Abstract—Immunofluorescent (IF) staining is the gold standard method to quantify and map protein expression. Several manual qualitative scoring systems are used based on fluorochrome cellular distribution and intensity but such systems are subject to significant intra- and inter-observer variability. New image analysis software are available but their application to the cancer field is still very difficult due to the presence of confounding factors for the analysis, namely cell-to-cell morphology variability.

This work describes an accurate quantitative methodology to characterize the level and localization of protein expression in heterogeneous cell populations, such as cancer cells, evaluating the distribution of the molecules at intra and inter-cellular spaces from IF images. The proposed method does not require segmentation of the cells which is not accurate in cases where the membrane is not visible due to lack of protein expression.

The method is based on one dimension (1D) internuclear (IN) and radial (RD) intensity profiles, computed from a set of IF images, which are representative of the entire cell population.

The central component of the method is a non-rigid alignment algorithm with an automatic outlier rejection strategy to compensate the differences between inter-nuclear and radial profiles due to cell size and shape variability.

The algorithm is designed in a Bayesian framework with regularization constraints to deal with the ill-posed nature of problem. Two different observation models for white noise are considered under the same unified framework: the additive Gaussian and the (multiplicative) Poissonian ones.

The application of the method is illustrated with real IF images of in vitro cell cultures where a protein responsible for stabilizing adhesion complexes between cadherins and catenins (p-120) is stained with a specific antibody.

Index Terms—bio-imaging; non-rigid alignment; biological quantification; protein expression; fluorescence

Fig. 1. Original immunofluorescence image showing p-120 catenin staining in a cell population. The catenin is labelled with green fluorescence and nuclei are stained with Dapi (blue).

I. INTRODUCTION

Immunofluorescence (IF) (Fig. 1), where only the molecules tagged with fluorescent dyes are observed [1], is extensively used in biological and medical research for qualitative analyses of protein expression [2], [3], [4].

Extensive work is available in the literature describing algorithms for processing this type of images, specially for denoising [5], morphologic analysis [6], [7], contour extraction [8], cell tracking [9], [10], shape characterization and detection [11], [12], [13], object counting and segmentation [14], [15]. Image analysis algorithms for biological quantification were only developed for quantification of cell volumes and single cell movement analysis [16]. The commonest methods for single cell analyses [17] includes Fluorescence Resonance Energy Transfer (FRET), Fluorescence Correlation Spectroscopy (FCS), Fluorescence Recovery After Photobleaching (FRAP) and Fluorescence Lifetime Imaging Microscopy (FLIM) which are able to quantify parameters of local dynamic processes within a single cell[18].

A class of automatic methods was also described for the analyses of an entire population of cells [19]. In this case, the commonest goals are counting, statistical characterization of object size and shape, and tracking of a large amount of objects [20]. In [21] a method to characterize the E-Cadherin molecule at the cell membrane of an heterogenous population to discriminate between treatment groups is described. In this case the membrane of the cells is segmented.

A recent overview on morphodynamic characterization and quantification of the cell from 2D and 3D imaging can be found in [22] and [23] respectively.

In heterogeneous populations of cells the key difficulty in imaging quantification is dealing with the size and shape variability of the cells. To cope with these difficulties an algorithm based on Spherical Harmonics Transforms (SPHARM) is described in [13]. In [24] the authors propose two methods, average fluorescence intensity method (AFIM) and amount of fluorescent pixels method (AFPM), for quantifying and evaluating the amount of surface receptors within a group of cells from fluorescence microscope images. However, only synthetic data was presented to test the accuracy of the methods.

Accurate segmentation of the membranes could be of great help to deal with this variability but it is difficult, time consuming [25] or even impossible when the membrane is not visible by lack of protein expression. In our images of transfected cells the expression of the molecule, e.g. p-120, can be very different from the wild type ones and the expression at the membrane can even be absent, thus invisible. In this case membrane segmentation is not feasible.

In this work we propose a method to characterize and quantify the distribution of tagged fluorescent molecules in the intra and intercellular space applied to a population of...
cells as shown in Fig. 1.

The clear benefit of this methodology is its ability to cope with the morphological variability of a cell population where a one dimensional (1D) representative profile of protein distribution is computed.

Here, a detailed mathematical formulation of the problem is presented where significant improvements with respect to [26] were introduced, such as, 1) the observation model that is here assumed Poissonian; 2) the denoising process is embedded in the main algorithm of geometrical compensation and 3) an adaptive mechanism of outliers rejection.

The method was illustrated with images of heterogeneous cell populations stained for p-120 catenin, a protein responsible for stabilizing cadherins-catenins adhesion complex. We obtained representative quantitative profiles of the level of expression of the molecule even in highly heterogeneous populations of cells.

II. BIOLOGICAL DATA

Cadherins are transmembrane proteins responsible for cell adhesion and forming adherent junctions. These proteins depend on Calcium ions $Ca^{2+}$ to function. They mediate calcium-dependent cell-cell adhesion, localized to the adherences-type junctions [27]. Generally, cadherins behave as receptors and ligands for other molecules, the intracellular domain binds directly to cytoplasmic catenins forming tight complexes which link them with the actin cytoskeleton. This feature provides the molecular basis for stable cell interactions maintaining cell and tissue structure. The cadherin/catenin complex, as well as the signalling pathways controlled by this structure, represent a major regulatory mechanism that influence cell growth, differentiation, motility, and survival [27], [28]. Two of the main cadherins, E-cadherin and P-cadherin, and its influence on cell interaction and tissue structure was studied through the expression of p-120, a protein responsible for the stabilization of the interaction between cadherins and catenins at the cell membrane. The co-expression of both cadherins has effects on several levels such as cell-cell adhesion, invasion capacity and stabilization of cadherin-catenin complex. In terms of cell aggregation, this feature can be observed with the existence of either cadherin while the invasive behaviour, however, can be due to the lack of expression or the co-expression of both cadherins. Furthermore, the expression of each cadherin individually increases significantly the interaction at the cell membrane between the cadherin and the catenins when compared to cells expressing both E-cadherin and P-cadherin [29]. This last observation is of extreme importance to this study. E- and P-cadherin heterodimers induce a more aggressive cell behaviour and have increased invasive and tumourogenic potential when compared with cells expressing only one of the cadherins. The efficiency in the stabilization of strong cadherin-catenin complexes at the cell membrane is significantly decreased. In fact, studies demonstrate that the expression of P-cadherin disrupts the normal invasive suppressor function of E-cadherin [29].

III. METHODOLOGY

From each IF image (Fig. 1) pairs of cells were selected for quantification of protein expression. This first step intended to exclude all negative cells that may represent technical pitfalls related to IF immunostaining. A specific software application was produced to assist the operator in this task.

The quantification of the expression level of the protein in a global and regional basis is performed by computing 1D intensity internuclear (IN) and radial (RD) profiles that represent the typical distribution of the molecule between or within cells in a given population.

The main purpose of the IN profiles is to measure the average expression level of the protein between pairs of neighboring cells including at the inter-cellular space. In case of cell adhesion molecules, as E-cadherin, or cell receptors this is of particular importance due to their outside-inside function (Fig. 2 a) and b)). These IN profiles are also able to capture the typical distribution of protein expression along the medial axis of cell pairs, namely at the cytoplasm, as well as to measure the average distance between two cells that are next to each other.

The RD profiles were tuned to measure the average expression level of the protein at the cytoplasm, mainly for quantification and mapping of aberrant foci of expression located outside of the internuclear axis and therefore impossible to be captured by the IN profiles. Results were obtained for various angles by extracting several RD profiles, anchored at the geometrical centers of the nuclei of selected cells (Fig. 2 c) and d)).

During the process of acquiring IN and RD profiles in the different cell lines, two main technical hitches can be found: i) population variability concerning cell size and shape and ii) segmentation of the cell boundaries. In some cells the protein does not stain at the membrane due to transient trafficking of protein at the cytoplasm or loss or decreased level of expression near the cell membrane, which is a common event occurring in cancer [29].

To cope with cell size and shape variability, a geometric compensation algorithm was designed in a Bayesian framework. Geometric compensation is a common procedure in several image modalities, mainly for registration purposes [30]. The general strategy in this type of algorithms consists on the estimation of a geometric transformation, rigid or non-rigid [31], by optimizing a metric of similarity to make the objects under alignment as similar as possible in terms of shape, size and orientation [32]. In microscopy imaging this is usually done for segmentation and tracking purposes [33], [34].

Here, the alignment algorithm is composed by the following steps: i) profile extraction from selected single cells (in case of RD) or pairs of cells (in case of IN) as illustrated in Fig. 2 a) and c); ii) image map building by stacking these profiles together in columns of the map after length normalization as shown in Fig. 2 b) and d); iii) geometric compensation of each 1D column profile aiming at minimizing the overall variability of the map along the lines (horizontal direction) with embedded denoising process of the map image with the algorithm described in [5] (where multiplicative noise is
described by a Poisson distribution) and iv) obtained profile analysis and metrics computation from the compensated map.

Since the aim of the strategy was to determine the typical level and pattern of protein expression at the intra and intercellular space, the segmentation of the cell was not a fundamental issue to consider in this work and therefore was not addressed in this study.

![Figure 2](image.png)

Fig. 2. Internuclear and Radial Profile Extraction. a) Pairs of selected nuclei, b) Final IN Map, c) Nucleus centered region of interest, d) Final RD profile

### A. Problem Formulation

Let \( Y = \{y_{i,j}\} \) be a \( L \times M \) map of intensity profiles where each one of its \( M \) columns, \( y_j \), corresponds to a \( L \) length single profiles extracted from a pair of selected cells, as illustrated in Fig. 2. Let also \( X = \{x_{i,j}\} \), with the same dimensions of \( Y \) with \( L \) length columns \( x_j \), be the corresponding locations of the intensities \( y_{i,j} \) along the profiles where \( x_{i,j} \in [0, 1] \).

Since the distance between cells is not constant, the length of the extracted profiles may be different, specially in the IN case. Therefore, they are interpolated and converted into \( L \) 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 are assumed to be evenly distributed in the interval \([0, 1]\) according \( x_{i,j} = i/(L - 1) \) with \( i = 0, \ldots, L - 1 \). Each \( N \) length column of \( Y \), \( y_j \), is assumed to be a distorted and non uniformly sampled version of an ideal continuous profile, representative of the entire population, \( g(x) : \Omega \rightarrow R \), where \( \Omega = [0, 1] \).

The distortion of each profile is described by the unknown monotonic function \( f_j(x) \). The core of the proposed algorithm resides in the adjustment of the initial locations of the observations, \( x_{i,j} \), in order to implicitly estimate the inverse of \( f_j(x) \), and consequently the real locations of the observations, \( x^*_i,j = f_j^{-1}(x_{i,j}) \).

The estimation of the vector of coefficients \( c \) as well as the compensated locations, \( x^*_i,j = f_j^{-1}(x_{i,j}) \), is formulated according the following optimization problem,

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

where the energy function to be minimized is:

\[
E(X, c, Y) = E_Y(X, c, Y) + E_c(c) + E_X(X).
\]

In this equation \( E_Y \) is the data fidelity term, \( E_c \) is the regularization term for \( c \) and \( E_X \) is the prior term for the observation locations that induces similarity between neighbouring profiles in the map of profiles promoting, by this way, the alignment and geometric compensation of the profiles.

1) 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 \( N \) ideal interpolation functions, \( \phi_k(x) = \text{sinc}(x/\Delta - k) \) with \( \Delta = (N - 1)^{-1} \) and \( k = 0, 1, \ldots, N - 1 \),

\[
g(x) = \sum_{k=0}^{N-1} c_k \phi_k(x) = \phi^T(x)c
\]

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

In the case of the IN profiles, because they describe the typical intensity profile from cell A to cell B which is the same from B to A the vector \( c \) should be symmetric which can be imposed making \( c = Pb \) where \( b \) is a \( (N/2) \) length vector and \( P \) is the following \( N \times (N/2) \) matrix

\[
P = \begin{bmatrix}
1 & 0 & 0 & \ldots & 0 \\
0 & 1 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 1 & 0 \\
0 & 0 & \ldots & 0 & 1 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 0 & 0 & \ldots & 0
\end{bmatrix}.
\]

The ideal profile is the following

\[
g(x) = \phi^T(x)Pb
\]

that is used to describe the IN ideal profile, under symmetry constraints, using \( P \) as defined in (4) and to describe the RD ideal profile, without symmetry constraint, making \( P = I_N \) where \( I_N \) is the \( N \times N \) identity matrix.

2) Data Fidelity Term: The common used additive white Gaussian noise (AWGN) model [35] leads to the following data fidelity term

\[
E_Y(X, c, Y) = \sum_{i,j} \omega_{i,j} (f(x_{i,j}) - y_{i,j})^2
\]
where $\omega_{i,j}$ are outlier indicators

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

(7)

that are adaptively computed along the iterative process of estimation. If along the iterative process the distance of the $j^{th}$ profile to the current estimation of $c$, $\|f(x_j) - c\|_2^2$, is larger than a given threshold, the indicators corresponding to that column are set to zero, $\omega_{i,j} = 0$ with $0 \leq i \leq L - 1$. In this case the profile is classified as an outlier and is not used to estimate $g(x)$. However, its locations, $x_j$, are still updated.

The AWGN model is not the more appropriated to describe the intensities in fluorescence images. Poisson distribution is preferred when the data is obtained with photon-limited and counting based image acquisition processes where small amount of detected radiation and huge optical and electronics amplification is involved [5], which is the case of IF images. Assuming independence between observations the data fidelity term is the symmetric of the log-likelihood function

$$E_Y(x, c, Y) = -\log P(Y|f(x, c)) = -\log \prod_{i,j} p(y_{i,j}|f(x_{i,j})) = \sum_{i,j} [f(x_{i,j}) - y_{i,j}\log(f(x_{i,j}))]$$

(8)

where $p(y|f(x, c))$ is the Poisson distribution $p(y|f(x)) = (f(x)^y/y!)e^{-f(x)}$ with parameter $f(x, c)$, which leads to the following data fidelity term,

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

(9)

3) Function regularization: The solution of (1) with respect to the vector of coefficients, $c$, that defines the function $g(x, c)$ representing the ideal underlying population profile, is regularized using a quadratic penalty term to force smoothness of $g(x)$ defined in (3),

$$E_c(c) = \beta \sum_{k=0}^{N-1} (c_k - c_{k-1})^2 = \beta c^T \Psi_N c$$

(10)

where

$$\Psi_N = \theta_N^T \theta_N$$

(11)

with the following $N \times N$ difference operator

$$\theta = \begin{bmatrix} 1 & -1 & 0 & \ldots & 0 \\ -1 & 1 & 0 & \ldots & 0 \\ 0 & -1 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 & -1 \end{bmatrix}.$$  

(12)

The smoothing regularization prior term, as before, for $b$ is,

$$E_b(b) = \beta \sum_{k=0}^{N/2-1} (b_k - b_{k-1})^2 = \beta b^T \Psi_{N/2} b$$

(13)

4) Locations regularization: The optimization of the data fidelity term (8) with respect to 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. To avoid this, the limits are kept fixed (not updated), $x_{0,j} = 0$ and $x_{L-1,j} = 1$, and a regularization term is introduced by imposing a tension force between the neighboring sample locations in each profile

$$E_X(X) = \gamma \sum_{i,j} (x_{i,j} - x_{i-1,j})^2$$

$$= \gamma Tr [X^T \Psi_N X]$$

(14)

where $Tr$ denotes the Trace operator and $\Psi_N$ is the $N \times N$ matrix where each column contains the vectors of coefficients, $x_{i,j}$, that are adaptively computed along the iterative process of estimation.

Assuming independence between observations the data fidelity terms (6) and (9) can be written as follows

$$E(X, c, Y) = \min_c E(X, c, Y) + \beta c^T \Psi_N c + \gamma Tr [X^T \Psi_N X]$$

(15)

B. Optimization

The vectors of coefficients, $c$, as well as the location of the observations $X = \{x_{i,j}\}$ are estimated along an iterative process where the steps for the minimization of a global energy function, $E(X, c, Y)$, with respect to $c$ and $X$ alternate until a stopping criterion is met,

$$c^{t+1} = \arg \min_c E(X^t, c, Y)$$

$$X^{t+1} = \arg \min_X E(X, c^{t+1}, Y)$$

(16)

(17)

The minimization step (16) is performed by solving $\nabla_c E(X^t, c, Y) = 0$. For gradient computation purposes the data fidelity terms (6) and (9) can be written as follows

$$E_Y(X, c, Y) = (\Phi^T(x)c - y)^T W^T W (\Phi^T(x)c - y)$$

(18)

where $x = vect(X) = \{x_k\}$ is the vectorization of matrix $X$, $\Phi(x)$ is a $LM \times N$ matrix where each column contains the vectors $\phi(x_k)$, $W = \{W_{i,j}\} = \{\omega_{i,j}\alpha_{i,j}\}$ is a $LM \times LM$ diagonal matrix with

$$\alpha_{i,j} = \begin{cases} 1 & \text{AWGN} \\ \frac{1}{(f(x_{i,j}) + \epsilon)} & \text{Poisson} \end{cases}$$

(19)

$\omega_{i,j}$ are the outlier indicators defined before and $\epsilon = 10^{-6}$ is a small constant to avoid division by zero. The minimization step (16) is performed as follows

$$\nabla_c E(X, c, Y) = \nabla_c E_Y(X, c, Y) + \nabla_c E_c(c)$$

(20)

$$= (\Phi(x)^T W^T W (\Phi^T(x)c + \beta \Psi_N^{-1} c) - \Phi(x)^T W y = 0$$

from which the following recursion is obtained

$$c^{t+1} = (\Phi(x^t)^T W^T (\Phi(x^t) + \beta \Psi_N^{-1} c) - \Phi(x)^T W y$$

(21)

where $x^t$ is the current estimate of $x$.

By including the symmetry constraint (5) used in the IN profiles in (21) the following coefficients are obtained

$$b^{t+1} = (P^T \Phi(x^t)^T W^T (\Phi(x^t) + \beta \Psi_{N/2}^{-1} P^T \Phi(x^t) W y$$

(22)

The minimization step (17), where the observation locations
are updated, is performed by solving the following equation,

\[ \frac{\partial E}{\partial x_{i,j}} = z_{i,j} + \gamma (x_{i,j} - \bar{x}_{i,j}) = 0, \quad (23) \]

where \( z_{i,j} = \frac{\alpha_{i,j} [f(x_{i,j})-y_{i,j}] f(x_{i,j})}{3 x_{i-1,j} + x_{i,j} + x_{i+1,j}} \) is the average values of the neighboring intensity locations. \( f(x_{i,j}) = \frac{df(x)}{dx} = \sum_k c_k \phi_k(x) = \Phi^T(x) c \) where \( \phi_k(x) = \frac{df(x)}{dx} \).

Using as before the fixed point approach, the new locations of \( j^{th} \) profile, \( x_j \), can be obtained as follows

\[ X^{t+1} = \frac{1}{3} \left( \Omega X^t - \frac{Z}{\gamma} \right) \quad (24) \]

where \( \Omega \) is the \( L \times L \) matrix,

\[ \Omega = \begin{bmatrix} 1 & 2 & 0 & 0 & \ldots & 0 \\ 1 & 1 & 1 & 0 & \ldots & 0 \\ 0 & 1 & 1 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 2 & 1 \end{bmatrix} \quad (25) \]

In (24) the term \( \Omega X^t \) represents a matrix containing the sum/average \( x_{i-1,j} + x_{i,j} + x_{i+1,j} \) for each column profile \( x_j \). This equation is derived directly from (23).

IV. EXPERIMENTAL RESULTS

In this section results obtained from real data to illustrate the application of the method are presented.

Normalization, detection of nuclei centers, geometric compensation of the sample profiles and typical profiles are computed from the data automatically extracted from the selected cells or pairs of cells without any operator intervention.

Real IF images of cells expressing the protein p-120 were used to illustrate the application and validate the proposed method for the analysis of the spatial distribution and the level of expression of molecules independently of cell morphology. Three experimental conditions were analysed: expression of p-120 in the presence of E-cadherin (Mock), expression of p-120 in the presence of both E-cadherin and P-cadherin (P-cad), and the expression of p-120 in the presence of both cadherins when a drug, dasatinib is administrated. Dasatinib is an oral tyrosine kinase inhibitor namely the Src family tyrosine kinase with the purpose of inhibit the degradation of p-120 molecules.

A. Data

In the analysis of real data, cell lines expressing p-120 were used. In all images the green plane displays the expression level of the molecules (in this case the protein catenin p-120) that we want to quantify and the blue channel contains the DAPI information at the nuclei. The red channel is empty.

The IN and RD profile maps, extracted from the cells and pairs of cells semi-automatically selected by the operator from the original IF image (Fig. 4 a)), are displayed in Figs. 4 b) and c) respectively. The profiles, before map building, were normalized to a constant length of \( L = 100 \), by interpolation, in order to produce the rectangular matrix/image of profiles. An example of the geometrically compensated versions of these maps with the proposed method are displayed in Figs. 4 d) and e).

**Internuclear map (IN)**

The columns of the IN non compensated map contain sets of parallel intensity profiles along the axis of several pairs of cells. These profiles are different, as shown in Fig. 4 b), due to cell population heterogeneity in terms of morphology and nucleus position. To compensate for these variations, the geometric alignment (non-rigid transformation) algorithm was applied and the results are displayed in Fig. 4 d). The compensated map shows an almost constant horizontal invariant pattern of fluorescence that represents p-120 expression near cell membrane where it normally concentrates. The typical profile, computed by averaging the columns of these maps, and the corresponding standard deviation information, before and after geometrical compensation are displayed in Fig. 5 a) and b) respectively.

When compared with the non compensated profile, displayed in Fig. 5 a), the compensated one, displayed in Fig. 5 b), presents a smaller variance at each location and a higher sharpness of the peak. This shows the ability of the
Fig. 4. Real immunofluorescent images of tagged p-120. a) Original plaque, b) IN profiles from selected cells, c) RD profiles from selected cells, d) geometric compensated IN map of profiles, e) geometric compensated RD map of profiles.

proposed method to quantify protein expression levels in cell populations minimizing the effects of its heterogeneity.

Radial map (RD)

RD profiles aim at representing the overall distribution of the p-120 to capture aberrant p-120 expression foci not located along the cell-cell linkage axis. Because cells are not perfect circles and the nuclei do not coincide with their geometrical centers its polar representation is not a perfect straight line, as shown in Fig. 4 c).

The non compensated and compensated RD profile maps are displayed in Figs. 4 c) and e) respectively. The non compensated and compensated average profiles as well as the corresponding variability (STD) are represented in Fig. 5 c) and Fig. 5 d) respectively.

As observed in the case of the IN profile, the compensated average profile presents a sharp peak that accurately represents the p-120 near the membrane.

Biological Experiments

On figure 6 the internuclear calculated profiles for each experimental condition are represented. On Mock cells, the internuclear ideal profile is a symmetric peak around the centre of the profile with a flat edge on both sides. On these cells were only E-cadherin is expressed, p-120 protein is expressed near membrane level, peak area, acting as mediator on the E-cadherin-catenins adhesions. Both edges corresponds to the centre of the nucleus and therefore relevant expression of p-120 is not expected. On Pcad cells the central peak disrupts into separate peaks. Besides peak has shifted, the intensity has also decreased indicating a decreased in protein concentration. This represents the migration of p-120 protein to cytoplasm in the presence of P-cadherin protein. When both E-cadherin and P-cadherin are expressed, the activation of the Src kinase pathway and a concomitant decrease expression of p-120 at the membrane occurs. On Pcaddas Cells the peak is present again near the cell membrane and it intensity has also increased. This findings indicates an increase in protein concentration. The dasatinib treatment causes an inhibition in Src kinase and therefore it is possible to see that the expression of p-120 protein in the membrane has been recovered. It is possible to observe that the dasatinib has driven the recuperation of the expression of p-120 by acting on the suppression of Src Kinase.

Fig. 5. Geometric compensation. Average profile (red) and error bars (blue) representing the standard deviation.

Fig. 6. Geometric compensation. Obtained profiles for the three experimental conditions. Mock cell are represented in blue, Pcad cells are represented in green and Pcaddas in red.
V. Conclusions

In this work we propose a methodology for the quantification of the level of expression and cellular distribution of proteins at intra and intercellular space in a population of cells from IF images. In order to allow the computation of a typical intensity distribution profile of molecules tagged with fluorescence dyes, excluding variations of cell morphology from the analysis, the process combines a series of steps. The processing pipeline was composed by steps of 1) cell selection; 2) profile extraction and length normalization; 3) geometrical compensation to cope with cell shape and size variability and 4) typical expression profile computation.

The selection of cell images is a semi-automated procedure conducted by the operator to retrieve from the slides information with biological meaning and not related to technical pitfalls. Geometric compensation procedures is a core step of our proposed methodology making it resistant to the geometric variability of the cells. This allows computation of IN and RD profiles to accurately describe the pattern and quantify the level of expression of the protein in cell populations.

The proposed method implements a non rigid transformation of each profile that is described by a finite dimension continuous 1D function. The coefficients of the basis functions describing the sample profiles are estimated by using smooth constraints that regularize the original intrinsically ill-posed estimation problem. The non rigid transformation is performed by moving the locations of the observations under similarity constraints of neighbouring profiles that were stacked in the 2D map of profiles. The method is designed as an iterative algorithm which optimizes an overall energy function composed by three terms: a data fidelity term, that pushes the solution to the observations, and two prior terms corresponding to smoothness and similarity constraints.

Examples of IF real images of cells expressing p-120 were used to illustrate the application of the method. On IN Profiles information is more explicit than RD when analysing the expression of p-120 protein. The maximum intensity of expression diminishes from Mock to Pcad cells corroborating the hypothesis that when both E-cadherin and P-cadherin are expressing both cadherins.

We envision that our methodology can be used in a computer aid diagnosis (CAD) framework for semi-automatic detection of dysfunctional proteins that can be useful in clinical practice for screening, diagnosis and theragnosis purposes in various diseases.

References


