

SCIENTIFIC REPORTS



OPEN

Geometric compensation applied to image analysis of cell populations with morphological variability: a new role for a classical concept

Joana Figueiredo^{1,2}, Isabel Rodrigues³, João Ribeiro³, Maria Sofia Fernandes^{1,2,3}, Soraia Melo^{1,2,4}, Bárbara Sousa^{1,2}, Joana Paredes^{1,2,4}, Raquel Seruca^{1,2,4} & João M. Sanches³

Immunofluorescence is the gold standard technique to determine the level and spatial distribution of fluorescent-tagged molecules. However, quantitative analysis of fluorescence microscopy images faces crucial challenges such as morphologic variability within cells. In this work, we developed an analytical strategy to deal with cell shape and size variability that is based on an elastic geometric alignment algorithm. Firstly, synthetic images mimicking cell populations with morphological variability were used to test and optimize the algorithm, under controlled conditions. We have computed expression profiles specifically assessing cell-cell interactions (IN profiles) and profiles focusing on the distribution of a marker throughout the intracellular space of single cells (RD profiles). To experimentally validate our analytical pipeline, we have used real images of cell cultures stained for E-cadherin, tubulin and a mitochondria dye, selected as prototypes of membrane, cytoplasmic and organelle-specific markers. The results demonstrated that our algorithm is able to generate a detailed quantitative report and a faithful representation of a large panel of molecules, distributed in distinct cellular compartments, independently of cell's morphological features. This is a simple end-user method that can be widely explored in research and diagnostic labs to unravel protein regulation mechanisms or identify protein expression patterns associated with disease.

Immunofluorescence (IF) microscopy is a widely used technique that uses fluorescent-labelled markers to visualize the distribution of proteins, glycoproteins and other molecular targets in intracellular structures, at the cellular level or at the tissue level^{1,2}. In the last years, different approaches have been developed to extract quantitative features from IF images and, in this way, to better understand the most complex cellular mechanisms^{3,4}. Image acquisition modalities, such as time-lapse microscopy, confocal laser scanning microscopy (CLSM) and spinning disk microscopy can offer quantitative analysis of a target protein; nonetheless, those techniques rely on measurements of total fluorochrome intensity, regardless of its distribution in an image or in a selected region⁵⁻⁷.

Recently, we have developed a bioimaging tool to assess the patterns of expression of *CDH1* germline missense variants associated to a cancer syndrome⁸. In that approach, the major analytical challenge was related to the heterogeneous morphology of cells in IF images. In fact, within the same cell population, it is possible to identify cells with very different shapes and sizes due to DNA replication error/mutations, epigenetic alterations, independent clonal evolution or different cell cycle stages^{9,10}.

Different morphological features will give rise to a high variability in the expression profiles, impairing the extraction of a representative overview/map of a particular target within a heterogeneous cell population. To overcome this constraint on endogenous cell-to-cell differences, we developed a geometric compensation model specific for *in situ* IF applications. Geometric compensation is a common procedure in several image analysis modalities, and typically consists on the estimation of rigid or non-rigid transformations to make the objects

¹Instituto de Investigação e Inovação em Saúde (i3S), Porto, Portugal. ²Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal. ³Institute for Systems and Robotics (ISR/IST), LARSyS, Bioengineering Department, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal. ⁴Medical Faculty of the University of Porto, Porto, Portugal. Correspondence and requests for materials should be addressed to R.S. (email: rseruca@ipatimup.pt) or J.M.S. (email: jmsr@tecnico.ulisboa.pt)

under alignment as similar as possible in terms of shape and size^{11–14}. This technique is essential for image reconstruction and fusion by improving the resolution of the raw data and increasing image analysis accuracy^{11–14}.

In this report, we describe an analytical pipeline which includes the extraction of internuclear (IN) and radial (RD) fluorescence profiles, and their accurate alignment. Specifically, the method applies a Bayesian non-rigid alignment algorithm and an automatic outlier rejection strategy to a large number of individual profiles, minimizing the undesired effect of size and shape variability within the cell population. In this way, the algorithm generates a final profile which is a distorted version of an ideal unknown profile, representative of all cells analysed within an IF image.

To experimentally validate our strategy, we applied this new algorithm to IF images of real cell cultures stained with E-cadherin, a key cell-cell adhesion molecule; tubulin, a major component of the eukaryotic cytoskeleton; and Mitotracker that labels mitochondria, which are complex cytoplasmic organelles responsible for the generation of energy in cells^{15–20}. Altogether, prototype markers of membrane, cytoplasmic and organelle-specific moieties, representative of distinct cellular compartments, were incorporated in our validation.

Results

Extraction of synthetic expression profiles. In imaging analysis, cellular morphological heterogeneity is a major challenge that needs to be addressed to obtain an accurate quantitative map of a tagged molecule in a cell population. Herein, we took advantage of an alignment algorithm to minimize the variability of synthetic and real fluorescence profiles, demonstrating the accuracy of our approach to achieve a typical picture and a precise expression profile of proteins in the populations analysed.

To achieve our goal, the strategy applied involved a number of specific steps. First, cells composing synthetic images mimicking heterogeneous cultures/tissues were automatically selected and connected using a bioimaging tool previously developed by our group (Fig. 1A,B)²¹. The algorithm generates a nucleus-nucleus network representing cell distribution across the image, in which the nodes are the geometric centres of the cell nuclei and the edges represent the neighbouring relation between them. The Delaunay tessellation algorithm automatically groups neighbour nodes in three element clusters (triangles), minimizing their total area and eccentricity. The networks are independent of the non-regular distribution of cells and avoid the need of repetitive and predictive patterns to define a neighbouring system. Noteworthy, for an accurate intensity mapping, cells should be confluent in a way that only neighbouring/adjacent cells are associated in triplets, forming a contiguous diagram. The presence of empty space between cells could lead to an erroneous interpretation of the data.

The connections retrieved from this networking process were then used to extract IN and RD profiles from cell pairs and individual cells, respectively. As showed in Fig. 1C, IN profiles consist of a set of quasi-parallel segments around the main axis linking the geometrical centres of two neighbour nuclei and measure fluorescence intensities occurring between two contiguous cells. This output is of particular relevance to evaluate proteins located at the plasma membrane or in specific cellular organelles.

In order to extract IN profiles, we have considered the diameters of the nuclei perpendicular to the IN axis, linking both geometric centres of the nuclei and their interceptions with the nuclei, x^A , x^B , y^A and y^B , where $x, y \in \mathbb{R}^2$. The starting and ending points, x_i and y_i , of the i^{th} straight line/profile within the beam are the following:

$$\begin{cases} x_i = x^A + \frac{i}{n-1}(x^B - x^A) \\ y_i = y^A + \frac{i}{n-1}(y^B - y^A) \end{cases} \quad (1)$$

where $i = 0 \cdots n - 1$ and n , the number of profiles within the beam, is typically $n = 10$.

The intensity of the j^{th} pixel of the i^{th} profile is $G(p_i(j))$, where $G(\cdot)$ is the image intensity of the channel of interest at the location

$$p_i(j) = x_i + \frac{j}{m_i - 1}(y_i - x_i) \quad (2)$$

Here, $j = 0 \cdots m_i - 1$. m_i is the length of the i^{th} profile, which is typically the distance between the geometric centres of the nuclei, c_A and c_B , $m_i = \|c_A - c_B\|$. Since the IN distances were different from pair to pair of cells and the next step was to package them in a single $N \times M$ matrix, where N is the total number of profiles and M is the length of them, an interpolation procedure with bi-cubic functions was performed over each extracted profile to normalize their length to M samples.

In contrast, RD profiles were designed to capture the intensity pattern throughout the cytoplasm of a single cell (Fig. 1D) and could be very useful for the analysis of cytoplasmic proteins. Indeed, the RD profiles correspond to a set of m equi-spaced angular profiles anchored at the centres of the nuclei. The length of each profile within the set depends on the neighbouring configuration in the vicinity of the cell. As shown in Fig. 1D, an interpolation spline passing by the adjacent cells defines the limit and the length of each profile. Further, as described for IN profiles, the different lengths of RD profiles were normalized using a bi-cubic interpolation operation, allowing the packaging of all profiles (from all cells) in a $N \times M$ matrix. Here, M is the normalized length of the profiles, $N = N_c m$ is the total number of RD profiles and N_c is the number of cells in the image.

The intensity of the j^{th} pixel from the i^{th} profile, extracted from the k^{th} cell, is $G(p_i(j))$ where $G(\cdot)$ is the image intensity at the location

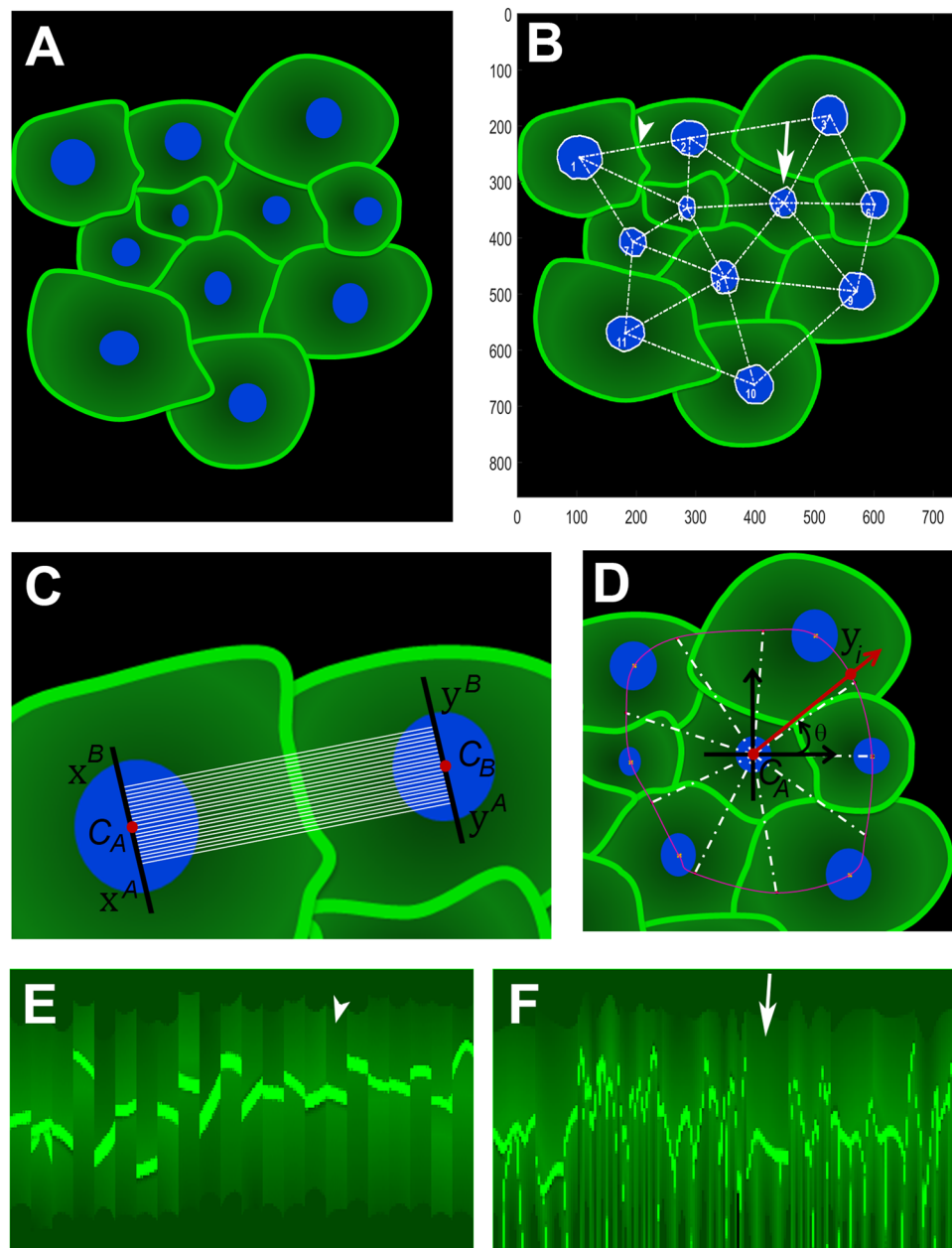


Figure 1. Profiling internuclear (IN) and radial (RD) expression in heterogeneous cell populations. **(A)** Synthetic image mimicking cellular heterogeneity, concerning size and morphology. **(B)** Automatic selection and networking of cells through nuclei segmentation and calculation of their geometric centroids. Arrowhead represents IN profiles and arrow corresponds to RD analysis. **(C)** IN profiles capture signal intensities occurring between two contiguous cells. **(D)** RD profiles encompass fluorescence patterns that spread from a nucleus centroid and cover the total area of a single cell. **(E,F)** IN and RD intensity maps, respectively, obtained from the original synthetic image presented in **(A)**.

$$p_i(j) = c_k + \frac{j}{r_i - 1}(y_i - c_k) \quad (3)$$

Here, $j = 0 \dots r_i - 1$. r_i is the length of the i^{th} profile, which corresponds to the distance from the centre of the cell to the interception of the corresponding radius vector (Fig. 1D) with the spline that passes through the cell centroids.

As demonstrated in Fig. 1E,F, IN and RD original maps were successfully obtained from synthetic IF images following this pipeline. However, their huge irregularity prevents obtaining a clear picture of the expression pattern represented in the original image.

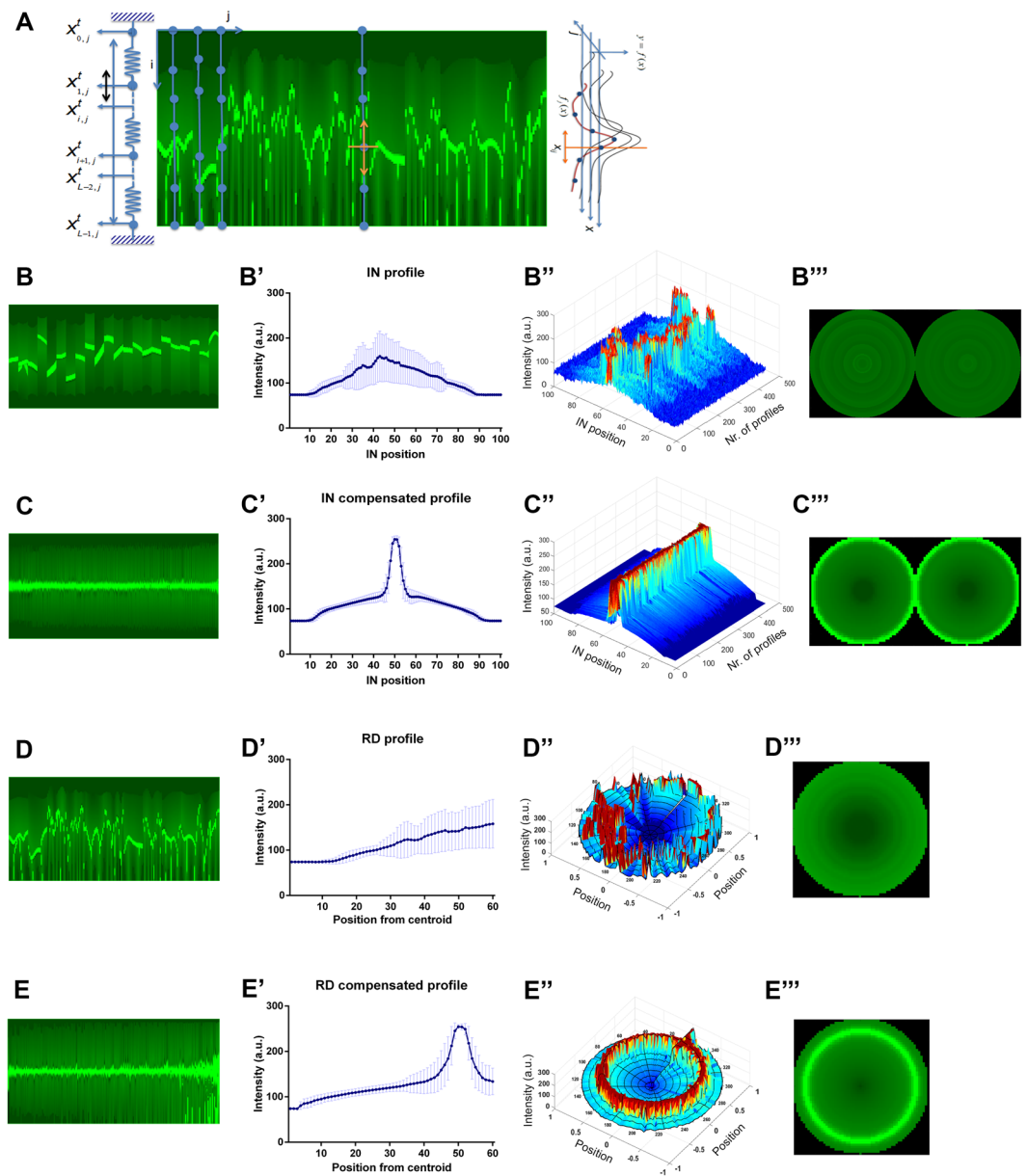


Figure 2. Geometrically compensated profiles reproduce signal patterning of synthetic images. (A) Strategy for geometric alignment. An iteration model of intensity's adjustment was applied to each profile from the intensity map, imposing a tension regularization term and, simultaneously, a similarity driven force. (B,C) non-compensated and compensated IN maps, respectively, extracted from the same synthetic image. (B',C') Mean and standard deviation (SD) of non-aligned and aligned IN profiles. (B'') and (C'') 3D overviews of all extracted IN profiles. (B''',C''') Virtual cell pairs illustrating signal distribution in non-aligned and aligned IN profiles. A similar analysis is presented in panels D and E for the corresponding non-compensated and compensated RD profiles. a.u., arbitrary units.

Alignment of synthetic profiles and image reconstruction. IN and RD profile maps obtained in the previous step are composed by stacks of profiles extracted from different cells and pairs of cells, with different sizes and shapes. Therefore, the panel of profiles is highly heterogeneous and, as expected, it is not possible to directly extract their typical protein distribution.

To overcome the variability of IN and RD maps, a geometric compensation method was developed (Fig. 2A), assuming that all profiles within each map are distorted versions of an unknown ideal profile that is representative of the protein distribution in the whole cell population.

For that purpose, we established an $N \times M$ matrix with $Y = \{y_{i,j}\}$ representing the map of profiles and $X = \{x_{i,j}\}$ holding the same dimensions of Y . The matrix contains the normalized locations of the intensities, $y_{i,j}$, along the profiles, where $x_{i,j} \in [0, 1]$. Importantly, the distance between cells is not constant, so the length of the extracted profiles may be different, especially in the case of IN profiles. Therefore, profiles were interpolated, using

a bi-cubic interpolation function, and converted into N length vectors suitable to be packed side-by-side in the matrix Y .

Matrix X contains the initial locations of the observations, $x_{i,j}$, that we have assumed to be evenly distributed in the interval $[0, 1]$ according to $x_{i,j} = i/(N - 1)$ with $i = 0, \dots, N - 1$. Each N length column of Y , y_j , was also assumed to be a distorted and non-uniformly sampled version of an ideal continuous profile, representative of the entire population, $f(x): \Omega \rightarrow R$, where $\Omega = [0, 1]$.

The distortion of each profile was described by the unknown monotonic function $g_j(x)$. The algorithm is based on the adjustment of the initial locations of the observations, $x_{i,j}$, in order to estimate the inverse of $g_j(x)$ and, consequently, the real locations of the observations, $x_{i,j}^* = g_j^{-1}(x_{i,j})$.

The ideal profile is, thus, a finite dimension continuous function, $f(x, \mathbf{c})$, where $\mathbf{c} = \{c_k\}$ is a vector of coefficients to be calculated. The estimation of \mathbf{c} , as well as the compensated locations, $x_{i,j}^* = g_j^{-1}(x_{i,j})$, were formulated according to the following optimization problem:

$$[\mathbf{c}, X]^* = \arg \min_{\mathbf{c}, X} E(X, \mathbf{c}, Y) \tag{4}$$

Here the energy function to be minimized is composed by three terms:

$$E(X, \mathbf{c}, Y) = E_Y(X, \mathbf{c}, Y) + E_c(\mathbf{c}) + E_X(X) \tag{5}$$

the data fidelity term E_Y ; the regularization term for \mathbf{c} , E_c ; and the prior term for the observed locations, E_X . This equation will induce similarity between neighbouring profiles within the map and, consequently, the alignment and geometric compensation of the profiles. A number of assumptions need to be highlighted and include:

Ideal Profile. The unknown function that describes the ideal profile to be estimated is assumed to be a finite dimension continuous function described by a linear combination of L ideal interpolation functions, $\phi_k(x) = \text{sinc}(x/\Delta - k)$ with $\Delta = (L - 1)^{-1}$ and $k = 0, 1, \dots, L - 1$.

$$f(x) = \sum_{k=0}^{L-1} c_k \phi_k(x) = \phi^T(x) \mathbf{c} \tag{6}$$

Here, $\phi(x) = \{\phi_0(x), \phi_1(x), \dots, \phi_{L-1}(x)\}^T$ is an L length column vector containing the interpolation functions computed at location x , and $\mathbf{c} = \{c_0, c_1, \dots, c_{L-1}\}^T$ is an unknown L length column vector of coefficients that needs to be estimated.

Because the IN profiles describe the typical intensity distributed from the cell A to the cell B, which is the same from B to A, the vector \mathbf{c} should be symmetric by imposing $\mathbf{c} = P\tilde{\mathbf{c}}$, where $\tilde{\mathbf{c}}$ is a $(L/2)$ length vector and P is the following $L \times (L/2)$ matrix

$$P = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & 0 & 0 & \dots & 0 \end{bmatrix} \tag{7}$$

The ideal profile can be described as following

$$f(x) = \phi^T(x) P\tilde{\mathbf{c}} \tag{8}$$

Noteworthy, the IN ideal profile undergo symmetry constraints, using P as defined in (7), while the RD ideal profile is not subjected to symmetry constraint, $P = I_L$, where I_L is the $L \times L$ identity matrix.

Data fidelity term. The common used additive white Gaussian noise (AWGN) model²² leads to the following data fidelity term

$$E_Y(X, \mathbf{c}, Y) = \sum_{i,j} \omega_{i,j} (f(x_{i,j}) - y_{i,j})^2 \tag{9}$$

where $\omega_{i,j}$ are outlier indicators,

$$\omega = \begin{cases} 1 & \text{valid observation} \\ 0 & \text{outlier} \end{cases} \tag{10}$$

The indicators are adaptively computed along the iterative process of estimation. In case the distance of j^{th} profile to the current estimation of $f(x)$, $\|f(x_j) - y_j\|_2^2$, is larger than a given threshold, the indicators corresponding to that column are set to zero, $\omega_{i,j} = 0$ with $0 \leq i \leq N - 1$. Hence, the profile is classified as an outlier and is not used to estimate $f(x)$. However, its location, x_j , is still updated and, in future iterations, it can be re-classified as valid data and be included in the estimation of $f(x)$.

From the models commonly used to describe intensities in fluorescence images, the Poisson distribution model was preferred, as the data is obtained with photon-limited and counting-based image acquisition processes, where a small amount of detected radiation and a huge optical/electronics amplification is involved²⁵. Further, assuming the independence between observations, the data fidelity term is symmetric of the log-likelihood function

$$E_Y(x, \mathbf{c}, Y) = -\log P(y|f(x, \mathbf{c})) = -\sum_{i,j} \log p(y_{i,j}|f(y_{i,j}, \mathbf{c})) \tag{11}$$

where $p(y|f(x, \mathbf{c}))$ is the Poisson distribution $p(y|f(x)) = (f(x)^y / y!)e^{-f(x)}$, with parameter $f(x, \mathbf{c})$, resulting in the following data fidelity term:

$$E_Y(x, \mathbf{c}, Y) = \sum_{i,j} \omega_j [f(x_{i,j}) - y_{i,j} \log(f(x_{i,j}))] \tag{12}$$

Function regularization. The solution of (4) that defines the function $f(x, \mathbf{c})$, representing the ideal population profile, was regularized using a quadratic penalty term to force smoothness of $f(x)$ defined in (6),

$$E_c(\mathbf{c}) = \alpha \sum_{k=0}^{L-1} (c_k - c_{k-1})^2 = \alpha \mathbf{c}^T \psi_L \mathbf{c} \tag{13}$$

Here,

$$\psi_L = \theta_L^T \theta_L \tag{14}$$

with the following $L \times L$ difference operator

$$\theta = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -1 & 1 \end{bmatrix} \tag{15}$$

Similarly, the smoothing regularization prior term for \mathbf{b} is

$$E_b(\mathbf{b}) = \alpha \sum_{k=0}^{L/2-1} (b_k - b_{k-1})^2 = \alpha \mathbf{b}^T \psi_{L/2} \mathbf{b} \tag{16}$$

Locations regularization. The optimization of the data fidelity term (11) concerning the observation locations, $x_{i,j}$, is an *ill-posed* problem that also needs to be regularized. A trivial solution would be the collapsing of all locations at the same point. So, to avoid that, the limits were kept fixed (not updated) - $x_{0,j} = 0$ and $x_{N-1,j} = 1$ - and a regularization term was introduced by imposing a tension force between the neighbouring locations in each profile

$$E_X(X) = \beta \sum_{i,j} (x_{i,j} - x_{i-1,j})^2 = \beta Tr[X^T \psi_N X] \tag{17}$$

Here, Tr denotes the *Trace* operator and ψ_N is the $N \times N$, as defined in (14) and demonstrated in Fig. 2A. At the end, the overall energy to be minimized is the following

$$E(X, \mathbf{c}, Y) = E_Y(X, \mathbf{c}, Y) + \alpha \mathbf{c}^T \psi_L \mathbf{c} + \beta Tr[X^T \psi_N X] \tag{18}$$

Optimization: The vectors of coefficients, \mathbf{c} , as well as the location of the observations $\mathbf{X} = \{x_{i,j}\}$ were estimated along an iterative process. Further, the steps for the minimization of a global energy function $E(X, \mathbf{c}, Y)$, alternate regarding \mathbf{c} and X until a stopping criterion is met,

$$\mathbf{c}^{t+1} = \underset{\mathbf{c}}{\operatorname{argmin}} E(X^t, \mathbf{c}, Y) \tag{19}$$

$$X^{t+1} = \operatorname{arg min}_X E(X, \mathbf{c}^{t+1}, Y) \tag{20}$$

The minimization step (19) is performed by solving $\nabla_{\mathbf{c}} E(X^t, \mathbf{c}, Y) = 0$. For gradient computation purposes, the data fidelity terms (9) and (12) can be defined as follows

$$E_Y(X, \mathbf{c}, Y) = (\phi^T(x)\mathbf{c} - y)^T \sum (\phi^T(x)\mathbf{c} - y) \tag{21}$$

Here, $\mathbf{x} = \text{vect}(X) = \{x_k\}$ is the vectorization of matrix X , $\phi(\mathbf{x})$ is an $L \times NM$ matrix where each column contains the vectors $\phi(x_k)$, and $\Sigma = \{\sigma_{i,j}\} = \{\omega_{i,j}\gamma_{i,j}\}$ is an $NM \times NM$ diagonal matrix with

$$\gamma_{i,j} = \begin{cases} 1 & \text{AWGN} \\ 1/(2f(x_{i,j}) + \epsilon) & \text{Poisson} \end{cases} \quad (22)$$

$\omega_{i,j}$ are the outlier indicators (as defined above) and $\epsilon = 10^{-6}$ is a small constant to prevent division by zero. The minimization step (19) is then performed:

$$\begin{aligned} \nabla_c E(X, \mathbf{c}, Y) &= \nabla_c E_Y(X, \mathbf{c}, Y) + \nabla_c E_c(\mathbf{c}) \\ &= (\phi(\mathbf{x})\Sigma\phi^T(\mathbf{x}) + \alpha\psi_L^T)\mathbf{c} - \phi(\mathbf{x})\Sigma\mathbf{y} = 0 \end{aligned} \quad (23)$$

This allows the generation of the following recursion

$$\mathbf{c}^{t+1} = (\phi(\mathbf{x}^t)\Sigma\phi^T(\mathbf{x}^t) + \alpha\psi_L^T)^{-1}\phi(\mathbf{x}^t)\Sigma\mathbf{y} \quad (24)$$

where \mathbf{x}^t is the current estimate of \mathbf{x} .

By including the symmetry constraint (8) used in the IN profiles, the following coefficients can be obtained

$$\tilde{\mathbf{c}}^{t+1} = (P^T\phi(\mathbf{x}^t)\Sigma\phi^T(\mathbf{x}^t)P + \alpha\psi_{L/2})^{-1}P^T\phi(\mathbf{x}^t)\Sigma\mathbf{y} \quad (25)$$

Subsequently, the minimization step (20), where the observation locations are updated, is applied by solving the following equation:

$$\frac{\partial E}{\partial x_{i,j}} = z_{i,j} + \beta(x_{i,j} - \bar{x}_{i,j}) = 0 \quad (26)$$

Here, $z_{i,j} = \gamma_{i,j}[f(x_{i,j}) - y_{i,j}] \dot{f}(x_{i,j})$ and $\bar{x}_{i,j} = (x_{i-1,j} + x_{i+1,j})/2$ are the average values of the neighbouring intensity locations. $\dot{f}(x_{i,j}) = \frac{df(x)}{dx} = \sum_k c_k \dot{\phi}_k(x) = \dot{\Phi}^T(x)\mathbf{c}$, being $\dot{\phi}_k(x) = d\phi(x)/dx$.

Upon the application of the fixed-point approach, the new locations of f^{th} profile, x_j , can be obtained as follows

$$x_j^{t+1} = \frac{1}{3} \left(\Omega x_j^t - \frac{1}{\beta} z_j \right) \quad (27)$$

Here, Ω is

$$\Omega = \begin{bmatrix} 1 & 2 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1 & 0 & \dots & 0 \\ 0 & 1 & 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 2 & 1 \end{bmatrix} \quad (28)$$

and the term Ωx_j^t represents a vector of the sum/average $x_{i-1,j} + x_{i,j} + x_{i+1,j}$ for each column profile x_j . This equation is driven directly from (26).

As demonstrated in the Fig. 2B,C, this compensation strategy originates IN maps showing an almost constant horizontal invariant pattern of fluorescence that represents high levels of protein expression at the membrane and lower levels homogeneously distributed at the cell cytoplasm. Confirming this observation, we verified that the maximum fluorescence intensity occurs at IN position 50 (Fig. 2C'), which corresponds to the plasma membrane shared between two contiguous cells (Fig. 2C''). Further, when compared with the non-compensated IN profile, the compensated profile presents a smaller variance at each position and a higher sharpness of the peak (Fig. 2B',B'',C' and C''), demonstrating a significant improvement in image interpretation and quantitative analysis.

Regarding RD profiles, we verified that these maps can capture more efficiently protein distribution along the cell cytoplasm, which is not evaluated with IN profiling only (Fig. 2C'',E''). In fact, RD profiles were able to acquire the fluorescence signal into the full extent of a cell, revealing low levels of the marker inside the cell and its accumulation in cell periphery (Fig. 2E'',E''').

These results indicate that the application of the alignment pipeline generates a rigorous profiling of image signals that can be statistically examined and virtually represented.

Algorithm validation in IF images of heterogeneous cell populations stained for membrane, cytoplasmic and mitochondria markers. To experimentally validate our analytical pipeline, we used real *in situ* immunofluorescence images of cell cultures stained for E-cadherin, tubulin and mitochondria, which were selected as the prototypes of membrane, cytoplasmic and organelle-specific markers¹⁵⁻¹⁸.

In Fig. 3A, a cell culture showing E-cadherin membranous expression is presented. In this cell population, non-aligned profiles yield intense E-cadherin peaks randomly distributed in the cytoplasm, precluding true meaningful conclusions about protein status, namely its mapping and level of expression (Supplementary Fig. S1). Nonetheless, after geometric compensation, the analysis of IN maps revealed a strong intensity peak at the cell-cell junction (87,69 a.u. at IN position 50), in accordance with E-cadherin normal appearance and

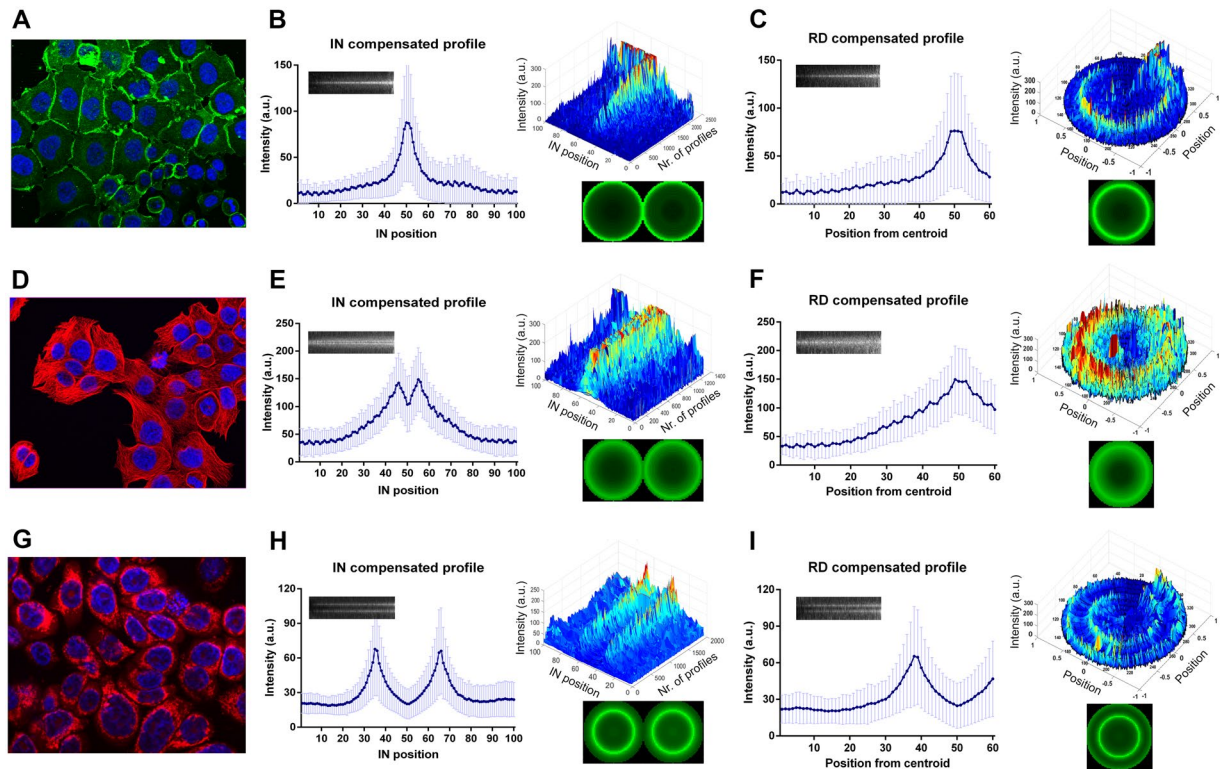


Figure 3. Method applicability in real immunofluorescence images of membrane, cytoplasmic and organelle-specific markers. (A) Immunofluorescence showing E-cadherin localization (green staining) in heterogeneous epithelial cells. Nuclei were counterstained with DAPI (blue). (B) E-cadherin IN profiles were extracted and geometrically compensated to evaluate protein distribution along contiguous cells. Compensated average intensity in each internuclear position \pm SD and its corresponding IN compensated map are presented in the linear graph. 3D graph showing the overview of all extracted profiles upon compensation. Virtual cell pair construction based on IN compensated profiles. (C) Average of E-cadherin compensated RD profiles \pm SD and its map are presented in the linear graph. Polar plot of all compensated RD profiles. 2D virtual cell illustrating E-cadherin distribution in the cell population. (D) Tubulin is stained in red and nuclei are marked in blue. (E) Average and map of tubulin IN compensated profiles. 3D graphical representation of IN compensated profiles and its virtual cell pair. (F) Average of tubulin RD profiles \pm SD. Dynamic overview of all compensated RD profiles and its virtual illustration. (G) Mitochondria staining in red and nuclei marked in blue. (H) Average and map of mitochondria IN compensated profiles. 3D graph of IN compensated profiles and its virtual representation. (I) Mitochondria radial analysis including RD profiles mean \pm SD. Overview of all compensated profiles and the respective virtual cell. a.u., arbitrary units.

adhesive functional role (Fig. 3B). In addition, besides the membrane expression of the protein, compensated RD maps also detect the presence of lower E-cadherin levels diffusely distributed in the cytoplasm (Fig. 3C). This cytoplasmic protein fraction corresponds to the E-cadherin that is continuously being synthesised, recycled and degraded by mechanisms of protein trafficking that occur in several cytoplasmic organelles/structures, namely endoplasmic reticulum, golgi complex and endosomes^{24–26}. Geometric compensation was also determinant for *in situ* image analysis of cytoplasmic and organelle specific proteins. For tubulin, a cytoskeleton component, it was possible to reveal a continuous increase in fluorescence from the position 20 till its maximum at the position 46, and a symmetric pattern between the positions 55 and 81 (Fig. 3D,E). Notably, at the position 50 that corresponds to the cell membrane, the protein level is reduced. This result is a precise overview of the IF image presented: a massive network is extended from the nucleus till a membrane-close region, where it polymerizes and accumulates. Indeed, microtubules are known to be present at the cytoplasm but not at the membrane¹⁸. The assessment of RD expression further confirmed these observations (Fig. 3F).

Regarding the mitochondria staining, the algorithm depicts a specific perinuclear expression pattern, either by IN or RD profiling (Fig. 3G–I). As showed in the dynamic IN and RD overviews, signals with maximum fluorescence intensities of 67,75 a.u. or 66,11 a.u. were restricted to a particular region of the cytoplasm that corresponds to positions 35 or 66 from the IN map (Fig. 3H,I). The remaining profile positions display much lower mean intensities (around 20 a.u.).

Overall, with this algorithm, we are able to achieve a precise and quantitative representation of the IF images of either membrane, cytoplasmic or organelle-specific markers, even in cases of highly heterogeneous cell cultures, such as those of cancer cells.

Discussion

Despite all the advances in the bioimaging field, immunofluorescence remains the technique of choice to determine the level and the spatial and temporal changes of fluorescent-tagged molecules^{3,27}. In the last years, an increased number of imaging methods have emerged, offering new possibilities to visualize and quantify fluorescent signals and allowing their dissemination in scientific research and clinical practice^{3,27}. However, although major progress was achieved, the quantitative analysis of *in situ* cell fluorescence images still faces important limitations, being the morphologic variability within cell populations a major one^{3,4}.

Cell morphologic variance/heterogeneity can result from a plethora of molecular events, such as changes in DNA sequence and status, distinct intercellular signalling, different cell cycle stages or altered cell behaviour^{9,10}. Indeed, the size and the morphology of a cell can change abruptly upon acquisition of a DNA mutation or upon modulation of a single protein^{28,29}.

In this work, we designed an analytical strategy to deal with cell shape and size variability that is based on a geometric compensation algorithm. With our approach, we were able to normalize cell size to a constant frame, and extract intensity profiles independently of cell morphological features. Ultimately, expression maps were generated reproducing an accurate level and a precise pattern of the target protein, in both synthetic and real cell cultures.

As a first step, we have computed two types of expression profiles, one focusing on cell-cell interactions (IN profiles) and the other directed to the distribution of a marker throughout the intracellular space of a single cell (RD profiles). The algorithm was shown to be successful in the extraction of both profile-types, nevertheless, their map compilation displayed scattered and undefined patterns of expression, which do not represent the typical profile of the cell population analysed. Therefore, an alignment model based on a geometric compensation algorithm was developed. By applying a classical image registration and alignment strategy, we set up a geometric compensation algorithm that detects common objects and reorient them in a way that corresponding data is paired^{12,30}. This method enforced a controlled normalization of intensity maps and revealed a final set of compensated profiles from which we can estimate the ideal distribution of the protein in the internuclear or radial axes. At the end, a compensated map epitomizing the pattern observed in the original synthetic image was produced. In fact, the benefits of registration procedures have been demonstrated in diverse medical software by improving the visual or quantitative interpretation of the results from magnetic resonance image (MRI), ultrasound, positron emission tomography (PET), single photon emission computed tomography (SPECT) or magnetic resonance spectroscopy (MRS)^{12,31}.

Subsequently, immunofluorescence of E-cadherin, tubulin and mitochondria was used to validate the applicability of our protocol in images of real cell cultures. Given that E-cadherin is the most important protein for the establishment and maintenance of cell-cell adhesion in epithelial tissues, it constitutes a classical example of a molecule strongly expressed at the plasma membrane^{15,16}. In contrast, tubulin is the basic structure of microtubules, a major cytoskeletal component and, therefore, the prototype of an abundant cytoplasmic protein³². The performance of the algorithm was also tested in images of well-defined subcellular compartments, such as mitochondria, which is the organelle responsible for the cell energetic supply¹⁹.

For all the markers analysed, the method was successful. The results demonstrated that our strategy enables an accurate quantification and mapping of membrane, cytoplasmic and organelle-specific proteins. Upon length normalization and stacking of fluorescent profiles together in columns, we were able to recognize high protein expression in the middle position of the E-cadherin profile – in accordance with its normal membrane appearance. For tubulin, a clearly distinct phenotype could be noticed: increasing fluorescence levels along the cell cytoplasm till membrane-proximal regions, without any site-specific preference. Indeed, it is well-known that the cytoskeleton is extended all over the cell and it polymerizes close to the plasma membrane, sustaining cell shape and connecting all intracellular organelles by a dynamic network^{18,32}. In contrast, in the case of mitochondria tracking, a sharp and intense peak is observed specifically in perinuclear positions, supporting its specialized and local metabolic function^{19,20}. Overall, we demonstrate that this method is suitable to a large panel of molecules distributed in very distinct cellular compartments.

Although the protocol yields a homogenous view of the cell population analysed, cell lines presenting different protein distribution patterns (mixed populations) can be evaluated in a single-cell based approach. Given that each centroid (nucleus centre) composing the triangular network is numbered, its corresponding data can be easily identified and extracted. As so, instead of using mean values of the whole population, we are able to analyze each cell clone separately.

Most of the available methods for quantification of immunofluorescence images evaluate total intensity or number of pixels present in the fluorochrome channel, disregarding the expression profile of the target protein along the distinct subcellular compartments. Nevertheless, the recognition of abnormal patterns of expression can provide valuable information for research and for diagnostic purposes. For example, if a protein that is normally expressed at the membrane is being accumulated at the cytoplasm, as a result of trafficking deregulation mechanisms, the fluorescence levels can be the same although its pattern can be remarkably different and indicative of protein dysfunction³³. In the last years, different bioimaging tools have been developed in an attempt to profile cell surface and intracellular markers along different cellular compartments^{34,35}. Mosaliganti and colleagues engineered an automated method to first reconstruct membrane signals and then segment out cells from 3D membranes for quantification³⁶. Still, the approach requires labelling of membrane boundaries and is only suitable for cell surface proteins³⁶. Another protocol generates a coupled multidimensional representation of spatial distribution for nuclear and membrane-bound proteins in a process highly dependent on nuclear and membrane segmentation, as well as on the continuity of fluorescent signals along cell surface boundaries³⁷. For the tracking of protein translocation between intracellular compartments, a system based on the average fluorescence changes over time was developed, using the variance instead of the raw fluorescence³⁸. The tool allows the detection of a change in a compartment, even if the total amount of the dye remains unchanged but this strategy is limited to comparative analysis with an initial image³⁸. In general, variation in background, signal discontinuity,

non-uniformity in the width and strength of the signal, in addition to cellular morphological heterogeneity constitute major issues hampering successful analyses^{34,35,37}. To overcome these limitations, cell segmentation procedures are usually required, increasing the complexity of the protocols and hindering their implementation and acceptance by biologists^{34–36}.

To the best of our knowledge, this is the first description of a pipeline for imaging analysis, using a geometric alignment strategy, to investigate protein phenotypic signatures. Importantly, our computational method copes with cell-to-cell morphological differences, avoiding complex segmentation procedures or the need of continuous readouts of fluorescent signals. In summary, we propose a novel application of geometric compensation for a robust and automatic quantitative expression analysis (level and mapping) of either membrane or cytoplasmic proteins. This is a simple end-user method that can be widely explored in research and diagnostic labs to unravel protein regulation mechanisms or identify protein expression patterns associated with disease.

Materials and Methods

Cell culture. MKN28, MDA-MB-468 and CHO cells stably transfected with a vector encoding the wild type E-cadherin (as described previously)³³ were cultured in RPMI, DMEM or α -MEM media (all from Gibco, Invitrogen), respectively, supplemented with 10% fetal bovine serum (HyClone, Perbio) and 1% penicillin/streptomycin (Gibco, Invitrogen). CHO Ecad cells were maintained under antibiotic selection with 5 μ g/ml blasticidin (Gibco, Invitrogen). All cell lines were grown at 37 °C and 5% CO₂ humidified air.

Fluorescence staining. Cells were seeded on 6-well plates on top of glass coverslips and grown for 48 h, in order to reach 70% confluence. For E-cadherin immunofluorescence, fixation was performed in ice-cold methanol for 20 minutes, while for tubulin, cells were fixed in 4% formaldehyde for 30 minutes. Cells fixed in formaldehyde were treated with 50 mM NH₄Cl for 10 minutes, washed with phosphate buffered saline (PBS), and permeabilized with 0.2% Triton X-100 in PBS for 10 minutes. Blocking was performed in 5% bovine serum albumin (BSA) in PBS for 30 minutes, at room temperature. Cells were, subsequently, incubated with E-cadherin mouse monoclonal (1:300 dilution, BD Biosciences) or anti- α -tubulin (1:1000, Sigma) antibodies for 1 h. The Alexa Fluor 488 goat anti-mouse or the Alexa Fluor 594 goat anti-mouse (both diluted 1:500, Invitrogen) were applied for 1 h in the dark, as secondary antibodies. MitoSOX Red (1.5 μ M, Molecular Probes) specifically targeting mitochondria-derived superoxide anion was applied at 37 °C for 30 minutes in live cells. In this case, fixation was performed thereafter in 4% formaldehyde. Coverslips were mounted on slides using Vectashield mounting medium with DAPI (Vector Laboratories). Images were acquired on a Carl Zeiss Apotome Axiovert 200 M Fluorescence Microscope (Carl Zeiss, Jena, Germany) with an AxioCam HRm camera, and processed with the Zeiss Axion Vision 4.8 software.

Synthetic images generation. Synthetic images mimicking heterogenous cell populations were generated in Matlab R2015b version. Geometric shapes, such as circles or ellipses, and free draw tools from the toolbox were used to produce reference patterns that simulate reorganization and distribution of cell-like objects. Intensity, contrast and hue were adjusted in each image, and Poisson noise was added in order to obtain synthetic images resembling fluorescence microscopy pictures.

Nuclei segmentation and network generation. Denoising and nuclei segmentation were performed as previously described²¹. Images were first subjected to a pre-processing pipeline of contrast enhancement and adjustment of image intensities in order to diminish background and increase signal-to-noise ratios. Specifically, the Otsu method and the Moore-Neighbour tracing algorithm, modified by Jacob's stopping criteria, were applied to each image for nucleus segmentation. In case of nuclei not properly segmented upon application of this protocol, nucleus manual fixation was performed using a computer-assisted mode. Nuclei geometric centre (v) was computed and its definition enabled the establishment of a segment connecting two neighbouring nuclei (ϵ) and, thus, the creation of an undirected graph, $G(v, \epsilon)$ by sequential association of other neighbours. A triangular network was designed using the Delaunay triangulation algorithm, which selects the mesh that maximizes the smaller angle of the triangles³⁹. Highly obtuse triangles, with $\varphi_k > \mu_\varphi + 3\sigma_\varphi$ or $\vartheta_k > \mu_\vartheta + 5\sigma_\vartheta$, were considered outliers and removed from the mesh.

Data analysis. Data analysis and scientific graphing was performed through Graph Pad Prism software version 6.01 (Graph Pad Software, San Diego, CA) and Matlab R2015b.

Computational power and process estimation. The analyses were performed in a regular computer with an Intel(R) Core(TM) i3 CPU M370@ 2.40 GHz, 6.00 GB RAM, 64-bit operating system and Windows 10 Home version 1709. The extraction of internuclear and radial profiles took less than 1 minute per image while the alignment and can take 5–20 minutes per image, depending on the number of cells that is present in each image. For large scale image analysis, intensity profiles should be first extracted for all images. Subsequently, a batch of mat-files containing non-aligned profile maps is submitted together for alignment under similar prior parameters.

References

- Lichtman, J. W. & Conchello, J. A. Fluorescence microscopy. *Nat Methods* **2**, 910–919, <https://doi.org/10.1038/nmeth817> (2005).
- Ntziachristos, V. Fluorescence molecular imaging. *Annual review of biomedical engineering* **8**, 1–33, <https://doi.org/10.1146/annurev.bioeng.8.061505.095831> (2006).
- Hamilton, N. Quantification and its applications in fluorescent microscopy imaging. *Traffic* **10**, 951–961, <https://doi.org/10.1111/j.1600-0854.2009.00938.x> (2009).
- Waters, J. C. Accuracy and precision in quantitative fluorescence microscopy. *J Cell Biol* **185**, 1135–1148, <https://doi.org/10.1083/jcb.200903097> (2009).

5. Muzzey, D. & van Oudenaarden, A. Quantitative time-lapse fluorescence microscopy in single cells. *Annu Rev Cell Dev Biol* **25**, 301–327, <https://doi.org/10.1146/annurev.cellbio.042308.113408> (2009).
6. Sandison, D. R., Williams, R. M., Wells, K. S., Strickler, J. & Webb, W. W. Quantitative Fluorescence Confocal Laser Scanning Microscopy (CLSM). *Handbook of Biological Confocal Microscopy*, 39–53 (1995).
7. Nakano, A. Spinning-disk confocal microscopy—a cutting-edge tool for imaging of membrane traffic. *Cell structure and function* **27**, 349–355 (2002).
8. Sanches, J. M. *et al.* Quantification of mutant E-cadherin using bioimaging analysis of *in situ* fluorescence microscopy. A new approach to CDH1 missense variants. *European journal of human genetics: EJHG* **23**, 1072–1079, <https://doi.org/10.1038/ejhg.2014.240> (2015).
9. Marusyk, A. & Polyak, K. Tumor heterogeneity: causes and consequences. *Biochim Biophys Acta* **1805**, 105–117, <https://doi.org/10.1016/j.bbcan.2009.11.002> (2010).
10. Rubakhin, S. S., Romanova, E. V., Nemes, P. & Sweedler, J. V. Profiling metabolites and peptides in single cells. *Nat Methods* **8**, S20–29, <https://doi.org/10.1038/nmeth.1549> (2011).
11. Fonseca, L. M. G. & Manjunath, B. S. Registration techniques for multisensor remotely sensed imagery. *Photogrammetric Engineering and Remote Sensing* **62**, 1049–1056 (1996).
12. Zitova, B. & Flusser, J. Image registration methods: a survey. *Image and Vision Computing* **21**, 977–1000 (2003).
13. Sanches, J. M. & Marques, J. S. Joint image registration and volume reconstruction for 3d ultrasound. *Pattern Recognition Letters* **24**, 791–800 (2003).
14. Li, S., Wakefield, J. & Noble, J. A. Automated segmentation and alignment of mitotic nuclei for kymograph visualisation. *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, 622–625 (2011).
15. van Roy, F. & Bex, G. The cell-cell adhesion molecule E-cadherin. *Cell Mol Life Sci* **65**, 3756–3788, <https://doi.org/10.1007/s00018-008-8281-1> (2008).
16. Paredes, J. *et al.* Epithelial E- and P-cadherins: role and clinical significance in cancer. *Biochim Biophys Acta* **1826**, 297–311, <https://doi.org/10.1016/j.bbcan.2012.05.002> (2012).
17. Janke, C. The tubulin code: molecular components, readout mechanisms, and functions. *J Cell Biol* **206**, 461–472, <https://doi.org/10.1083/jcb.201406055> (2014).
18. Musch, A. Microtubule organization and function in epithelial cells. *Traffic* **5**, 1–9 (2004).
19. Westermann, B. Mitochondrial fusion and fission in cell life and death. *Nat Rev Mol Cell Biol* **11**, 872–884, <https://doi.org/10.1038/nrm3013> (2010).
20. Schmidt, O., Pfanner, N. & Meisinger, C. Mitochondrial protein import: from proteomics to functional mechanisms. *Nat Rev Mol Cell Biol* **11**, 655–667, <https://doi.org/10.1038/nrm2959> (2010).
21. Mestre, T. *et al.* Quantification of topological features in cell meshes to explore E-cadherin dysfunction. *Scientific reports* **6**, 25101, <https://doi.org/10.1038/srep25101> (2016).
22. Moon, T. K. & Stirling, W. C. *Mathematical Methods and Algorithms for Signal Processing*. Prentice-Hall (2000).
23. Rodrigues, I. C. & Sanches, J. M. Convex total variation denoising of Poisson fluorescence confocal images with anisotropic filtering. *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society* **20**, 146–160, <https://doi.org/10.1109/TIP.2010.2055879> (2011).
24. Bryant, D. M. & Stow, J. L. The ins and outs of E-cadherin trafficking. *Trends Cell Biol* **14**, 427–434, <https://doi.org/10.1016/j.tcb.2004.07.007> S0962-8924(04)00172-2 [pii] (2004).
25. Delva, E. & Kowalczyk, A. P. Regulation of cadherin trafficking. *Traffic* **10**, 259–267, <https://doi.org/10.1111/j.1600-0854.2008.00862.x> (2009).
26. Yap, A. S., Crampton, M. S. & Hardin, J. Making and breaking contacts: the cellular biology of cadherin regulation. *Curr Opin Cell Biol* **19**, 508–514, <https://doi.org/10.1016/jceb.2007.09.008> (2007).
27. Odell, I. D. & Cook, D. Immunofluorescence techniques. *The Journal of investigative dermatology* **133**, e4, <https://doi.org/10.1038/jid.2012.455> (2013).
28. Yin, Z. *et al.* A screen for morphological complexity identifies regulators of switch-like transitions between discrete cell shapes. *Nat Cell Biol* **15**, 860–871, <https://doi.org/10.1038/ncb2764> (2013).
29. Suriano, G. *et al.* E-cadherin germline missense mutations and cell phenotype: evidence for the independence of cell invasion on the motile capabilities of the cells. *Hum Mol Genet* **12**, 3007–3016, <https://doi.org/10.1093/hmg/ddg316> [pii] (2003).
30. Brown, L. G. A survey of image registration techniques. *ACM Computing Surveys* **24**, 325–376 (1992).
31. Fischer, B. & Modersitzki, J. Ill-posed medicine—an introduction to image registration. *Inverse Problems* **24**, <https://doi.org/10.1088/0266-5611/24/3/034008> (2008).
32. Wade, R. H. & Hyman, A. A. Microtubule structure and dynamics. *Curr Opin Cell Biol* **9**, 12–17 (1997).
33. Figueiredo, J. *et al.* The importance of E-cadherin binding partners to evaluate the pathogenicity of E-cadherin missense mutations associated to HDGC. *European journal of human genetics: EJHG* **21**, 301–309, <https://doi.org/10.1038/ejhg.2012.159> (2013).
34. Dufour, A. *et al.* Signal Processing Challenges in Quantitative 3-D Cell Morphology: More than meets the eye. *IEEE Signal Processing Magazine* **32**, 30–40, <https://doi.org/10.1109/MSP.2014.2359131> (2015).
35. Ortiz-de-Solórzano, C., Muñoz-Barrutia, A., Meijering, E. & Kozubek, M. Toward a Morphodynamic Model of the Cell: Signal processing for cell modeling. *IEEE Signal Processing Magazine* **32**, 20–29, <https://doi.org/10.1109/MSP.2014.2358263> (2015).
36. Mosaliganti, K. R., Noche, R. R., Xiong, F., Swinburne, I. A. & Megason, S. G. ACME: automated cell morphology extractor for comprehensive reconstruction of cell membranes. *PLoS computational biology* **8**, e1002780, <https://doi.org/10.1371/journal.pcbi.1002780> (2012).
37. Han, J. *et al.* Multidimensional profiling of cell surface proteins and nuclear markers. *IEEE/ACM transactions on computational biology and bioinformatics* **7**, 80–90, <https://doi.org/10.1109/TCBB.2008.134> (2010).
38. Calmettes, G. & Weiss, J. N. A quantitative method to track protein translocation between intracellular compartments in real-time in live cells using weighted local variance image analysis. *PLoS One* **8**, e81988, <https://doi.org/10.1371/journal.pone.0081988> (2013).
39. Okabe, A., Boots, B., Sugihara, K., Chiu, S. N. & Kendall, D. G. *Definitions and Basic Properties of Voronoi Diagrams, in Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. John Wiley & Sons, 43–112 (2000).

Acknowledgements

This work was supported by FEDER funds through the Operational Programme for Competitiveness Factors (COMPETE) and National Funds through the Portuguese Foundation for Science and Technology (FCT), under the projects PTDC/BIM-ONC/0171/2012, PTDC/BIM-ONC/0281/2014, PTDC/BBB-IMG/0283/2014; Post-Doctoral grants SFRH/BPD/87705/2012-JF and SFRH/BPD/104208/2014-BS; and Doctoral grant SFRH/BD/108009/2015-SM. We acknowledge the Programa IFCT (FCT Investigator) for funding JP research. We also thank to the American Association of Patients with Hereditary Gastric Cancer “No Stomach for Cancer” for funding the projects “Today’s present, tomorrow’s future on the study of germline E-cadherin missense mutations” and “Today’s Present, Tomorrow’s Future on the Study of Germline E-Cadherin Missense Mutations: A Step Forward on Providing Informed Genetic Counseling to Everyone”.

Author Contributions

J.F. was responsible for the conception of the experimental system, analysis and interpretation of data, and wrote the manuscript. I.R. developed the algorithm and analysed the data. J.R., M.S.F., S.M. and B.S. were involved in the acquisition and analysis of data. J.P. analysed the data and contributed to the design of the project and manuscript preparation. R.S. and J.M.S. were responsible for the conception and design of the experimental system, the analysis and interpretation of data, and the review of the manuscript. All authors approved the manuscript final version.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-28570-z>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018