a single object are aggregated into the global adequacy degree (GAD) of an object into a class using connectors.

Cell signature classification. The database with all cells was divided into training "Trainbase" (100 RWPE-1 cells and 100 PC3-GFP cells combined into one database) and test databases (144 RWPE-1 cells and 144 PC3-GFP cells, each cell line has its own test database). The seven extracted features were normalized from 0 to 1. In the end, the "Trainbase" contains the normalized values of the seven features of the 16 FCs of the 100 RWPE-1 cells and the 100 PC3-GFP cells. Each FC is assigned to a class (1 or 2 for RWPE-1 or PC3-GFP). And the test databases contain the normalized values of the seven features of the 16 FCs of the 144 other RWPE-1 cells and 144 other PC3-GFP cells. None of these FCs are assigned to a class. The population of the "Trainbase" is the context for the algorithm. LAMDA serves as a classifier for the algorithm and makes each new feature value in the test base fuzzy. These fuzzy sets represent the degree of membership of the features for one or another of the classes.[63]

Statistical analysis. Histograms have been fitted to a Gaussian distribution ($f(x) = \frac{1}{\theta\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\theta})^2}$). In a Gaussian law, the centre (*x*) of the distribution is the mean and the median. The width at ($e^{-\frac{1}{2}}$) of the distribution corresponds to twice the standard deviation. Values are given as means ± standard deviations (SD). Distributions consider the values measured or calculated on each force curve for all cells of the same cell line. Statistical t-tests were performed on Originlab up to a p-value of $p \leq 0.001$. The values given in the table are the calculated medians of the 16 FCs per cell.



Graphical abstract: Automated bio-AFM generation of large mechanome data set and their analysis by machine learning to classify cancerous cell lines.

References

(1) Pillet, F.; Chopinet, L.; Formosa, C.; Dague, É. Atomic Force Microscopy and Pharmacology: From Microbiology to Cancerology. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2014**, *1840* (3), 1028–1050. https://doi.org/10.1016/j.bbagen.2013.11.019.

(2) Parot, P.; Dufrêne, Y. F.; Hinterdorfer, P.; Le Grimellec, C.; Navajas, D.; Pellequer, J.-L.; Scheuring, S. Past, Present and Future of Atomic Force Microscopy in Life Sciences and Medicine. *Journal of Molecular Recognition* **2007**, *20* (6), 418–431. https://doi.org/10.1002/jmr.857.

(3) Deng, X.; Xiong, F.; Li, X.; Xiang, B.; Li, Z.; Wu, X.; Guo, C.; Li, X.; Li, Y.; Li, G.; Xiong, W.; Zeng, Z. Application of Atomic Force Microscopy in Cancer Research. *Journal of Nanobiotechnology* **2018**, *16* (1), 102. https://doi.org/10.1186/s12951-018-0428-0.

(4) Muller, D. J. AFM: A Nanotool in Membrane Biology. *Biochemistry* **2008**, *47* (31), 7986–7998. https://doi.org/10.1021/bi800753x.

(5) Smolyakov, G. et al. Elasticity, Adhesion, and Tether Extrusion on Breast Cancer Cells Provide a Signature of Their Invasive Potential. *ACS Applied Materials & Interfaces*. 2016.

(6) Dague, E.; Alsteens, D.; Latgé, J.-P.; Verbelen, C.; Raze, D.; Baulard, A. R.; Dufrêne, Y. F. Chemical Force Microscopy of Single Live Cells. *Nano Lett.* **2007**, *7* (10), 3026–3030. https://doi.org/10.1021/nl071476k.

(7) Dufrêne, Y. F.; Hinterdorfer, P. Recent Progress in AFM Molecular Recognition Studies. *Pflugers Arch - Eur J Physiol* **2008**, *456* (1), 237–245. https://doi.org/10.1007/s00424-007-0413-1.

(8) Rankl, C.; Kienberger, F.; Wildling, L.; Wruss, J.; Gruber, H. J.; Blaas, D.; Hinterdorfer, P. Multiple Receptors Involved in Human Rhinovirus Attachment to Live Cells. *Proceedings of the National Academy of Sciences* **2008**, *105* (46), 17778–17783. https://doi.org/10.1073/pnas.0806451105.

(9) Cross, S. E.; Jin, Y.-S.; Rao, J.; Gimzewski, J. K. Nanomechanical Analysis of Cells from Cancer Patients. *Nature Nanotech* **2007**, *2* (12), 780–783. https://doi.org/10.1038/nnano.2007.388.

(10) Stolz, M.; Gottardi, R.; Raiteri, R.; Miot, S.; Martin, I.; Imer, R.; Staufer, U.; Raducanu, A.; Duggelin, M.; Baschong, W.; Daniels, A. U.; Friederich, N.; Aszodi, A.; Aebi, U. Early Detection of Osteoarthritis and Articular Cartilage Aging in Mice and Patient Biopsies Using Atomic Force Microscopy. *Nature Nanotechnology* **2009**, *4*.

(11) Smolyakov, G.; Cauquil, M.; Severac, C.; Lachaize, V.; Guilbeau-Frugier, C.; Sénard, J.-M.; Galés, C.; Dague, E. Biophysical Properties of Cardiomyocyte Surface Explored by Multiparametric AFM. *Journal of Structural Biology* **2017**, *198* (1), 28–37. https://doi.org/10.1016/j.jsb.2017.03.001.

(12) Dague, E.; Genet, G.; Lachaize, V.; Guilbeau-Frugier, C.; Fauconnier, J.; Mias, C.; Payré, B.; Chopinet, L.; Alsteens, D.; Kasas, S.; Severac, C.; Thireau, J.; Heymes, C.; Honton, B.; Lacampagne, A.; Pathak, A.; Sénard, J.-M.; Galés, C. Atomic Force and Electron Microscopic-Based Study of Sarcolemmal Surface of Living Cardiomyocytes Unveils Unexpected Mitochondrial Shift in Heart Failure. *Journal of Molecular and Cellular Cardiology* **2014**, *74*, 162–172. https://doi.org/10.1016/j.yjmcc.2014.05.006.

(13) Plodinec, M.; Loparic, M.; Monnier, C. A.; Obermann, E. C.; Zanetti-Dallenbach, R.; Oertle, P.; Hyotyla, J. T.; Aebi, U.; Bentires-Alj, M.; Lim, R. Y. H.; Schoenenberger, C.-A. The Nanomechanical Signature of Breast Cancer. *Nature Nanotech* **2012**, *7* (11), 757–765. https://doi.org/10.1038/nnano.2012.167.

(14) Zanetti-Dällenbach, R.; Plodinec, M.; Oertle, P.; Redling, K.; Obermann, E. C.; Lim, R. Y. H.; Schoenenberger, C.-A. Length Scale Matters: Real-Time Elastography versus Nanomechanical Profiling by Atomic Force Microscopy for the Diagnosis of Breast Lesions. *BioMed Research International* **2018**, *2018*, e3840597. https://doi.org/10.1155/2018/3840597.

(15) Wang, H.; Zhang, H.; Da, B.; Lu, D.; Tamura, R.; Goto, K.; Watanabe, I.; Fujita, D.; Hanagata, N.; Kano, J.; Nakagawa, T.; Noguchi, M. Mechanomics Biomarker for Cancer Cells Unidentifiable through Morphology and Elastic Modulus. *Nano Lett.* **2021**, *21* (3), 1538–1545. https://doi.org/10.1021/acs.nanolett.1c00003.

(16) Proa-Coronado, S.; Séverac, C.; Martinez-Rivas, A.; Dague, E. Beyond the Paradigm of Nanomechanical Measurements on Cells Using AFM: An Automated Methodology to Rapidly Analyse Thousands of Cells. *Nanoscale Horiz.* **2019**, *5* (1), 131–138. https://doi.org/10.1039/C9NH00438F.

(17) Severac, C.; Proa-Coronado, S.; Formosa-Dague, C.; Martinez-Rivas, A.; Dague, E. Automation of Bio-Atomic Force Microscope Measurements on Hundreds of C. Albicans Cells. *JoVE* **2021**, No. 170, 61315. https://doi.org/10.3791/61315.

(18) Hedjazi, L.; Aguilar-Martin, J.; Le Lann, M.-V.; Kempowsky-Hamon, T. Membership-Margin Based Feature Selection for Mixed Type and High-Dimensional Data: Theory and Applications. *Information Sciences* **2015**, *322*, 174–196. https://doi.org/10.1016/j.ins.2015.06.007.

(19) Hedjazi, L.; Aguilar-Martin, J.; Le Lann, M.-V. Similarity-Margin Based Feature Selection for Symbolic Interval Data. *Pattern Recognition Letters* **2011**, *32* (4), 578–585. https://doi.org/10.1016/j.patrec.2010.11.018.

(20) Kempowsky-Hamon, T.; Valle, C.; Lacroix-Triki, M.; Hedjazi, L.; Trouilh, L.; Lamarre, S.; Labourdette, D.; Roger, L.; Mhamdi, L.; Dalenc, F.; Filleron, T.; Favre, G.; François, J.-M.; Le Lann, M.-V.; Anton-Leberre, V. Fuzzy Logic Selection as a New Reliable Tool to Identify Molecular Grade Signatures in Breast Cancer – the INNODIAG Study. *BMC Med Genomics* **2015**, *8*, 3. https://doi.org/10.1186/s12920-015-0077-1.

(21) Martínez Rivas, A.; Dague, E.; Proa, C. S.; Séverac, C.; González, Q. G. K. Process Using Atomic Force Microscopy for Mass Physical and Mechanical Analysis in Materials, Biomaterial Arrangements and Structures. WO2019112414 (A1), June 13, 2019. https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20190613&DB=EPODOC&locale=en_EP&CC=WO&NR=2019112414A1&KC=A1&ND=5 (accessed 2022-12-14).

(22) Lagraulet, A.; Foncy, J.; Berteloite, B.; Esteve, A.; Blatche, M.-C.; Malaquin, L.; Vieu, C. InnoStamp 40^{TM} and InnoScan 1100AL^{TM}: A Complete Automated Platform for Microstructured Cell Arrays. *Nat Methods* **2015**, *12* (9), i–iii. https://doi.org/10.1038/nmeth.f.383.

(23) Bluemke, D. A. Radiology in 2018: Are You Working with AI or Being Replaced by AI? *Radiology* **2018**, *287* (2), 365–366. https://doi.org/10.1148/radiol.2018184007.

(24) Sharma, G.; Carter, A. Artificial Intelligence and the Pathologist: Future Frenemies? *Archives of Pathology & Laboratory Medicine* **2017**, *141* (5), 622–623. https://doi.org/10.5858/arpa.2016-0593-ED.

(25) O'Neill, C. Is AI a Threat or Benefit to Health Workers? *CMAJ* **2017**, *189* (20), E732–E732. https://doi.org/10.1503/cmaj.1095428.

(26) Recht, M.; Bryan, R. N. Artificial Intelligence: Threat or Boon to Radiologists? *Journal of the American College of Radiology* **2017**, *14* (11), 1476–1480. https://doi.org/10.1016/j.jacr.2017.07.007.

(27) Marx, V. May Mechanobiology Work Forcefully for You. *Nat Methods* **2019**, *16* (11), 1083–1086. https://doi.org/10.1038/s41592-019-0621-6.

(28) Wu, P.-H.; Aroush, D. R.-B.; Asnacios, A.; Chen, W.-C.; Dokukin, M. E.; Doss, B. L.; Durand-Smet, P.; Ekpenyong, A.; Guck, J.; Guz, N. V.; Janmey, P. A.; Lee, J. S. H.; Moore, N. M.; Ott, A.; Poh, Y.-C.; Ros, R.; Sander, M.; Sokolov, I.; Staunton, J. R.; Wang, N.; Whyte, G.; Wirtz, D. A Comparison of Methods to Assess Cell Mechanical Properties. *Nature Methods* **2018**, *15* (7), 491–498. https://doi.org/10.1038/s41592-018-0015-1.

(29) Krieg, M.; Fläschner, G.; Alsteens, D.; Gaub, B. M.; Roos, W. H.; Wuite, G. J. L.; Gaub, H. E.; Gerber, C.; Dufrêne, Y. F.; Müller, D. J. Atomic Force Microscopy-Based Mechanobiology. *Nat Rev Phys* **2019**, *1* (1), 41–57. https://doi.org/10.1038/s42254-018-0001-7.

(30) Maki, K.; Han, S.-W.; Hirano, Y.; Yonemura, S.; Hakoshima, T.; Adachi, T. Real-Time TIRF Observation of Vinculin Recruitment to Stretched α-Catenin by AFM. *Sci Rep* **2018**, *8*, 1575. https://doi.org/10.1038/s41598-018-20115-8.

(31) Hirvonen, L. M.; Marsh, R. J.; Jones, G. E.; Cox, S. Combined AFM and Super-Resolution Localisation Microscopy: Investigating the Structure and Dynamics of Podosomes. *Eur J Cell Biol* **2020**, *99* (7), 151106. https://doi.org/10.1016/j.ejcb.2020.151106.

(32) Nelsen, E.; Hobson, C. M.; Kern, M. E.; Hsiao, J. P.; O'Brien III, E. T.; Watanabe, T.; Condon, B. M.; Boyce, M.; Grinstein, S.; Hahn, K. M.; Falvo, M. R.; Superfine, R. Combined Atomic Force Microscope and Volumetric Light Sheet System for Correlative Force and Fluorescence Mechanobiology Studies. *Sci Rep* **2020**, *10*, 8133. https://doi.org/10.1038/s41598-020-65205-8.

(33) Schillers, H.; Rianna, C.; Schäpe, J.; Luque, T.; Doschke, H.; Wälte, M.; Uriarte, J. J.; Campillo, N.; Michanetzis, G. P. A.; Bobrowska, J.; Dumitru, A.; Herruzo, E. T.; Bovio, S.; Parot, P.; Galluzzi, M.; Podestà, A.; Puricelli, L.; Scheuring, S.; Missirlis, Y.; Garcia, R.; Odorico, M.; Teulon, J.-M.; Lafont, F.; Lekka, M.; Rico, F.; Rigato, A.; Pellequer, J.-L.; Oberleithner, H.; Navajas, D.; Radmacher, M. Standardized Nanomechanical Atomic Force Microscopy Procedure (SNAP) for Measuring Soft and Biological Samples. *Sci Rep* **2017**, *7* (1), 5117. https://doi.org/10.1038/s41598-017-05383-0.

(34) Lekka, M.; Laidler, P.; Gil, D.; Lekki, J.; Stachura, Z.; Hrynkiewicz, A. Z. Elasticity of Normal and Cancerous Human Bladder Cells Studied by Scanning Force Microscopy. *European Biophysics Journal* **1999**, *28* (4), 312–316. https://doi.org/10.1007/s002490050213.

(35) Zemła, J.; Bobrowska, J.; Kubiak, A.; Zieliński, T.; Pabijan, J.; Pogoda, K.; Bobrowski, P.; Lekka, M. Indenting Soft Samples (Hydrogels and Cells) with Cantilevers Possessing Various Shapes of Probing Tip. *Eur Biophys J* **2020**, *49* (6), 485–495. https://doi.org/10.1007/s00249-020-01456-7.

(36) Ramos, J. R.; Pabijan, J.; Garcia, R.; Lekka, M. The Softening of Human Bladder Cancer Cells Happens at an Early Stage of the Malignancy Process. *Beilstein J. Nanotechnol.* **2014**, *5* (1), 447–457. https://doi.org/10.3762/bjnano.5.52.

(37) Parihar, K.; Nukpezah, J.; Iwamoto, D. V.; Janmey, P. A.; Radhakrishnan, R. Data Driven and Biophysical Insights into the Regulation of Trafficking Vesicles by Extracellular Matrix Stiffness. *iScience* **2022**, *25* (8), 104721. https://doi.org/10.1016/j.isci.2022.104721.

(38) Pogoda, K.; Pięta, E.; Roman, M.; Piergies, N.; Liberda, D.; Wróbel, T. P.; Janmey, P. A.; Paluszkiewicz, C.; Kwiatek, W. M. In Search of the Correlation between Nanomechanical and Biomolecular Properties of Prostate Cancer Cells with Different Metastatic Potential. *Archives of Biochemistry and Biophysics* **2021**, *697*, 108718. https://doi.org/10.1016/j.abb.2020.108718.

(39) Bandara, C.; Dorrance, K.; Gay, T.; Srividya, N.; Muralidharan, S.; Park, B.-J.; Abu-Lail, N. Characterization of Nanomechanical Properties of Normal and Cancer Prostate Cells. *Washington State University Academic Showcase*. 2010.

(40) Faria, E. C.; Ma, N.; Gazi, E.; Gardner, P.; Brown, M.; Clarke, N. W.; Snook, R. D. Measurement of Elastic Properties of Prostate Cancer Cells Using AFM. *Analyst* **2008**, *133* (11), 1498–1500. https://doi.org/10.1039/B803355B.

(41) Lekka, M.; Gil, D.; Pogoda, K.; Dulińska-Litewka, J.; Jach, R.; Gostek, J.; Klymenko, O.; Prauzner-Bechcicki, S.; Stachura, Z.; Wiltowska-Zuber, J.; Okoń, K.; Laidler, P. Cancer Cell Detection in Tissue Sections Using AFM. *Archives of Biochemistry and Biophysics* **2012**, *518* (2), 151–156. https://doi.org/10.1016/j.abb.2011.12.013.

(42) Deliorman, M.; Janahi, F. K.; Sukumar, P.; Glia, A.; Alnemari, R.; Fadl, S.; Chen, W.; Qasaimeh, M. A. AFM-Compatible Microfluidic Platform for Affinity-Based Capture and Nanomechanical Characterization of Circulating Tumor Cells. *Microsyst Nanoeng* **2020**, *6*, 20. https://doi.org/10.1038/s41378-020-0131-9.

(43) Fischer, T.; Hayn, A.; Mierke, C. T. Effect of Nuclear Stiffness on Cell Mechanics and Migration of Human Breast Cancer Cells. *Frontiers in Cell and Developmental Biology* **2020**, *8*.

(44) Coceano, G.; Yousafzai, M. S.; Ma, W.; Ndoye, F.; Venturelli, L.; Hussain, I.; Bonin, S.; Niemela, J.; Scoles, G.; Cojoc, D.; Ferrari, E. Investigation into Local Cell Mechanics by Atomic Force Microscopy Mapping and Optical Tweezer Vertical Indentation. *Nanotechnology* **2015**, *27* (6), 065102. https://doi.org/10.1088/0957-4484/27/6/065102.

(45) Favre, M.; Polesel-Maris, J.; Overstolz, T.; Niedermann, P.; Dasen, S.; Gruener, G.; Ischer, R.; Vettiger, P.; Liley, M.; Heinzelmann, H.; Meister, A. Parallel AFM Imaging and Force Spectroscopy Using Two-Dimensional Probe Arrays for Applications in Cell Biology. *J. Mol. Recognit.* **2011**, *24* (3), 446–452. https://doi.org/10.1002/jmr.1119.

(46) Sadeghian, H.; Herfst, R.; Dekker, B.; Winters, J.; Bijnagte, T.; Rijnbeek, R. High-Throughput Atomic Force Microscopes Operating in Parallel. *Review of Scientific Instruments* **2017**, *88* (3), 033703. https://doi.org/10.1063/1.4978285.

(47) Dujardin, A.; Wolf, P. D.; Lafont, F.; Dupres, V. Automated Multi-Sample Acquisition and Analysis Using Atomic Force Microscopy for Biomedical Applications. *PLOS ONE* **2019**, *14* (3), e0213853. https://doi.org/10.1371/journal.pone.0213853.

(48) Yablon, D.; Chakraborty, I. *Machine learning to enhance atomic force microscopy analysis and operation*. Wiley Analytical Science.

(49) Alldritt, B.; Hapala, P.; Oinonen, N.; Urtev, F.; Krejci, O.; Federici Canova, F.; Kannala, J.; Schulz, F.; Liljeroth, P.; Foster, A. S. Automated Structure Discovery in Atomic Force Microscopy. *Science Advances* **2020**, *6* (9), eaay6913. https://doi.org/10.1126/sciadv.aay6913.

(50) Carracedo-Cosme, J.; Romero-Muñiz, C.; Pérez, R. A Deep Learning Approach for Molecular Classification Based on AFM Images. *Nanomaterials* **2021**, *11* (7), 1658. https://doi.org/10.3390/nano11071658.

(51) Müller, P.; Abuhattum, S.; Möllmert, S.; Ulbricht, E.; Taubenberger, A. V.; Guck, J. Nanite: Using Machine Learning to Assess the Quality of Atomic Force Microscopy-Enabled Nano-Indentation Data. *BMC Bioinformatics* **2019**, *20* (1), 465. https://doi.org/10.1186/s12859-019-3010-3.

(52) Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach Learn* **1995**, *20* (3), 273–297. https://doi.org/10.1007/BF00994018.

(53) Sokolov, I.; Dokukin, M. E.; Kalaparthi, V.; Miljkovic, M.; Wang, A.; Seigne, J. D.; Grivas, P.; Demidenko, E. Noninvasive Diagnostic Imaging Using Machine-Learning Analysis of Nanoresolution Images of Cell Surfaces: Detection of Bladder Cancer. *Proceedings of the National Academy of Sciences* **2018**, *115* (51), 12920–12925. https://doi.org/10.1073/pnas.1816459115.

(54) Vyas, V.; Solomon, M.; D'Souza, G. G. M.; Huey, B. D. Nanomechanical Analysis of Extracellular Matrix and Cells in Multicellular Spheroids. *Cell Mol Bioeng* **2019**, *12* (3), 203–214. https://doi.org/10.1007/s12195-019-00577-0.

(55) Taubenberger, A. V.; Girardo, S.; Träber, N.; Fischer-Friedrich, E.; Kräter, M.; Wagner, K.; Kurth, T.; Richter, I.; Haller, B.; Binner, M.; Hahn, D.; Freudenberg, U.; Werner, C.; Guck, J. 3D Microenvironment Stiffness Regulates Tumor Spheroid Growth and Mechanics via P21 and ROCK. *Adv. Biosys.* **2019**, *3* (9), 1900128. https://doi.org/10.1002/adbi.201900128.

(56) Abuhattum, S.; Mokbel, D.; Müller, P.; Soteriou, D.; Guck, J.; Aland, S. An Explicit Model to Extract Viscoelastic Properties of Cells from AFM Force-Indentation Curves. *iScience* **2022**, *25* (4), 104016. https://doi.org/10.1016/j.isci.2022.104016.

(57) Ren, K.; Gao, J.; Han, D. AFM Force Relaxation Curve Reveals That the Decrease of Membrane Tension Is the Essential Reason for the Softening of Cancer Cells. *Front Cell Dev Biol* **2021**, *9*, 663021. https://doi.org/10.3389/fcell.2021.663021.

(58) Schiavone, M.; Sieczkowski, N.; Castex, M.; Trevisiol, E.; Dague, E.; François, J. M. AFM Dendritips Functionalized with Molecular Probes Specific to Cell Wall Polysaccharides as a Tool to Investigate Cell Surface Structure and Organization. *Cell Surf* **2019**, *5*, 100027. https://doi.org/10.1016/j.tcsw.2019.100027.

(59) Elie-Caille, C.; Lascombe, I.; Péchery, A.; Bittard, H.; Fauconnet, S. Molecular and Nanoscale Evaluation of N-Cadherin Expression in Invasive Bladder Cancer Cells under Control Conditions or GW501516 Exposure. *Mol Cell Biochem* **2020**, *471* (1), 113–127. https://doi.org/10.1007/s11010-020-03771-1.

(60) Harris, M. C.; Cislo, D.; Lenz, J. S.; Umbach, C.; Lindau, M. AFM/TIRF Force Clamp Measurements of Neurosecretory Vesicle Tethers Reveal Characteristic Unfolding Steps. *PLoS One* **2017**, *12* (3), e0173993. https://doi.org/10.1371/journal.pone.0173993.

(61) Sundar Rajan, V.; Laurent, V. M.; Verdier, C.; Duperray, A. Unraveling the Receptor-Ligand Interactions between Bladder Cancer Cells and the Endothelium Using AFM. *Biophys J* **2017**, *112* (6), 1246–1257. https://doi.org/10.1016/j.bpj.2017.01.033.

(62) Fredonnet, J.; Foncy, J.; Lamarre, S.; Cau, J.-C.; Trévisiol, E.; Peyrade, J.-P.; François, J. M.; Séverac, C. Dynamic PDMS Inking for DNA Patterning by Soft Lithography. *Microelectronic Engineering* **2013**, *111*, 379–383.

(63) Moeendarbary, E.; Valon, L.; Fritzsche, M.; Harris, A. R.; Moulding, D. A.; Thrasher, A. J.; Stride, E.; Mahadevan, L.; Charras, G. T. The Cytoplasm of Living Cells Behaves as a Poroelastic Material. *Nat Mater* **2013**, *12* (3), 253–261. https://doi.org/10.1038/nmat3517.

Acknowledgements

We deeply thank C. Blatché and S. Assie-Souleille and the carac platform of LAAS-CNRS for the cell culture room support. The work was partly supported by the French Renatech network. And, in particular, we thank V. Fabre for helping us with SEM images. We also thank the BIOSOFT joint laboratory for the microcontact-printing part. This work was supported by grants from the Agence Nationale de la Recherche (ANR) AutoBioTip - ANR-20-CE42-0017. This work, bearing the reference EUR CARe N°ANR-18-EURE-0003, benefited from a government grant managed by the Agence Nationale de la Recherche under the Programme Investissements d'Avenir.

Table of contents

Graphical abstract: Automated bio-AFM generation of large mechanome data set and their analysis by machine learning to classify cancerous cell lines

Figure 1|. Influence of 4 parameters (cell patterning, indenter geometry, indenter velocity and number of measurements (FCs)/cell) on the histograms of the stiffness of PC3 cells obtained by force-distance measurements.

Figure 2|. Preparation of cells array.

Figure 3|. Cell biomechanical features extracted from the measured FCs and used as inputs for the machine learning algorithm.

Figure 4|. Conventional comparison of the 7 biomechanical features measured for the non-malignant (RWPE-1 or Hs 895.Sk) and the malignant (PC3-GFP or Hs 895.T) cell lines.

Figure 5|. Performance of LAMDA for the classification of PC3-GFP and RWPE-1 cells (A, B) and of Hs 895.T and Hs 895.Sk cells (C, D) as a function of the threshold number of FCs/cell assigned to one or the other category. The red bar represents a possible assignment of the threshold in order to avoid misclassification of cancerous cells (PC3-GFP and Hs 895.T cells).

Figure 6|. Performance validation of the LAMDA algorithm on RWPE-1 and PC3-GFP cell lines co-incubated on the same fibronectin array.

Supplementary Data

#1: Influence of 2 x 2 measurements per cell on k_{ext} of cells

#2: Script

#3: Video

#4: Data (314 RWPE-1, 314 PC3-GFP, 176 Hs 895.Sk and 152 Hs 895.T cells)

#5: Percentage common values

#6: Influence of the size of the fibronectin square patterns on the 7 features

#7: Comparison of the seven features used in the recognition model between Hs 895.Sk and Hs 8985.T cell lines seeded on 40 µm square fibronectin patterns

#8: Choice of features used by LAMDA to classify the mechanobiological data

#9: Training base size

#10: Algorithm choice