Electronic Supplementary Information:

Cell deformability heterogeneity recognition by unsupervised machine learning from in-flow motion parameters

Maria Isabella Maremonti[a], David Dannhauser[a*], Valeria Panzetta[a], Paolo Antonio Netti[a] and Filippo Causa[a*]

[a]Interdisciplinary Research Centre on Biomaterials (CRIB) and Dipartimento di Ingegneria Chimica, dei Materiali e della Produzione Industriale, Università degli Studi di Napoli "Federico II", Piazzale Tecchio 80, 80125 Naples, Italy.
*e-mail: causa@unina.it and david.dannhauser@unina.it

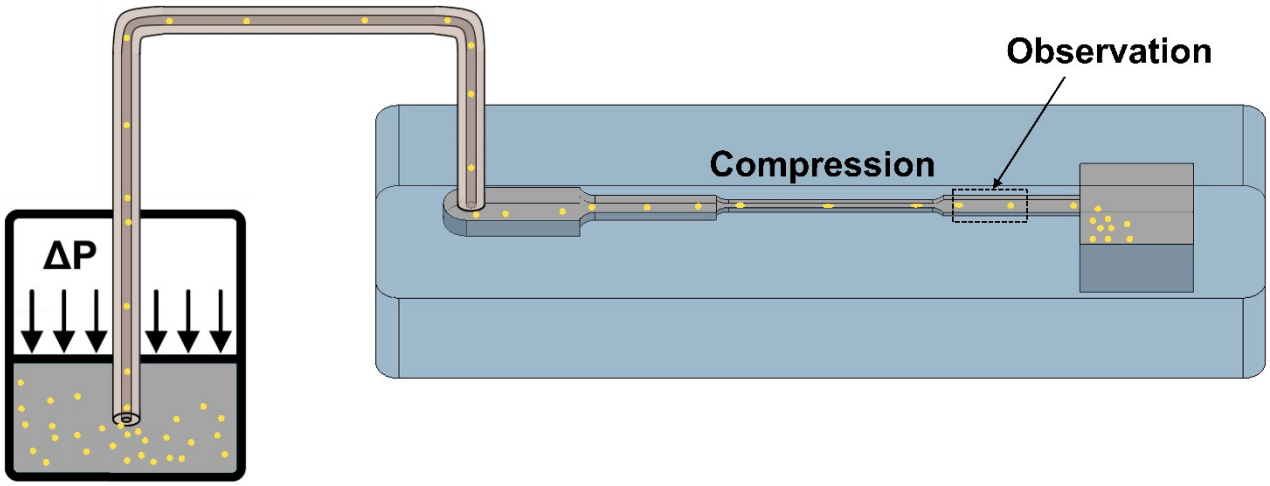

Fig. S1 Sketch representation of the microfluidic setup for inducing in-flow compressive forces. The sample is forced to pass into a capillary and then into the rectangular cross-sections, moving as a pressure-driven flow. A pressure drop (ΔP) is applied in order to move cells that flow into the capillary and then gradually arrive into the compression section where cells deform. Immediately after an expansion region where the compression section ends, an enlarged section for cell observation has been conceived to perform the overall dynamic analysis. Then, cells are guided at the exit of the device.

Table S1. Theoretical estimation of the in-flow compressive forces (F_{EMax}), where d_1 is the major cell diameter.

Cell Type	d_1 (μm)	F_{EMax} at PEO 05 (μN)
MCF-10A	13.634 ± 0.168	17.00
MDA-MB-231	14.798 ± 0.240	23.63

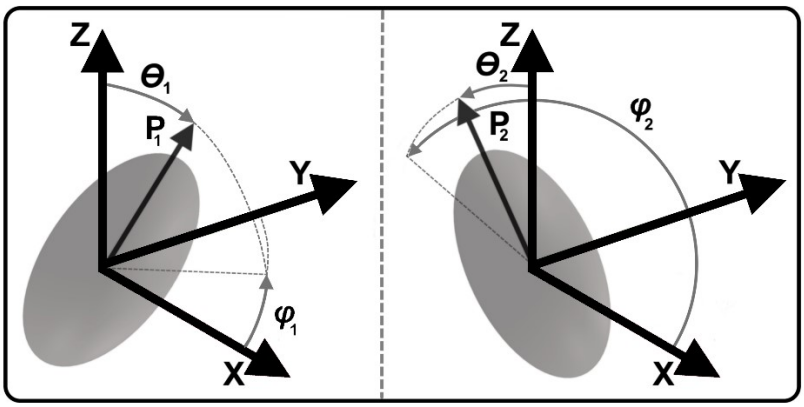

Fig. S2 Reference system for the orientation angles measurement (φ_1 and φ_2) with respect to the vector positions of P1 and P2, identifying the centroid of the cell.

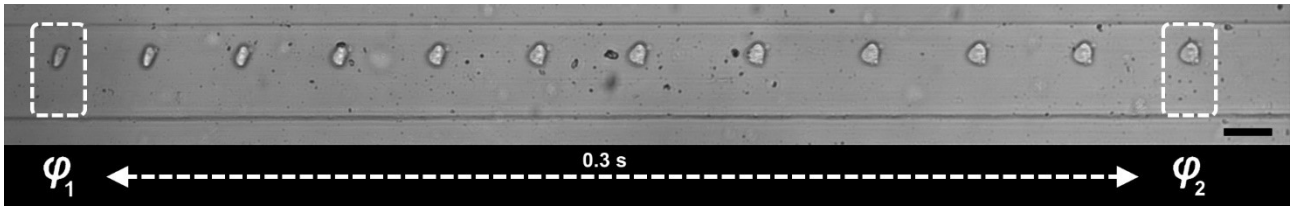

Fig. S3 Illustrative example of cell T for MDA-MB-231, looking at the two acquisition positions where φ_1 and φ_2 are measured. Between them, we collected the φ variation during the 0.3 s of acquisition. Scale bar 30 µm.

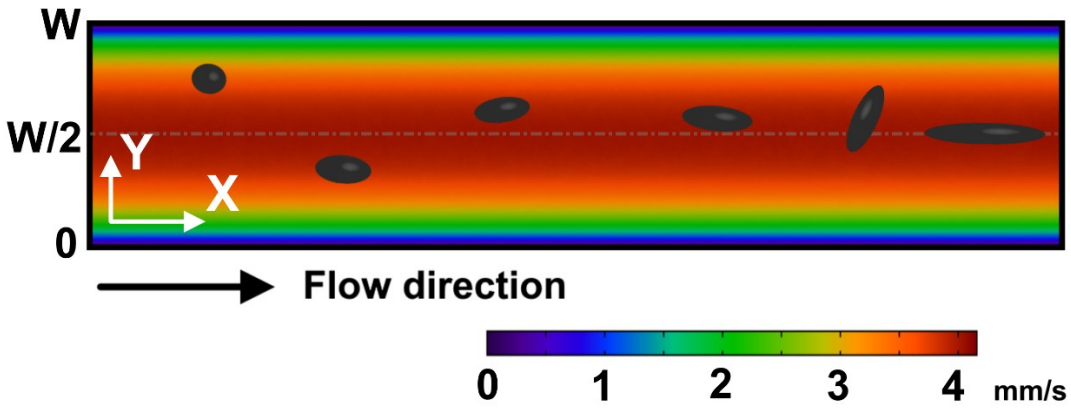

Fig. S4 Simulation results of the velocity profile developed into the observation section, where the cell alignment positions are detected. Black objects are sketch representations of cells with the different shapes and positions that we found into the measurements.

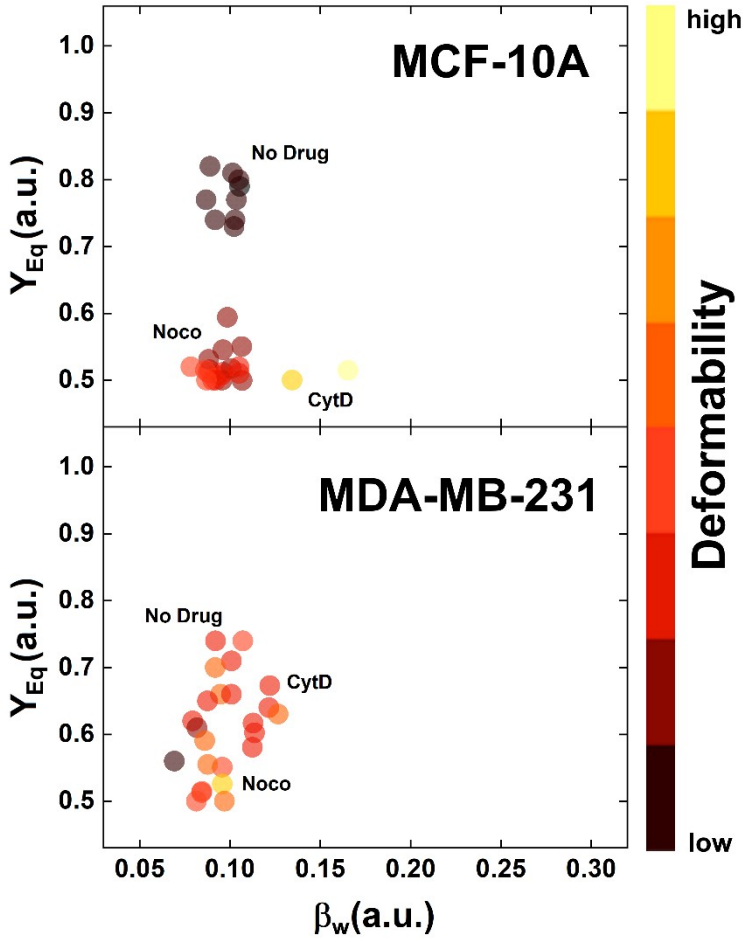

Fig. S5 Equilibrium positions (Y_{Eq}) variation with respect to the blockage ratio, computed with respect to the section width (β_W). As expected, the in-flow position is dominantly determined by the variable deformability instead of the cell dimension which is the same for both the investigated cell lines. The variable deformability has been induced by treatments of cytochalasin D (CytD) and nocodazole (Noco).

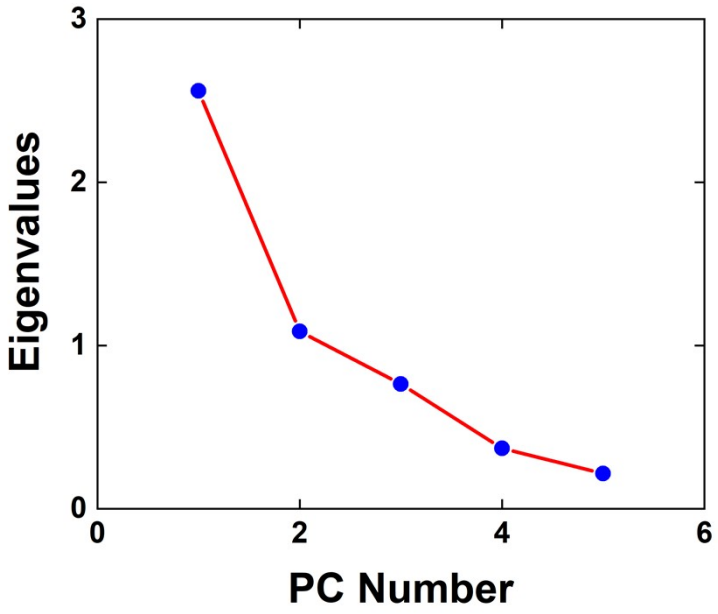

Fig. S6 Scree plot of the PCA analysis for the features of interest. For the Elbow rule, we consider at least three the components necessary to describe the set of data of interest.[1]

Table S2. Mean values and standard deviation of $\varphi(t)$ for each instant of time with Noco treatment.

Time (s)	MCF-10A T N= 4 (φ-Orientation)	MDA-MB-231 S N=3 (φ-Orientation)
0.03	-48.97 ± 20.94	63.08 ± 21.82
0.06	-47.49 ± 18.71	57.36 ± 23.13
0.09	-61.84 ± 22.80	52.39 ± 34.55
0.12	-64.65 ± 12.06	-41.05 ± 38.93
0.15	-33.74 ± 66.39	-50.11 ± 35.68
0.18	-42.15 ± 81.25	-53.53 ± 23.79
0.21	74.78 ± 12.84	
0.24	77.34 ± 16.35	
0.27	73.86 ± 12.64	
0.30	72.26 ± 12.03	

Table S3. Mean values and standard deviation of $\varphi(t)$ for each instant of time with CytD treatment.

Time (s)	MCF-10A TT N= 6 (φ-Orientation)	MDA-MB-231 T N=4 (φ-Orientation)
0.03	35.38 ± 12.11	82.60 ± 7.83
0.06	36.82 ± 12.99	80.96 ± 4.66
0.09	34.55 ± 11.14	82.26 ± 0.82
0.12	39.19 ± 14.53	83.64 ± 2.71
0.15	30.32 ± 7.39	79.09 ± 7.88
0.18	35.88 ± 10.08	-79.91 ± 10.54
0.21	38.93 ± 10.79	-79.98 ± 12.11
0.24	33.45 ± 10.45	-80.87 ± 11.58
0.27	36.21 ± 11.61	-80.97 ± 10.85
0.30		-90.00 ± 16.71

Table S4. Kruskal-Wallis statistical test for the experimental data. The difference is reported among the cells in No drug and treated conditions (Noco and CytD) for *AR*, *CD* and $\Delta\varphi$. In particular, we compare the two treatment outcome with the No drug condition. p<0.001*, p<0.01**, p<0.05*, p>0.05^{ns}.**

Cells	*AR*	*CD*	$\Delta\varphi$
MCF-10A - No Drug vs. Noco	***	***	***
MDA-MB-231 - No Drug vs. Noco	**	***	***
MCF-10A - No Drug vs. CytD	***	***	***
MDA-MB-231 - No Drug vs. CytD	**	***	ns

We quantified the separation or isolation of our resulting clusters from PCA by employing classical rules generally used for clustering applications. We computed the 'cluster diameter' and the 'cluster separation' as follows:[1]

$$Cluster\ Diameter = \ max_{j=r}^{i=r}\{d_{i,j}\}\ with\ i,j = 1,2,...,n$$

$$Cluster\ separation = min_{j=s}^{i=r}\{d_{i,j}\}\ with\ i,j = 1,2,...,n\ and\ r \neq s$$

Where *r* and *s* are clusters and *d* represent the distance between the points. To define clusters which are well separated and defined, except for outlier points that reduce such separation degree, the following formula has to be satisfied:

$$Cluster\ diameter\ \leq \max\{Cluster\ separation\}$$

We report in the following table the computed separation distance among clusters. We identified four clusters, one for each motion dynamics (Rolling (R), Tumbling (T), Tank-Treading (TT), Swinging (S)). In particular, the cases of No Drug MDA-MB-231, Cytochalasin D MDA-MB-231 and Nocodazole MCF-10A fall in the bigger T cluster.

Table S5. We report the computed separation distance among clusters. Precisely, we identified a cluster for each motion dynamics R, T, TT, S differentiating the cell classes and their intrinsic rheological/mechanical properties.[1]

Clusters to compare	**Separation satisfied:** $Cluster\ diameter\ \leq \max\{Cluster\ separation\}$
$R_{MCF\text{-}10A}$-$T_{MCF\text{-}10ANo}$	1.33 < 3.95
$R_{MCF\text{-}10A}$-$S_{MDA\text{-}MB\text{-}231No}$	1.33 < 2.91
$R_{MCF\text{-}10A}$-$TT_{MCF\text{-}10ACyt}$	1.33 < 3.6
$R_{MCF\text{-}10A}$-$T_{MDA\text{-}MB\text{-}231ACyt}$	1.33 < 1.51
$R_{MCF\text{-}10A}$-$T_{MDA\text{-}MB\text{-}231}$	1.33 < 2.99
$T_{MCF\text{-}10ANo}$-$S_{MDA\text{-}MB\text{-}231No}$	1.42 < 1.78
$T_{MCF\text{-}10ANo}$-$TT_{MCF\text{-}10ACyt}$	1.42 < 2.86
$T_{MCF\text{-}10ANo}$-$T_{MDA\text{-}MB\text{-}231}$	1.42 < 1.51
$TT_{MCF\text{-}10ACyt}$-$T_{MDA\text{-}MB\text{-}231}$	3.03 < 4.95
$TT_{MCF\text{-}10ACyt}$-$T_{MDA\text{-}MB\text{-}231Cyt}$	3.03 < 4.48
$T_{MDA\text{-}MB\text{-}231}$-$S_{MDA\text{-}MB\text{-}231No}$	1.97 < 2.79
$TT_{MCF\text{-}10ACyt}$-$S_{MDA\text{-}MB\text{-}231No}$	3.03 ~ 2.94
$T_{MCF\text{-}10ANo}$-$T_{MDA\text{-}MB\text{-}231Cyt}$	1.42 > 0.82
$T_{MDA\text{-}MB\text{-}231}$-$T_{MDA\text{-}MB\text{-}231Cyt}$	1.97 > 1.47

References

1 Desgraupes, Bernard. Clustering indices. University of Paris Ouest-Lab Modal' X, 2013, 1: 34 and Aggarwal, Charu C., et al. Data mining: the textbook. New York: springer, 2015.