TÉCNICO LISBOA

Molecular distribution characterization with geometric invariant profiles in fluorescence microscopy

Sofia Raquel Pereira de Sousa Esménio

Thesis to obtain the Master of Science Degree in

Biomedical Engineering

Examination Committee

Chairperson: Professor João Pedro Estrela Rodrigues Conde
Supervisor: Professor João Miguel Raposo Sanches
Co-Supervisor: Doctor Maria Raquel Campos Seruca
Member of the Committee: Professor Jorge dos Santos Salvador Marques

July 2013

” I believe that whatever we do or live for has its causality; it is good, however, that we cannot see through to it.”

Albert Einstein

Acknowledgments

First of all, I would like to thank Prof. João Miguel Sanches, for the opportunity of joining this project, and for all the guidance and support. I would also like to thank my co-supervisors Joana Figueiredo and Raquel Seruca, of the IPATIMUP, for all the corrections, clinical explanations and specially for the development of the *in vitro* assays from which emerged the images that were processed in this work.

To my colleagues at room 6.14, Vânia Relvas, Catarina Barata, Margarida Ruela, Joana Pinto, Carlos Santiago, Sven Rothlubbers, Rodrigo Abreu, Miguel Rodrigues thank you for all the support. Thank you for the laughts, the good music, the english words/synonyms, the *Matlab* help, the dailly fun facts and company.

A special thanks to João Marques, for the advices/corrections throughout this work, and also for the thesis template. Thank you David Afonso for the terrific help implementing the GUI interface.

To everyone of my biomedical collegues and friends for always being there to cheer me up. Your help over the last years has been inestimable, and I only wish to cherish it even further. Thank you.

A special thank you to my mother and father for all the advices and support both emotionally and financial during the entire course, and to my familly for helping me with their sense of humour, concern and affection.

Abstract

Structural and mechanical properties of the tissues are dependent on the physical linkage between cells. Alterations in the adhesion properties between cells endow them with an invasive and migratory phenotype, causing E-cadherin as a key molecule in cell adhesion mechanism to play an important role as an invasion suppressor. Mutations on the E-cadherin gene (CDH1), lead to changes in E-cadherin expression and function which are implicated in key steps of tumor progression. Functional E-Cadherin is synthesized at the Golgi, transported and stabilized at the membrane and continuously recycled. When dysfunctional molecules are produced, e.g due to mutations, E-cadherin trafficking may be affected, changing the cellular localization of the protein and consequently the expression pattern.

This work proposes a method to characterize E-Cadherin distribution in the cell, specially along the junction line of cell-to-cell pairs, for functional discrimination purposes. To correctly characterize E-cadherin phenotype, a multi-step process was developed. This process combines a Graphical user interface (GUI) developed for nuclei selection, a segmentation and centroid detection step, Radial Profiles (RDPs) and Internuclear Profiles (INPs) extraction, and a Geometric Compensation (GC) algorithm. In this work, the profiles extracted are anchored in the geometrical centers of the cell nuclei selected. This way no information about membrane location is used in the preprocessing, which is an important point as membrane segmentation may be difficult, inaccurate and time consuming. The GC algorithm, the key issue of this thesis, is developed to compensate differences in shape and size in these profiles that are not directly related with the traffic dynamics. Additionally, quantitative measures are proposed for characterization of the distribution, e.g mean intensity at the membrane, mean and variance of molecules concentration in the cytoplasm. The discriminative power of these features in recognizing functional and dysfunctional molecules is shown.

Keywords

Bioimaging, Geometric Compensation, Biological Quantification, Bayesian Estimation, Fluorescence microscopy, E-cadherin phenotype

Resumo

As propriedades estruturais e mecânicas dos tecidos dependem da ligação física entre células. Alterações na adesão entre células dota as células de um fenótipo invasivo e migrador, fazendo com que a E-caderina como molécula principal no mecanismo de adesão celular desempenhe um papel importante como supressor de invasão. Mutações no gene da E-caderina (CDH1) levam a alterações na expressão e função da mesma que estão envolvidas em etapas chave da progressão tumoral. A E-caderina funcional é sintetizada no Golgi, transportada e estalizada na membrana e continuamente reciclada. Quando são produzidas moléculas disfuncionais, por exemplo devido a mutações, o tráfico de E-caderina pode ser afetado, mudando a posição celular das proteínas e consequentemente o padrão de expressão. Este trabalho propõe um método para caracterizar a distribuição de E-caderina nas células, especialmente ao longo da linha de junção celular, para uma posterior discriminação funcional. Para caracterizar o fenótipo da E-caderina, um processo de vários passos foi desenvolvido. Este processo combina: uma interface gráfica para a seleção de núcleos, um passo de segmentação e detecção de centróide, extração de perfis radiais e internucleares e um algoritmo de compensação geométrica. Neste trabalho, os perfis são extraídos apartir dos centros geométricos dos núcleos selecionados. Deste modo nenhuma informação acerca da localização membranar é utilizada no preprocessamento, que é importanto visto que a segmentação membranar pode ser difícil, imprecisa e demorada. O algoritmo de compensação geometrica, tema essencial nesta tese, é desenvolvido para compensar nestes perfis diferenças biológicas na forma e tamanho não relacionados com a dinâmica do tráfico molecular. Adicionalmente, são propostas medidas quantitativas para caracterização da distribuição, por exemplo intensidade média na membrana, média e variância de concentração de moléculas no citoplasma. O poder discriminativo destas características no reconhecimento de moléculas funcionais e disfuncionais é apresentado.

Palavras Chave

Processamento de imagens biológicas, Compensação Geométrica, Quantificação Biológica, Estimativa Bayesiana, Microscopia de fluorescência, E-caderina fenótipo

Contents

List of Figures

List of Tables

Abbreviations

DAPI 4',6-diamidino-2-phenylindole

ROI Region of Interest

BSA Bovine serum albumin

PBS Phosphate Buffered Saline

GUI Graphical user interface

GC Geometric Compensation

FM Fluorescense microscopy

ERAD Endoplasmic Reticulum Associated Degradation

FOI Fluorochrome of interest

GFP Green fluorescence protein

IF Immunofluorescence

MI Maximum Intensity

TV Total Variation

MSTD Maximum Standard Deviation

STD Standard Deviation

MMR Maximum Mean Ratio

WT Wild Type

RD Radial

RDP Radial Profile

RDM Radial Map

IN Internuclear

INM Internuclear Map

INP Internuclear Profile

FC Flow Cytometry

AJ Adherens Junctions

GCr Gastric Cancer

HDGC Hereditary Diffuse Gastric Cancer

DGC Diffuse Gastric Cancer

AA acid residues

GEFM Global E-cadherin Form Map

CAD Computer Aid Diagnosis

1

Introduction

Contents

1.1 Motivation

Hereditary Diffuse Gastric Cancer (HDGC) is an autossomal-dominant, inherited cancer syndrome in which the individuals affected with this disorder develop Diffuse Gastric Cancer (DGC) at a young age. This syndrome is characterized clinically by either 1) ≥ 2 documented cases of DGC in 1^{st} or 2^{nd} degree relatives with ≥ 1 diagnosed with age ≤ 50; 2) > 3 documented cases of DGC in 1^{st} or 2^{nd}degree relatives independent of the age of onset. CDH1 heterozygous alterations are the only germline event known in HDGC [4].

CDH1 gene is localized on chromosome 16q22.1 and is the gene responsible for encoding E-cadherin. E-cadherin is a Ca^{2+} dependent cell-adhesion glycoprotein predominantly expressed at the basolateral membrane of epithelial cells, essential in the maintenance of cell differentiation and epithelial tissues architecture. The downregulation of E-cadherin results in a loss of cell adhesion, invasion and metastasis. Given that in HDGC CDH1 dysregulation leads to tumor formation and progression, CDH1 is considered to act as a tumor suppressor gene [1].

Studies of penetrance show that CDH1 mutation carriers lifetime risk of developing DGC is approximately 70% and affected woman have an additional $20\% - 40\%$ risk of developing lobular breast cancer [5]. It is estimated that HDGC comprise $1 - 3\%$ of all Gastric Cancer (GCr), being gastric cancer the sixth most common cancer and the third leading cause of cancer death worldwide (Globocan, assessed 1^{st} May 2013). The Globocan results for cancer incidence and mortality are displayed in fig. 1.1. Although the overall rates of GCr have declined in the last years, DGCs represent a percentage that have not diminished in incidence.

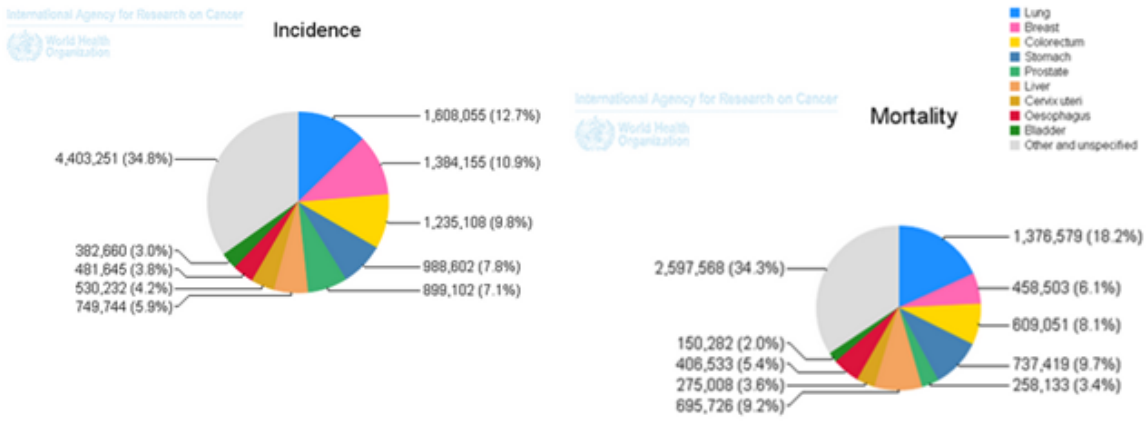


Figure 1.1: Cancers incidence and mortality (Globocan, assessed 1^{st} May 2013).

The major problem with HDGC syndrome carriers is that it is not possible to determine which CDH1 mutation carriers will develop DGC and the majority of DGCs become symptomatic only when their situation is incurable. Nowadays, due to the inadequacy of clinical screening and the high cancer penetrance of DGCs a prophylactic gastrectomy is recommended to asymptomatic carriers of CDH1 truncating germline mutations, to reduce cancer mortality. [6]. In this case a total gastrectomy is recommended because cancer cells in HDGC are multifocal and are distributed throughout the entire stomach. Although it is a lifesaving procedure, it has significant risks, namely: an estimated overall

mortality of $2\% - 4\%$; a 100% risk of long-term morbidity including diarrhea, dumping, weight loss and difficulty eating. The major problem is that this prophylactic gastrectomy procedure has been performed on several patients who had no reported evidence of gastric cancer on on screening biopsies [5].This highlights the importance of a reliable genetic screening that identifies non-functional E-cadherin molecules that induce an increased risk of gastric cancer in individuals that are carriers of germline CDH1 mutations.

This work has the objective of creating a method that allows the digital quantification and characterization of the expression pattern of membrane proteins at inter-cellular level, namely the cell membrane. In the proposed method several quantitative objective features that reflect E-cadherin expression level are extracted to characterize the distribution both at membrane and cytoplasmic level. One challenge that still remains is to clarify the pathogenicity of E-cadherin missense mutants associated to HDGC. This topic is addressed in this work, since the germline E-cadherin mutant forms implicated in HDGC were used as disease model in the *in vitro* assays developed. This procedure was developed to give information that reflects the functional activity of this particular protein, leading to a better understanding of the functional results of some structure alterations. This information is going to be used in a Computer Aid Diagnosis (CAD) framework for semi-automatic detection of dysfunctional proteins that can be used in clinical practice for screening and diagnosis .If this is possible these results may impact the quality of the information given within the genetic counseling sessions and in the decision making plan how to survey carriers of E-cadherin mutations improving their quality of life and decreasing Health Care associated clinical costs.

1.2 State of The Art

Familial forms of Gastric Cancer are known since the 1800s when multiple cases were detected on Bonaparte family. However, its was only on 1998 that Guilford et al. recognized the molecular basis of the first syndrome with proven inherited defect. The Hereditary Diffuse Gastric Cancer (HDGC) syndrome is associated with CDH1 germline mutations [7]. CDH1 gene is responsible for encoding E-cadherin, a surface protein member of the cadherin superfamily, discovered in 1977 by Takeichi with Ca^{2+} adhesion potential [8]. It was in 1988 that Takeishi and Nagafuchi discover the first direct evidences that the function of this cell-adhesion molecule is regulated by the cytoplasmic region. [9]. E-cadherin is an adhesive molecule responsible for the maintenance of cell differentiation and normal architecture of epithelial tissues. The revocation of its function induces a decrease of cell adhesiveness and loss of adherens junctions, leading to an abnormal morphogenesis and architecture of the epithelial tissues. Several studies, at the end of the 20^{th} century (1990s), were carried to understand how E-cadherin downregulation influenced tumorigenesis and tumor progression [10–12]. Further evidences supporting that E-cadherin abolished expression leads to tumorigenesis came from the study of families with an aggregation of DGC [6, 13, 14]. In HDGC, 25% of the cases CDH1 are missense mutations whose pathogenic significance is still difficult to established. To date more than thirty CDH1 germline missense mutations were found related with HDGC.

CDH1 germline missense mutations pathogenic relevance is still a clinical burden. To settle this, *in silico* models were developed over the time to dissect each mutation's molecular impact [15–17]. Although *in silico* models are very useful, their potential is limited, and so *in vitro* studies transfecting cadherin cell lines were carried to understand the functional results of the protein structure alterations [18].

To quantify E-cadherin's expression some techniques were used in combination with the *in vitro* assays. One of this techniques is Flow Cytometry (FC). In 1994, Tang et all used FC in combination with specific antibodies to carry a surface analysis and prove that human melanocytes, in culture environment, express E-cadherin with the same characteristics as reported for other cell types [19]. Furthermore, FC has been used to study E-cadherin influence on the cell surface localization of the promigratory 5T4 antigen, oncofetal antigen that correlates with poorer clinical outcome in colorectal, gastric, and ovarian carcinomas [20]. Likewise, Ferreira et all used FC to measure the percentage of apoptotic cells testing the hypothesis that E-cadherin deregulation could induce cell resistance to apoptosis [21]. Similarly, to discover the molecular mechanisms underlying the Chemical Chaperones effects in E-cadherin regulation, FC technique was used to assess E-cadherin in cell surface [22]. The major limitation of FC is the loss of morphology [23].

An alternative to circumvent this limitation is Image Stream, a tool for multiparameter cell analysis that combines the features of FC and Fluorescense microscopy (FM) with modern methodology for image analysis: allowing the analysis of a large number of cells based on their fluorescence features, providing a statistical analysis of these features and detailed morphometric cellular analysis, integrating morphometric and photometric features of the examined cells [24, 25]. The main limitation of this method is that it was designed to analyse suspending cells in a stream of fluid, preventing the assessment of cells in culture not measuring intercellular surfaces.

In this work FM images collected from *in vitro* assays were used to characterize the E-cadherin distribution within the cell in order to perceive and quantify E-cadherin expression differences between Wild Type (WT) images and seven different CDH1 germline missense variants HDGC related. FM appears in the beginning of the 20^{th} century (1911) when the first working fluorescence microscope was developed by Oskar Heimstädt to study autofluorescence in organic and inorganic compounds [26]. During the past one hundred years there were several breakthroughs in this area, namely: the developement of the 'secondary fluorescence' technique (staining with fluorescent dyes) in 1930s by Max Haitingen; the discovery of antibodies' labeling, in 1940s by Albert Coons, that allowed to specifically label proteins and subcellular structures [27]; the development of Green fluorescence protein (GFP) as a fluorescent label in 1990s by Tsien, Chalfie, and Shimomura [28–30].

In the FM images collected the cell membranes have conformations with a large geometric variability. When dealing with geometric abnormalities, Geometric Compensation (GC) is a useful practice. Geometric compensation is a common procedure in several image modalities, mainly for registration purposes [31]. The main goal is the comparison and alignment of objects with a wide range of shapes and sizes. The general strategy in this type of algorithms consists in the estimation of a geometric transformation, rigid or non-rigid [32], by optimizing a metric of similarity in order to make the objects

under alignment as similar as possible from shape and size points of view [33]. GC in images of microscopy is commonly used for segmentation [34] and tracking [35] puposes.

1.3 Original Contributions

This dissertation contributions begin with the characterization and quantification of WT E-cadherin and HDGC related missense mutant forms, distribution at intra- and inter-cellular levels, namely at the cell membrane. This information reflect the functional activity of this particular protein, leading to a better understanding of the functional results of some structure alterations.

In addition, a Graphical user interface (GUI) was implemented in Visual Basic/C^{++}, where a user can manually select in an image relevant points of study. This interface allow a specialist to select the representative cells of an image excluding from the analysis all negative cells, that worsen the results. This GUI is useful in bioimaging.

Another original aspect of this work is the development of internuclear and radial profiles, anchored in nuclei centers, to characterize molecules distribution at intra- and inter-cellular levels.

Finally, a registration method, a Geometric Compensation (GC) algorithm, was created to correct geometric heterogeneities in cells shape and size. This method models an image columns as a finite dimension continuous field, estimated from the initial observations, and imposing similarity between adjacent columns, allow the adjustment of each column observations, smoothing the solution. GC in quantification of biological features is not commonly described in the literature.

A full description of the images processing is provided. Finally, an analysis of each E-cadherin form phenotype is presented as well as a statistical analysis of some selected features to perceive the quantitative differences between E-cadherin WT phenotype and HDGC related remaining mutations phenotype.

This work resulted in an:

- Poster presentation with the title *E-cadherin spatial characterization with radial distribution profile for mutation detection.*,18^{th} edition of the Portuguese Conference on Pattern Recognition, Coimbra, October 2012.

- Oral communication with the title *E-cadherin radial distribution characterization for mutation detection purposes*, 6^{th} Iberian Conference on Pattern Recognition and Image Analysis , Madeira , June 2013.

- Sofia Esménio, J.Miguel Sanches and Raquel Seruca. Leitura digital da expressão proteica nas interfaces inter- intracellulares. Pedido de patente, modelo de utilidade ou de topografia de produtos semicondutores n^o 106982.

- Sofia Esménio, Joana Figueiredo, Raquel Seruca and J.Miguel Sanches, Geometrical compensation for protein expression profiling in fluorescence microscopy, Transactions on image processing (submitted).

1.4 Thesis Outline

This dissertation starts with the description of the concepts with biological background on Chapter 2. Initially, an explanation of E-cadherin's structure, role, tumor influence and clinical implications is provided. Moreover, a subsection is dedicated to FM, approaching subjects as: FM characteristics, immunofluorescence and limitations. Finally, both the specimen used in the *in vitro* assays developed for this project and the assays methodology are presented.

In Chapter 3, the preprocessing procedure and GC algorithm are described thoroughly. First, the several steps of the preprocessing procedure are described. Initially, a user manual to the GUI implemented is provided. Moreover, the segmentation and centroid calculation steps are further detailed. Finally, both processes of profiles collection (Radial (RD) and Internuclear (IN)) are explained. Results of both profiles approaches are shown. The Chapter ends with the mathematical explanation of the GC algorithm. Results of synthetic data and real data processed with this algorithm are displayed.

After that, in Chapter 4 , the features collected in the Internuclear Maps (INMs) and Radial Maps (RDMs) characterization process as well as the methods used in the statistical analysis are further detailed. First, the features collected for either the Mean Profiles and the Standard Deviation (STD) profiles are explained. Moreover, the statistical test implemented, Mann-Whitney U test, to test the discriminative potential of each of the features collected on the mean profiles, is described. Finally, the k-means algorithm which is used to help clustering the results from the feature extraction is explained.

The characterization of the WT and mutant forms of E-cadherin profiles is presented in further detail in Chapter 5. This Chapter is divided in an IN and RD analysis. Initially the geometrically compensated profiles are displayed and discussed. After that, the features selected to characterize the profiles, e.g mean intensity of the image, Maximum Intensity (MI), position of the MI, are extracted and the results are displayed and analyzed. Finally, a statistical analysis based on the Mann-Whitney U test, is presented.

In Chapter 6, thesis conclusions and future work are presented.

2

Biological Background

Contents

2.1 E-cadherin - Structure, Role and Tumor influence

E-cadherin (epithelial cadherin) was first identified by Takeichi in 1977 as a surface Ca^{2+} dependent protein with cell-cell adhesion potential [36]. E-cadherin is the major component of Adherens Junctions (AJ), intercellular junctions crucial for epithelial adhesion and barrier function in a wide variety of tissues and organisms [37]. The role of E-cadherin in tumor development is now well described. Experimental evidence supports a complex role both in suppressing invasion and metastasis formation. [17, 36].

2.1.1 E-cadherin - Gene, Structure and Role

E-cadherin belong to the subfamily of classical cadherins also known as type I cadherins. Cadherins are a superfamily of adhesion molecules which is mainly composed by three different kinds: 1) classical cadherins; 2) non-classical cadherins and 3) protocadherins. Classic Cadherins are Ca^{2+} dependent, homophilic, cell-adhesion molecules, often associated with various forms of AJ [2, 38].

Once mature, E-cadherin protein structure can be organized in three major structural domains: a cytoplasmic domain of about 150 amino acid residues (AA), a single transmembrane domain and an extracellular domain of about 550 AA, comprising five tandemly repeated domains exclusive to cadherins, the EC1-EC5 domains. The EC1 is the main responsible for the adhesive properties [38] Both the gene and an encoded E-cadherin protein are displayed in Fig. 2.1.

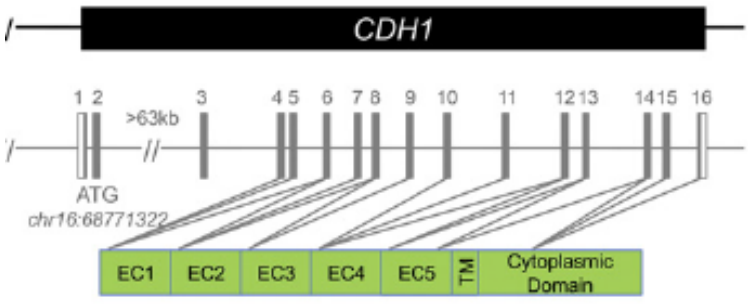


Figure 2.1: CDH1 gene structure and the encoded E-cadherin protein. E-cadherin molecules are encoded by the CDH1 gene, noted to the human chromosome 16q22.1. Adapted from [1].

E-cadherin has a major role in the formation, maintenance and homeostasis of epithelia. Its continued expression and functional activity are required for cells to remain tightly associated in the epithelium. In the absence of E-cadherin, the other adhesion cells and junction proteins are unable to support intercellular adhesion [39]. The epithelial adhesion happens through homophilic interactions between adjacent cadherins, first among adjacent cells (trans-interaction) and then within the same cell by lateral association (cis-interaction), leading to the formation of zipper-like structures [36].

To exhibit functional adhesion activity, cadherins must form complexes with cytoplasmic plaque proteins, called catenins, and with the actin cytoskeleton. The cytoplasmic domain of E-cadherin interacts with $\beta-$, $\alpha-$, and $\gamma-$ (plakoglobin) catenins (βctn, αctn, γctn), establishing the cadherin-

catenin complex. [36] The stability of the cadherin-catenin complex, and its linkage to actin filaments (through αctn), forms the core of the AJ (see Fig. 2.2 b) [39] . The AJs are crucial for the initiation and maintenance of a homeostatic intercellular space and cell to cell interaction, inhibiting individual epithelial cell motility and providing homeostatic tissue architecture in a wide variety of tissues and cell populations. E-cadherin is associated with the zonula adherens of the epithelial junctional complex, that help cells form a tight, polarized cell layer that can perform barrier and transport functions [2] (see Fig. 2.2a).

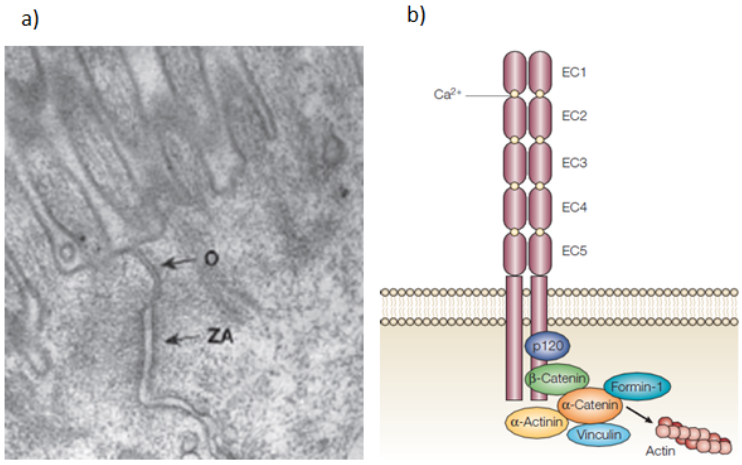


Figure 2.2: The adherens junction and the classic cadherin-catenin complex. a) Electron micrograph of a zonula adherens (ZA) junction of epithelia, where (O) represent a tight junction. b)The cadherin-catenin protein complex. Adapted from [2]

This capacity to maintain the overall state of adhesion between epithelial cells is what makes E-cadherin act as an important suppressor of epithelial tumor cell invasiveness and metastasis. Loss of E-cadherin expression and function leads to an enhance in cell invasiveness. E-cadherin deficiencies or mutations correlate with the invasiveness and metastasis of certain human tumors. [17, 36, 39]

2.1.2 E-cadherin - Influence in tumor progression and Clinical implications

Alterations in the adhesion properties between cells endow them with an invasive and migrator phenotype. E-cadherin plays an important role as an invasion suppressor protein, since its loss of expression, abnormal function, or both, leads to an increased ability of cells to invade neighboring tissues. Changes in E-cadherin expression have been implicated in key steps of tumour progression, including detachment of tumor cells from the primary site, intravasation into the blood stream, extravasation into distant target organs, and formation of secondary lesions or metastasis. [17].

Surveillance to CDH1 carriers is controversial due to the inadequacy of current screening modalities, since the lack of sensitivity for HDGC makes early diagnosis extremely challenging. HDGC genetic testing for CDH1 mutations is recommended to: 1) families with ≥ 2 cases of DGC and at least 1 case diagnosed before age 50 years; 2) families with ≥ 3 cases of DGC at any age; 1) isolated individuals diagnosed with DGC before age 35 years; 4) isolated individuals with both DGC and lobular breast cancer; 5) families with 1 member with DGC and another with either lobular breast cancer or signet ring cell colon cancer [5].

To circumvent current screening modalities limitations, IGCLC recommend that prophylactic gastrectomy should be offered to asymptomatic carriers of CDH1 truncating germline mutations. Indeed, the identification of small foci of signet-ring cell gastric carcinoma, early DGCs, in prophylactic gastrectomy specimens, from asymptomatic mutation carriers not yet displaying endoscopic evidence of disease, raises concerns regarding the efficacy of the current surveillance protocols [4, 16]. In this case the prophylactic gastrectomy recommended is a total gastrectomy as the signet ring cell cancer in HDGC are multifocal and distributed throughout the entire stomach. This procedure has significant risks namely: an estimated overall mortality is $2\% - 4\%$; a 100% risk of long-term morbidity including diarrhea, dumping, weight loss and difficulty eating [5].

Although its life saving potential this prophylactic gastrectomy procedure has been performed on several patients who had no reported evidence of GC on pathology being unclear [5]. This highlights the importance of genetic screening for identification of at risk individuals. Despite its limitations, endoscopic surveillance is still recommended for mutation carriers younger than 20 years old or for those who decline or delay prophylactic surgery, and should be carried out annually [4]. In hereditary forms of GC, carriers of CDH1 germline missense mutations represent a major burden in terms of genetic counseling and clinical management, and thus, there have been increasing efforts to predict the pathogenic significance of CDH1 germline missense variants. [16]

2.2 Fluorescence microscopy

Specificity, high sensitivity and versatility are the main contributions that FM can provide to the study of fixed and living cells. This efficient approach has an inherently greater sensibility and range than methods based upon changes in optical density or chemiluminescent emission. One of the ongoing contributions of fluorescence microscopy is to allow the study of single molecules enabling the understanding and characterization of both static and dynamic cellular processes [40].

2.2.1 Fluorescence Characteristics

Fluorescence as a phenomenon is part of a larger family of related luminescent processes in which a susceptible substance absorbs light, only to reemit light (photons) from electronically excited states after a given time. Fluorescence outcome is the emittance of a photon with a longer wavelength. When light of a particular wavelength hits a fluorescent sample, the atoms, ions or molecules therein absorb a specific quantum of light, which pushes a valence electron from the ground state. The energy of photons involved in fluorescence and generally a quantum of light can be expressed via Planck's law:

$$E = h.\nu = h.\frac{c}{\lambda}$$

where E is the quantum's energy (J), h is Planck's constant $(J.s)$, ν the frequency (s^{-1}), λ is the wavelength of the photon (m), and c is the speed of light in a vacuum. $(m.s^{-1})$. [41]

Fluorescence has several characteristics, some of which should be compensated to obtain better results. The essential characteristics of fluorescence are: quantum yield, fluorescence lifetime, anisotropy, fluorescence quenching, fluorescence intermittency, resonance energy transfer and charge-transfer complexes.

Quantum yield (brightness of fluorochrome's emission): is given by the ratio of the number of emitted to absorbed photons, which is determined by the rate constants of emission (Γ) and the sum of all non-radiative decay processes (k_{nr}) that depopulate the excited state. [41]

Fluorescence lifetime (fluorochrome's fluorescence lifetime (τ)): is the average time an electron spends in the excited state before returning to ground state. The initial fluorescence intensity, I_0, suffer an exponential decay, I_t, over time, t, that can be expressed as: $I_t = I_0 e^{(\frac{-t}{\tau})}$. This decay through radiative, Γ, and non-radiative processes, k_{nr}, time is the fluorescence lifetime, $\tau = \frac{1}{\Gamma + k_{nr}}$ [41].

Anisotropy (different properties along different axes): in a pool of randomly oriented fluorochromes, only those fluorochromes with transition dipole moments that are aligned parallel to the polarization direction of the excitation beam, linearly polarized, will be excited (photoselection) [41].

Fluorescence quenching: is the phenomenon in which the interaction of the fluorochrome with a molecule, the quencher, reduces the quantum yield or the lifetime [41].

Fluorescence intermittency: when the fluorochrome randomly alternates between a fluorescent ("on") and dark state ("off") despite continuous excitation illumination. The random nature and power law dynamics generally frustrates and precludes comparison of results between independent experiments. [41]

Resonance energy transfer: is a photophysical process in which the excited state energy from a donor fluorochrome is transferred via a non-radiative mechanism to a ground state acceptor chromophore via weak long-range dipole-dipole coupling. [41]

Charge-transfer complexes: are nanosecond short-lived homodimers (excimer) or heterodimers (exciplex) of two molecules of which at least one is in the excited state that show redshifted emission compared with the monomer's emission. [41]

2.2.2 Imunofluorescence

Immunofluorescence (IF) is widely used in both scientific research and clinical laboratories. This technique utilizes antibodies chemically conjugated to fluorescent dyes to detect specific target antigens. This labeled antibodies bind either directly (direct immunofluorescence) or indirectly (indirect immunofluorescence) to the antigen of interest, resulting in antigen fluorescence detection. In direct immunofluorescence the antibody chemically conjugated with the fluorescent dye bind directly to the antigen of interest. On the other hand, indirect immunofluorescence is a two-step technique in which a primary unlabeled antibody binds to the molecule of interest and is afterwards detected by a fluorophore-labeled second antibody. The indirect immunofluorescence technique is more complicated and time consuming because it involves another antibody and so it requires a second incubation period. However, it is more sensitive as more than one secondary antibody can bind to each primary

antibody, which amplifies the fluorescence signal [42].

2.2.3 Immunofluorescence Limitations

This section introduces some of immunofluorescence most common limitations, namely: autofluorescence, photobleaching and fluorescence overlap.

Autofluorescence (background fluorescence): fluorescence that does not originate from the Fluorochrome of interest (FOI) but rather from cellular components with fluorescent properties. Biological autofluorescence in mammalian cells can be problematic in the detection of fluorescence probes in tissues and cells [43].

Photobleaching: photochemical destruction of a fluorophore due to the generation of reactive oxygen species in the specimen as a byproduct of fluorescence excitation. It is thought that the primary causative mechanism appears to be photosensitization of singlet oxygen (O_2) generation by the dye triplet-excited state and reference. [43]

Fluorescence Overlap: occur when measuring fluorescence of more than one color. This signal overlaps must be electronically removed or each detector will overestimate the actual signal. [43]

2.3 In Vitro Assays

In this work *in vitro* assays were developed containing WT E-cadherin and E-cadherin missense mutants forms associated with HDGC. The E-cadherin mutant forms, derived from CDH1 germline missense mutations, presented in this work were: T340A (1018), A463V (1901), R749W (2245), E757K (2269), E781D (2343), P799R (2396), V832M (2494). In the following image (fig. 2.3) the relative position of the mutation and the nucleotide (DNA) involved for each germline mutation form is shown.

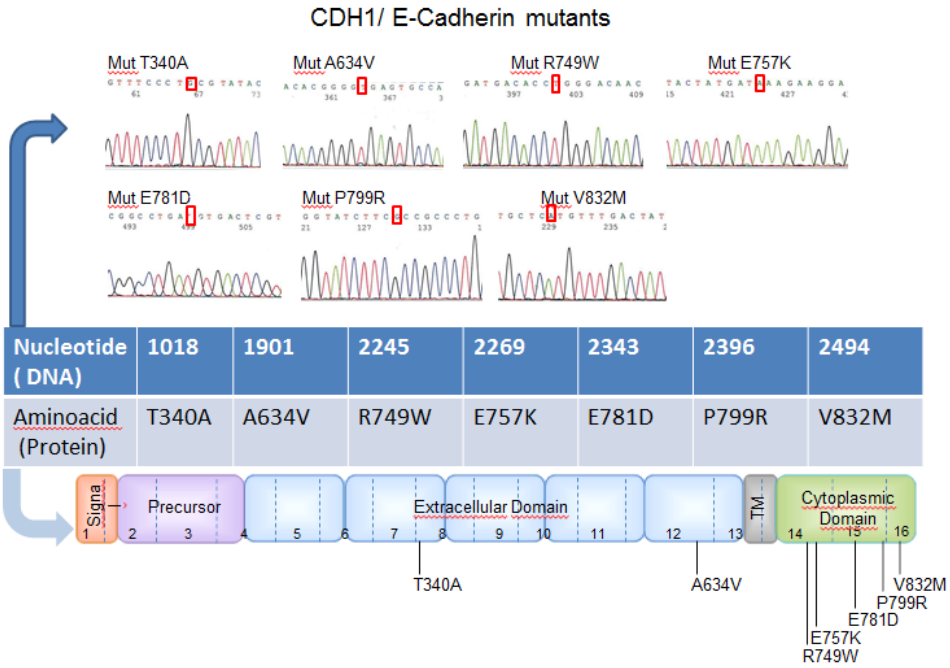


Figure 2.3: Germline missense CDH1 mutations and respective E-Cadherin mutant forms.

According with the site of mutation the functional impact is diferent. As can be seen in fig. 2.3, T340A and A643V are extracellullar mutations, R749W and E757K are juxta-membrane mutations and finally E781D, P799R and V832M are intracellular mutations. In some cases the functional result of the mutation is not well known. However in the juxta-membrane mutations its known that the mutation interferes with the conformation of the protein stopping the connection with $p-120$, compromising the stability of the protein in the membrane. Also in R749W the conformational changes activate the Endoplasmic Reticulum Associated Degradation (ERAD) mechanism. Finnally in the intracellular mutations, E781D, P799R,V832M, the mutation interfers with the connection with $\beta - catetin$, affecting the transport of the protein to the membrane. To study WT E-cadherin expresion and the functional impact of each one of these mutations in vitro assays were developed with WT E-cadherin and the previously mentioned mutant forms, derived from CDH1 germline missense mutations.

In this *in vitro* assays CHO cells transfected with vectors encoding the wild type E-cadherin or the mutant forms were seeded on 6-well plates on top of glass coverslips and grown to at least 80 % conguence. Fixation was performed in ice-cold methanol for 20 minutes, followed by washing and blocking in 5% Bovine serum albumin (BSA) and Phosphate Buffered Saline (PBS) for 30 minutes, at room temperature. The mouse monoclonal E-cadherin antibody (BD Biosciences) was used at 1:300 dilution in PBS with BSA 5% and incubated for 1 hour at room temperature. An Alexa Fluor 488 goat anti-mouse (1:500, Invitrogen) [44] was applied for 1 hour in dark as secondary antibody. The coverslips were mounted on slides using Vectashield with 4',6-diamidino-2-phenylindole (DAPI) (Vector Laboratories). Images were acquired on a Carl Zeiss Apotome Axiovert 200M Fluorescence Microscope, using 40x objectives. Images were taken with an Axiocam HRm camera and processed with the Zeiss Axion Vision 4.8 software.

The images obtained and consequently processed in this work were the following: 3 containing WT

E-cadherin (№ 37,38,39), 1 containing T340A (1018) E-cadherin (№ 41), 2 containing A463V (1901) E-cadherin (№ 45,48), 2 containing R749W (2245) E-cadherin (№ 49,50), 2 containing E757K (2269) E-cadherin (№ 57,59), 2 containing E781D (2343) E-cadherin (№ 63,64), 2 containing P799R (2396) E-cadherin (№ 65,67) and 1 containing V832M (2494) E-cadherin (№ 68).

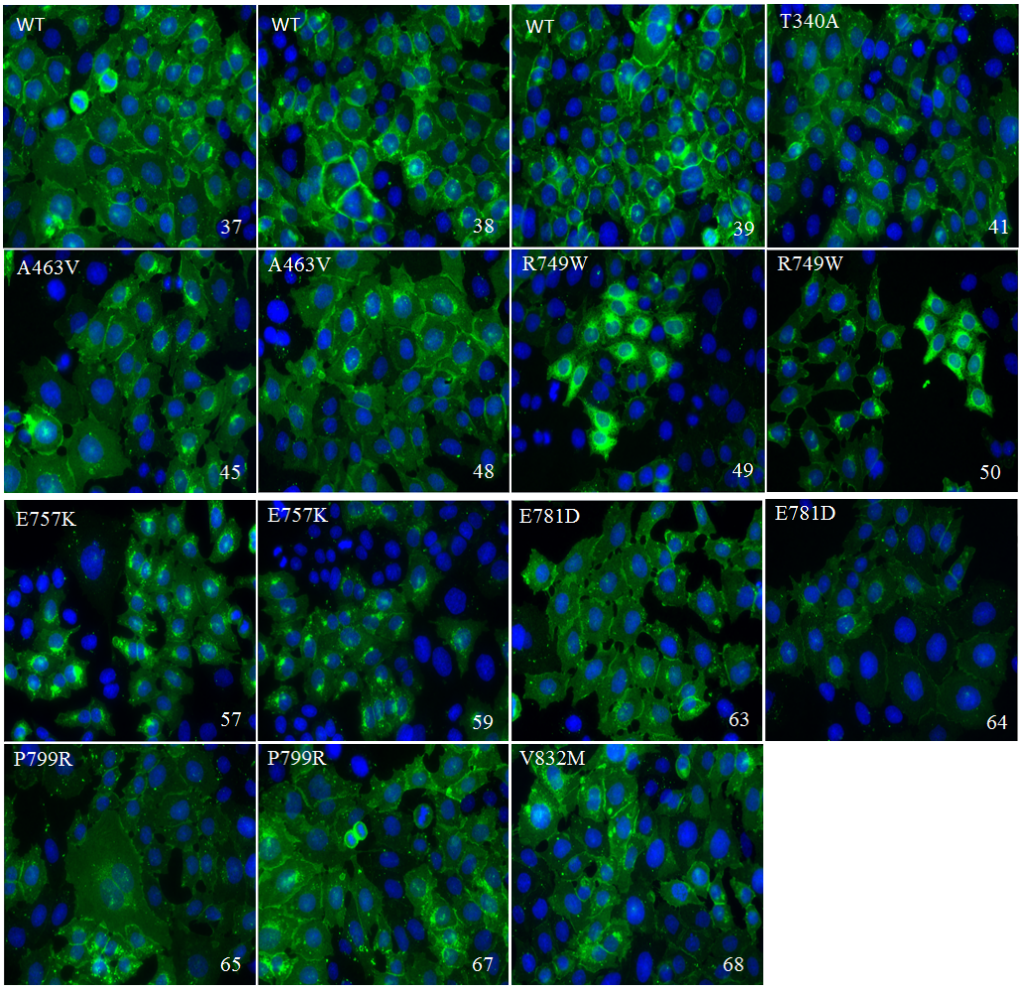


Figure 2.4: Images collected from the *in vitro* assays developed with CHO cells transfected with vectors encoding the wild type E-cadherin and diverse mutant forms.

3

Image processing and Geometric Compensation

Contents

3.1 Image processing

The preprocessing algorithm is composed by the following steps:

- **Nuclei Selection**, using a GUI (see section 3.1.1) to select the pairs of nuclei to be evaluated. This process is crucial for the success of the method because cells where the transfecting process did not perfectly occur should not be taken into account in profile definition.
- **Segmentation and centroid detection**, of the GUI selected nuclei coordinates, to ensure that the profiles collected are nucleus centered and independent of the GUI user accuracy (see section 3.1.2).
- **Internuclear and Radial Maps building**, inserting all the profiles of each kind (Internuclear or Radial) together in a respective length normalized map (see sections 3.1.4 and 3.1.5). One internuclear profile is created for each pairs of adjacent nuclei positively selected on *Nuclei Selection* and one radial nucleus centered profile is created for each nucleus selected (see section 3.1.3).
- **GC**,of the maps collected using 1D continuous column profiles to minimize the overall variability of each map in horizontal direction.

The geometric compensation of the maps collected from the image is the key step of the proposed method. Both maps contain geometrical abnormalities in the collected information either due to the process of collection or to cellular membrane shape and size heterogeneities. This method models these maps columns as a finite dimension continuous field estimated from the intensities of the profiles and allow the locations of each observation to adjusted, up and down, along the corresponding column under global smoothness and consistency criteria that avoid disruption. This algorithm intends to geometrically compensate the created maps so the information collected is rigorous and independent of cellular membrane geometry 3.2).

3.1.1 Graphical user interface (GUI)

In this work a Graphical user interface (GUI) was developed to allow an expert to manually select in each FM image the nuclei to evaluate. The biologist selects the cells in each image that are representative of the common protein status within that in vitro cell culture. This semi-automated approach allow the expert to exclude all negative cells from the analysis, that may represent technical pitfalls of immunofluorescence or protein degradation. This interface was developed to select pairs of adjacent nuclei. However, the interface stores the information (pair of coordinates) in a single, orderly in time, vector. Thus, this interface can be used in single nuclear selection. The GUI received the named of IDnuclei and was designed using *Visual Studio* (C^{++} language), with the CImg Library [45]. Follows the description of the GUI.

The initial Menu is shown in Fig.3.1 a). Initially an image should be loaded to the program, pressing *Load Image* and selecting the directory of the image. If correctly loaded the photo will appear in the initial menu and the user can inspect if it is the correct one. (see Fig.3.1 b))

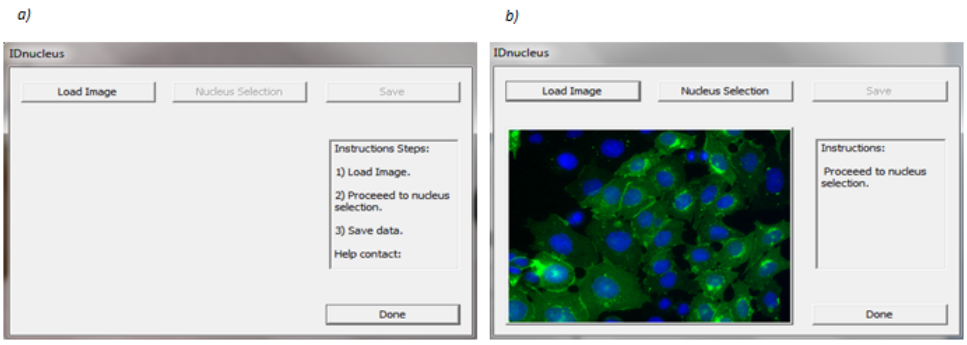


Figure 3.1: GUI a) Initial Menu, b) Menu After loading Image

After loading the image the user can decide to proceed to the nucleus selection proceeding (pressing *Nucleus Selection*) or in case it was not the correct image change it repeating the previous steps (pressing *Load Image* and selecting the directory). The *Nucleus Selection* option opens the image selected in full screen where the expert can now proceed to the selection of the pairs of nuclei desired. (example in Fig.3.2).

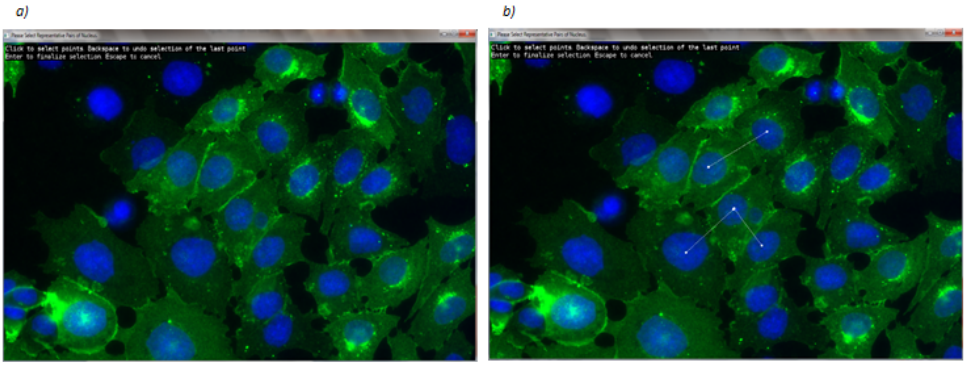


Figure 3.2: Selection of Nuclei Menu. a) Initial. b) After some selections

In the *Nucleus Selection* menu or full screen mode there are a few possible actions. The Left button of the mouse confirms the position of a nucleus saving its coordinates in the image. The Backspace button erases the last selection done. The Escape button aborts the entire selection. The Enter button finishes the selection properly. After finishing the selection the option to save data (Fig.3.3) appears allowing the data to be saved. The data will be saved in a folder with the name of the image analyzed in the directory of the program. The folder contains the initial image and a .txt document with all the coordinates (x,y) of the nuclei selected.

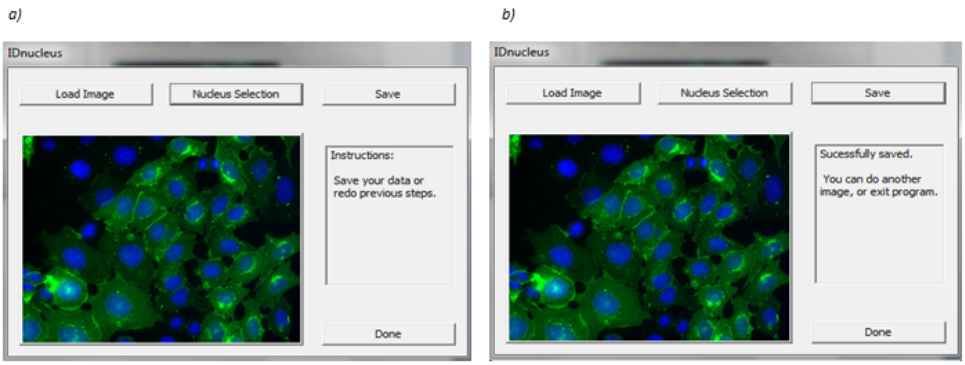


Figure 3.3: GUI a) After selection Menu, b) Final Menu

After saving the data, the user can close the program or proceed to another selection repeating all the steps from the begining.

3.1.2 Segmentation and Centroid Detection

Matlab®software uses the most commonly used color representation based on the classical three-color theory of Thomas Young, the RGB system. This means that a color image is represented as a 3D matrix, wherein each 1D matrix contain the information to one specific color. In this work, the color images processed contained E-cadherin molecules immunostained with a green fluorescent dye and cell nuclei tagged with blue fluorescent proteins. Thus, not only the information collected by the red channel was discarded but also the other two matrices were converted in gray-scale and processed individually. The blue matrix contains the information concerning cell nuclei and the green matrix the information concerning the E-cadherin molecules distribution.

As membrane segmentation may be difficult, inaccurate, time consuming, and not essential for intensity distribution characterization purposes, the profiles collected to study E-cadherin inter- and intra-membrane distribution were anchored in the geometrical centers of the cell nuclei. The cells and pairs of cells, selected by the operator as being representative of the common protein status within that in vitro cell culture, coordinates were used as nuclei reference points. A correction step was added to ensure that the results are independent of the GUI user accuracy and current conditions. The correction consisted on a segmentation step combined with a centroid detection.

In segmentation, Otsu thresholding method, described in section 3.1.2.A(see fig. 3.4 a)) was combined with a morphological procedure, the watershed algorithm (described in section 3.1.2.B). As Otsu thresholding alone is unable to distinguish contiguous nuclei and watershed in FM images leads to a lot of over segmentation due to the uneven intensity, owing to auto-fluorescence from the tissue and fluorescence from out of focus objects, the methods were implemented combined to improve segmentation (see fig. 3.4 a, b)) [46]. To ensure the separation of contiguous nuclei the segmentation result was eroded [1] increasing the inter space between nuclei.

The centroid detection was implemented using the *Matlab@* algorithm *Regionprops*. This algorithm measure several properties of binary images being one of the properties the center of the mass

[1] Binary erosion is the process of eliminating all the boundary points from an object, leaving it smaller in area by one pixel all around the perimeter

of a region (see fig. 3.4 d)). To convert watershed results in a binary image and simplify the centroid detection the Canny edge detector algorithm, described in section 3.1.2.C, was applied (see fig. 3.4 c)).

In few cases even non-spherical nuclei can be decomposed in half. To compensate in the end the algorithm study the existence of another centroid detected and calculate the mean position of them. An example of this is shown in fig. 3.4 d e e).

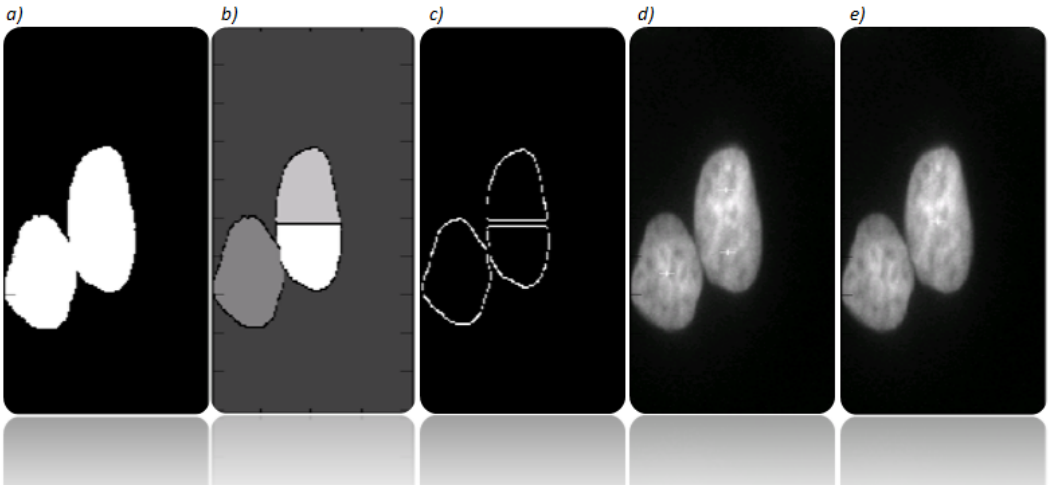


Figure 3.4: Segmentation and centroid detection process of a ROI. a) Otsu's thresholding, b) Watershed algorithm analysis, c) Erosion and Canny edge detection d) *Regionprops* algorithm Centroid detection e) Final ROI .

3.1.2.A Otsu's thresholding

Intensity thresholding was one of the first approaches to cell segmentation. And is still one of the most common approaches in cell segmentation, when combined with other methods. [47]. Among the numerous thresholding methods for nucleus segmentation Otsu's segmentation is still widely popular due to its simplicity and effectiveness in cases with good contrast.[48].

Otsu method is a nonparametric and unsupervised automatically threshold selection method. The algorithm assumes that the image to be thresholded contains two classes of pixels: background pixels, (Class 0) and object's pixels, (Class 1) and the objective is to calculate the optimum threshold minimizing the weighted sum of within-class variances of the two classes of the two classes of pixels that compose the image. [49] The within class variance, the between class variance and the total variance are:

$$\sigma_W^2 = \omega_0\sigma_0^2 + \omega_1\sigma_1^2$$

where $w_0 = \sum_{i=1}^{k} p_i = w(k)$ and $w_1 = \sum_{i=k+1}^{L} p_i = 1 - w(k)$ are the probability of class ocurrence, $\sigma_0^2 = \sum_{i=1}^{k}(i-\mu_0)^2 \frac{p_i}{\omega_0}$ and $\sigma_1^2 = \sum_{i=k+1}^{L}(i-\mu_1)^2 \frac{p_i i}{\omega_1}$ the variances of each class, and $\mu_0 = \sum_{i=1}^{k} \frac{ip_i}{\omega_0}$ and $\mu_1 = \sum_{i=k+1}^{L} \frac{ip_i}{\omega_1}$ the mean levels of each class. Otsu shows that minimizing the intra-class variance is the same as maximizing inter-class variance:

$$\sigma_B^2 = \omega_0\omega_1(\mu_1 - \mu_0)^2 = \frac{[\mu_T\omega(k) - \mu(k)]^2}{\omega(k)[1-\omega(k)]}$$

The optimal threshold to separate C_0 and C_1 is k^* is $\sigma_B^2(k^*) = \max_{1\leq k<L} \sigma_B^2(k)$. [50]

3.1.2.B Watershed Algorithm

The watershed transform is a gray scale mathematical morphology method, originally proposed by Digabel and Lantuéjoul [51] and later improved by Beucher and Lantuéjoul [52]. This binary morphological method is used to separate slightly touching objects that result from the segmentation process [3].

Rather than simply thresholding the image at an optimum gray level, the watershed approach begins with a threshold that isolate the individual objects properly, and gradually raises it to the optimum level. The process begins with a low threshold that segments the image into the proper number of objects. Afterwards, the threshold is raised gradually, one gray level at a time, expanding the object boundaries. The restraining of no merge between objects transform the points of first contact into boundaries. The process is terminated before the threshold reaches the gray level of the background [3]. The watershed algorithm is briefly shown in Fig. 3.5

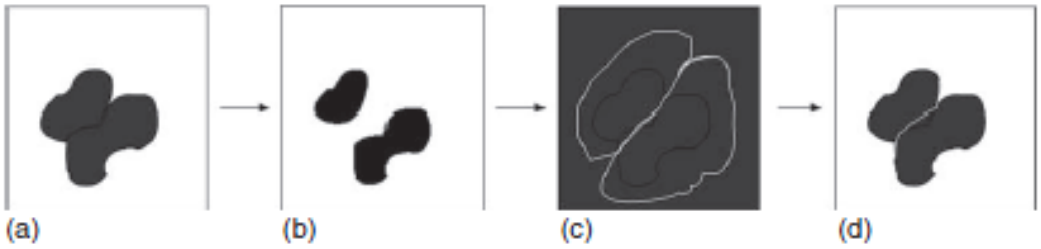


Figure 3.5: Illustration of separating touching objects. [3] (a) A binary segmented image. (b) After a few erosions and inversion. (c) The exoskeleton. (d) Separated objects

This morphological processing technique for image segmentation is still a quite popular method of growing region [47, 48].

3.1.2.C Canny edge detector

In this work the Canny edge detector was used to convert the Watershed segmentation results in a binary image, stabilizing the *Regionprops* algorithm. Edge detection consist on evidencing pixels in an image at which the gray level changes sharply taking into account the slope and direction of the transition. This operators perform a 2-D spatial gradient measurement using convolution with a pair of horizontal and vertical derivative kernels, g_x and g_y. Each pixel in the image I(x, y) is convolved with both kernels, one estimating the gradient in the x direction and the other in the y direction. The output of these convolutions is combined forming the estimated absolute magnitude of the gradient $|G|$ and its orientation θ at each pixel. This gradient magnitude is:

$$|G| = \sqrt{G_x^2 + G_y^2}$$

where Gx and Gy are the output of the estimated derivative function in the x and y directions, respectively, $G_x = I(x, y) * g_x$ and $G_y = I(x, y) * g_y$. Also the gradient drecton can be computed as $\theta = arctan\frac{G_y}{G_x}$.

Canny edge detector is peculiar as its operator is actually a multistage-edge detection algorithm. First, as this detector is susceptible to noise in raw unprocessed data, the image is smoothed through a convolution with a Gaussian kernel. Afterwards, a first-derivative operator (usually Sobel) is applied to obtain the spatial gradient measurements. In this process the pixels with gradient magnitudes that form local maxima in the gradient direction are determined and the local gradient maxima produces ridges in the edge map . A *nonmaximum suppression* step tracks along the top of these ridges using a dual threshold mechanism, that starts on a ridge higher than the upper threshold and proceed out from that point in both directions until a point on the ridge falls bellow the lower threshold. The underlying assumption is that important edges are along continuous paths in the image. This allow to discard noisy pixels that don't form paths. Finnaly every point who is not along the tracked ridges is set to zero and the result is a binary image where each pixels is labeled either as an edge point or a non edge point. [3]

3.1.3 Profiles collection

The motivation of this work is to qualify and quantificaty the distribution of E-cadherin molecules within cells. As explained in Chapter 2, conformacional changes in E-cadherin due to mutations result in alterations in E-cadherin distribution. In cell, WT E-Cadherin molecules are mainly concentrated at the cell membrane, where E-cadherin performs its important role in the adhesion complex and in the physical linkage between cells. In the wild-type context and in homeostatic situations, the level of E-cadherin expression at the cytoplasm is low and uniformly distributed and this is due to its normal recycling. To characterize the distribution of E-cadherin expression two types of profiles were developed: Radial Profiles (RDPs), anchored at the nucleus geometrical center of each individual cell collected from different angles, and Internuclear Profiles (INPs), corresponding to sets of parallel intensity lines extracted between pairs of neighbouring cell. The method and the purpose of collection of each profile will be described in the next Sections 3.1.4 and 3.1.5.

3.1.4 Internuclear profiles

The Internuclear Profiles (INPs) were developed to study the existence of a well-defined, equidistant, membrane between pairs of adjacent nuclei positively selected on Nuclei Selection. The equidistant of the membrane together with an increase in intensity values are traits of a healthy linkage between cells which mean presence of WT E-cadherin. These profiles are collected concerning the center coordinates of the selected pairs of adjacent nuclei, and consist of a set of parallel intensity lines extracted from the original image of the region between the selected pairs of nuclei (see Fig. 3.6). Each profile created, is inserted in an Internuclear Map (INM) created to include every profile withdrawn from the same image (see Fig.3.7). The collection method is explained in the next section.

3.1.4.A Methodology

These profiles are collected concerning the estimated centroid coordinates of the selected pairs of adjacent *nuclei*. First the distance between the nuclei that compose the pair, ρ , and the orientation

angle, θ ($\vec{k}$ direction) are calculated (see Fig. 3.6 a)). Thus, the INP dimension is $\mathrm{A} \times \rho$, being A a normalization value (A=100). Afterwards, two more orientation angles, perpendicular to θ, are computed: θ_{up} and θ_{down} (3.6 b)). These angles, as the two ways of $\vec{k1}$ direction, allow the collection of upper and lower parallel profiles. The profiles are collected starting on the line described by $\vec{k1}$ direction ($mathrm1^st$ *nucleus*) and and have a length of ρ. These profiles are collected in $\vec{k}$ parallel directions (3.6 c)). In Fig.3.6 c) the increase in darkness simulates the increase in the iteration step of withdrawn. The final INP is represented in Fig.3.6 d).

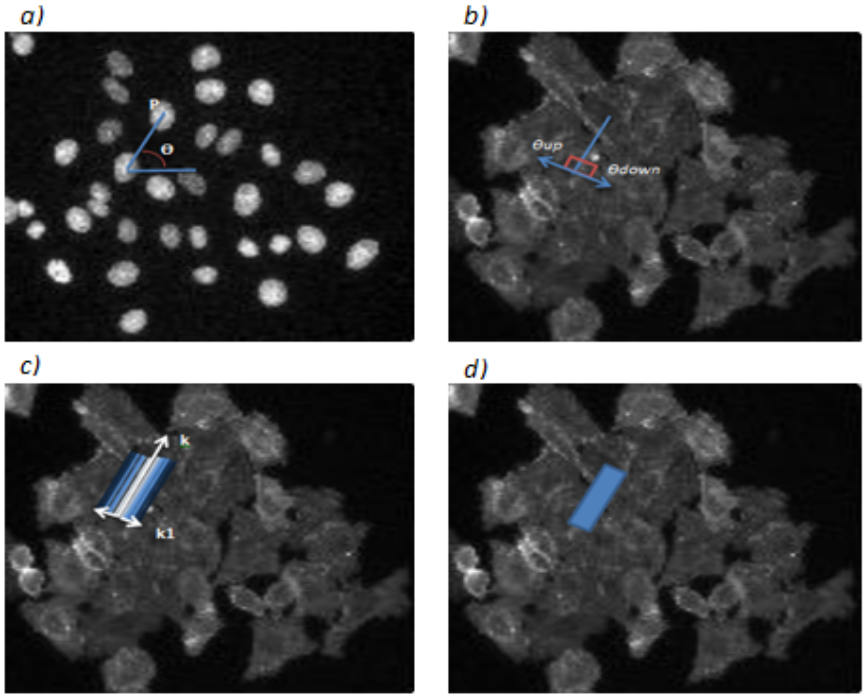


Figure 3.6: Internuclear Collection. a) Definition of θ direction andρ vector (nuclei centered) b) Definition of θ_{up} and θ_{down} directions c) Parallel Profiles collection d) Final INP

Finally, every profile is inserted in an Internuclear Map (INM) created to include every profile withdrawn from the same image. The profiles are inserted in order, occupying the next ρ columns of the map. The map dimensions will be $\mathrm{N} \times \mathrm{M}$, being $\mathrm{N} = \mathrm{A}$ and $\mathrm{M} = \mathrm{N^o}_{\mathrm{profiles}} \times \rho$. The insertion is shown in fig. 3.7.

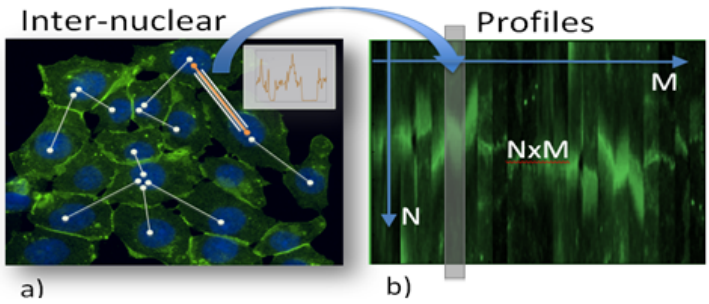


Figure 3.7: Internuclear Map. a) E-cadherin distribution Profiles b) Final INM.

3.1.4.B Initial Internuclear Maps (INMs)

In this work, several *in vitro* assays of WT E-cadherin and HDGC related mutant forms were developed (see section 2.3). Each assay developed resulted in one FM image, that has a corresponding INM comprising all the INPs collected. The INMs correspondents to each assay developed (see fig.2.4), are displayed on fig.3.8.

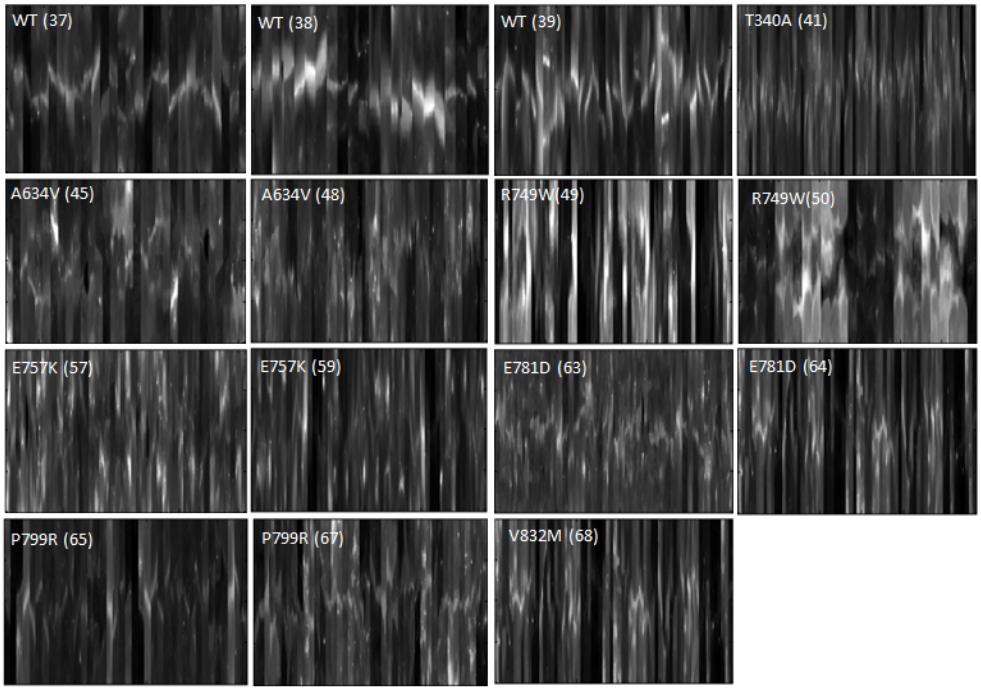


Figure 3.8: INMs and GEFMs collected from the *in vitro* assays developed with WT E-cadherin and HDGC related mutant forms.

To understand the intensity behavior between neighboring *nuclei* in each INM, the mean profile in horizontal direction was computed. To study behavior resemblances in INMs built from FM images containing identical E-cadherin forms, INMs regarding the same E-cadherin form were displayed together. The results are displayed in Fig.3.9 and Fig. 3.10.

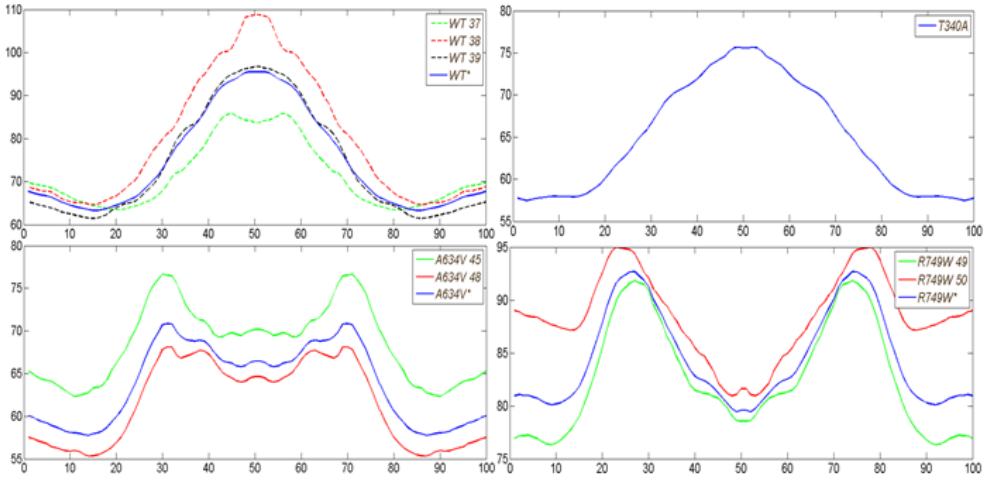


Figure 3.9: Comparison of the IN horizontal mean profiles concerning each form of E-cadherin, in this case WT, T340A, A634V, R749W.

WT, A634V, R749W INMs' mean profiles are similar in both intensity and shape among INMs concerning the same E-cadherin form. On the other hand, Fig. 3.10 shows that E757K, E781D and P799R INMs' mean profiles have some significant differences in either profile's shape and/or intensity. Finally, T340A and V832M have only one INMs so no resemblance study is possible.

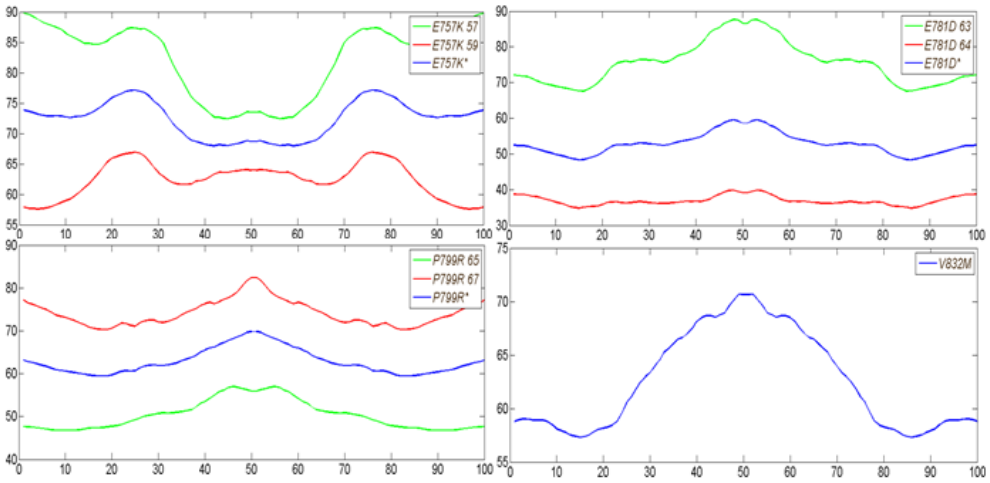


Figure 3.10: Comparison of the IN horizontal mean profiles concerning each form of E-cadherin, in this case E757K, E781D, P799R, V832M

To simplify the comparison between different forms of E-cadherin, before GC, Global E-cadherin Form Maps (GEFMs) were developed combining the INMs of the same form of E-cadherin. The mean and STD of this GEFMs are displayed in Fig. 3.11 and Fig.3.12.

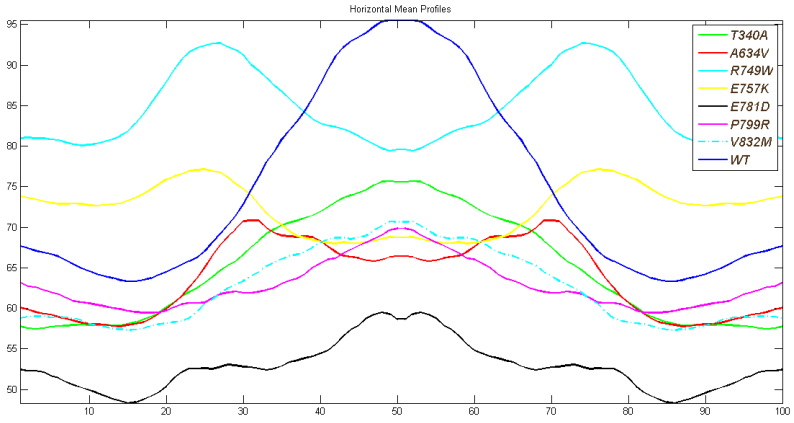


Figure 3.11: Comparison of the IN horizontal mean profiles of each form of E-cadherin map

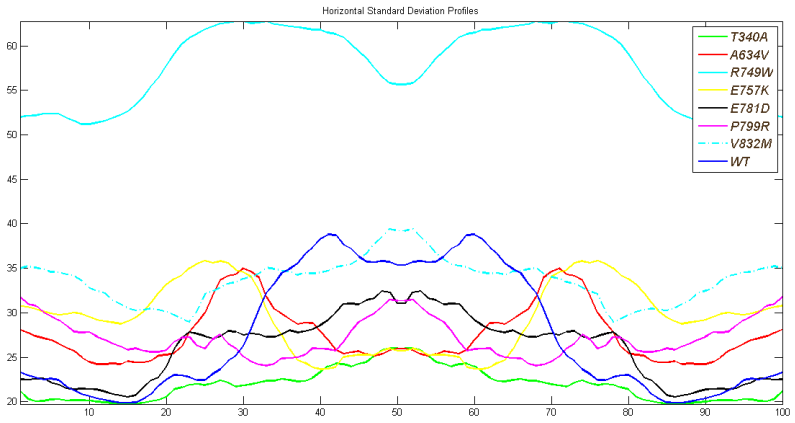


Figure 3.12: Comparison of the IN horizontal standard deviation profiles of each form of E-cadherin map

Regarding the average profiles of the GEFMs is possible to observe that: WT, T340A, V832M, P799R and E781D profiles have similar approximately Gaussian shape; WT has the overall higher maximum intensity; E781D has the lower overall maximum intensity; T340A, V832M, P799R profiles are quite similar both in shape has in overall intensity; on the contrary, R749W and E757K profiles are completely different of the others, as they display two high lateral intensity peaks. Regarding the STD of the GEFMs is possible to observe that WT, T340A, A634V, E757K, E781D, P799R and V832M STD are similar in intensity within a certain range. R749W GEFM display a much higher STD.

3.1.5 Radial profiles

The Radial Profiles (RDPs) store nucleus-to-membrane radial environment surrounding each nucleus, to study E-cadherin molecules distribution within the cell. In the wild-type context and in homeostatic situations, E-cadherin molecules are mainly concentrated at the membrane being the expression at the cytoplasm low and uniformly distributed, due to its normal recycling. The existence of abnormal values of E-cadherin being decomposed in the cytoplasm suggest E-cadherin with abnormal conformation and therefore a gene mutation. These profiles contain a set of intensity lines of the original image extracted from the radial surroundings relatively to the center coordinates of each of the selected *nuclei* rearranged in regards with the center of the cell (see fig. 3.13). The method of collection is explained in the next section.

3.1.5.A Methodology

Initially, a square Roi is collected from the E-cadherin distribution matrix centered in the center coordinates of the selected nuclei. The ROI size is $2\mathrm{N} \times 2\mathrm{N}$, being $\mathrm{N} = \mu_{\mathrm{dist}}$, where μ_{dist} is the mean of the distance vector obtained after the IN collection. (see section 3.1.4.A). This way its guaranteed that this ROI contain the information from nucleus-to-membrane in all directions. The information is rearranged in different radial coordinates, ρ ($\rho = \mathrm{N}$) and θ, with the referential in the center (see Fig. 3.13 a)). The final RDP dimension is $N \times \theta_n$, being θ_n equal to $\theta_n = \frac{2\times\pi}{\theta_{step}}$. This way the number of lines gathered depend on θ_{step}, which is the incremental step of θ. In this case θ_{step} was considered

$\frac{\pi}{180}$ (1°). The final RDM has dimension $\mathrm{N} \times \theta_{\mathrm{n}}(360)$. The intensity lines collection and rearrangement are displayed in Fig. 3.13.

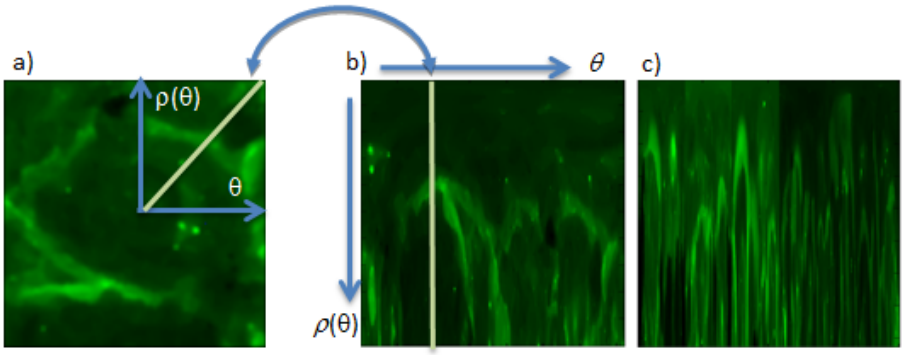


Figure 3.13: Radial profiles. a) E-cadherin distribution Nucleus centered ROI, b) Profile obtained from the polar collection of data, c) Final RDM.

In this process each line collected is then placed on the k^{th} column of the final RDP, being k both the number of the actual iteration and a reference to the θ value in which the line was gathered. This fact allow the reconstruction of the cell profile from the RD profile. This process is used to reconstruct the different cells comprising each form of E-cadherin present after maps GC. The process is shown in Fig. 3.14

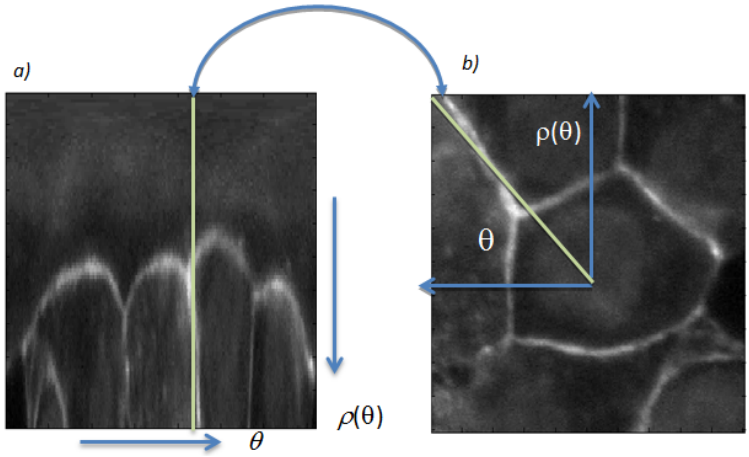


Figure 3.14: Radial profiles. a)Nucleus centered region of interest of E-cadherin distribution. b) Profile obtained from the polar collection of data.

Each profile, after being created, is inserted in a RDM that combines every profile withdrawn from the same image creating a RDM representative of the original image. The insertion method is similar with the one described in section 3.1.4.A. However, in this case the map dimensions will be $N = A$, being A the same normalization factor, and $M = N^{\underline{o}}{}_{profiles} \times \theta_n$.

3.1.5.B Initial Radial Maps (RDMs)

In this work, several in vitro assays of WT E-cadherin and HDGC related mutant forms were developed (see section 2.3). Each assay developed resulted in one FM image, that has a corresponding

RDM comprising all the RDPs collected. The RDMs obtained processing the several images collected from the in vitro assays developed (see Fig.2.4) are displayed on Fig.3.15.

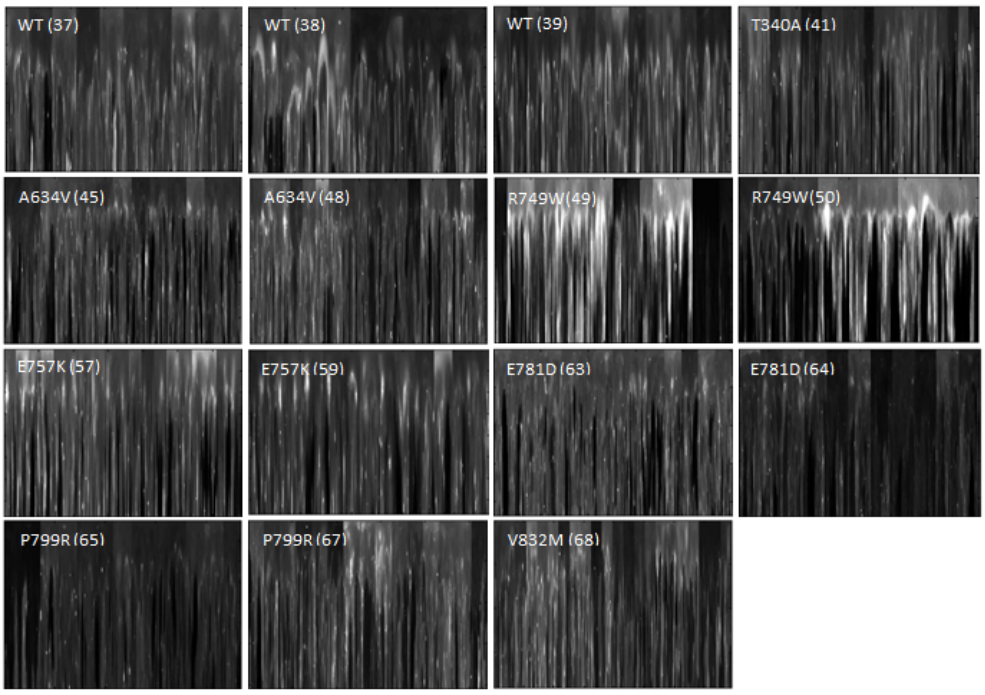


Figure 3.15: RDMs and GEFMs collected from the *in vitro* assays developed with WT E-cadherin and HDGC related mutant forms.

To understand the intensity behavior within the cell, RDMs' mean profile in lines direction were computed, where RDMs regarding the same E-cadherin form were displayed together, to study behavior resemblances in RDMs built from FM images containing identical E-cadherin forms. These profiles are displayed in Fig. 3.16 and Fig.3.17.

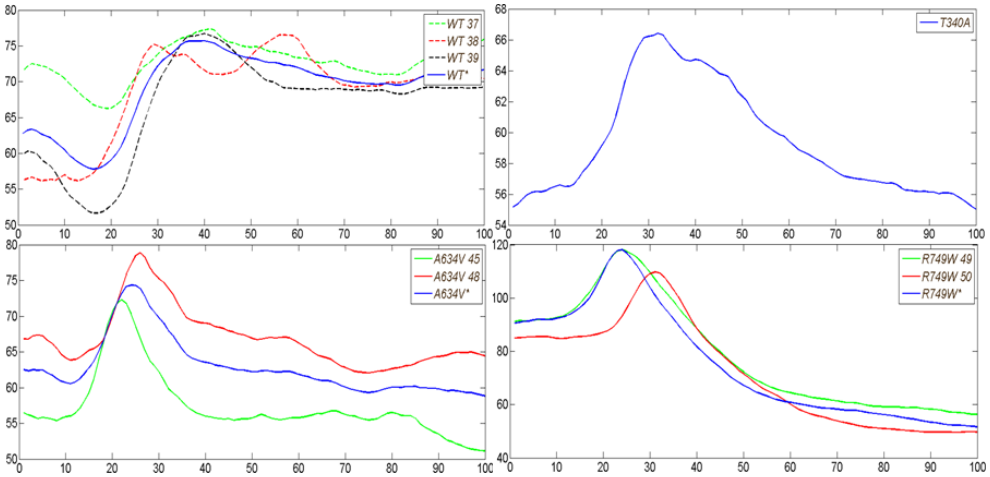


Figure 3.16: Comparison of the RD horizontal mean profiles concerning each form of E-cadherin, in this case WT, T340A, A634V, R749W.

With respect to WT, A634V, R749W, E757K forms of E-cadherin RDMs, mean profiles are similar in both intensity and shape, within a certain variance, among RDMs concerning the same E-cadherin form. On the other hand, E781D, P799R RDMs mean profiles have some significant differences

in either profiles shape and/or intensity. Finally, T340A and V832M have only one RDMs so no resemblance study is possible.

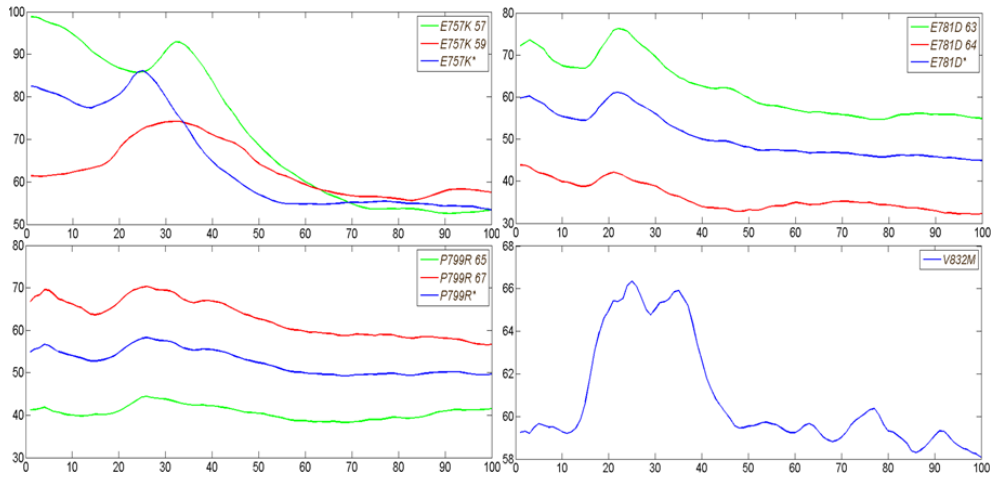


Figure 3.17: Comparison of the RD horizontal mean profiles concerning each form of E-cadherin, in this case E757K, E781D, P799R, V832M

Similarly to what was done before, GEFMs were developed combining all the RDMs concerning the same E-cadherin type. The mean and STD profiles of the GEFMs are displayed in Fig. 3.18 and Fig.3.19.

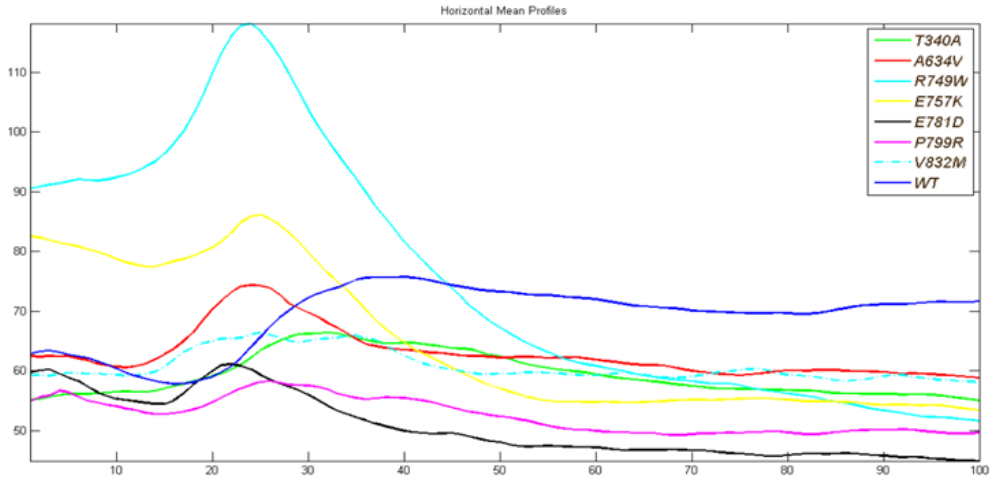


Figure 3.18: Comparison of the RD horizontal mean profiles of each form of E-cadherin map

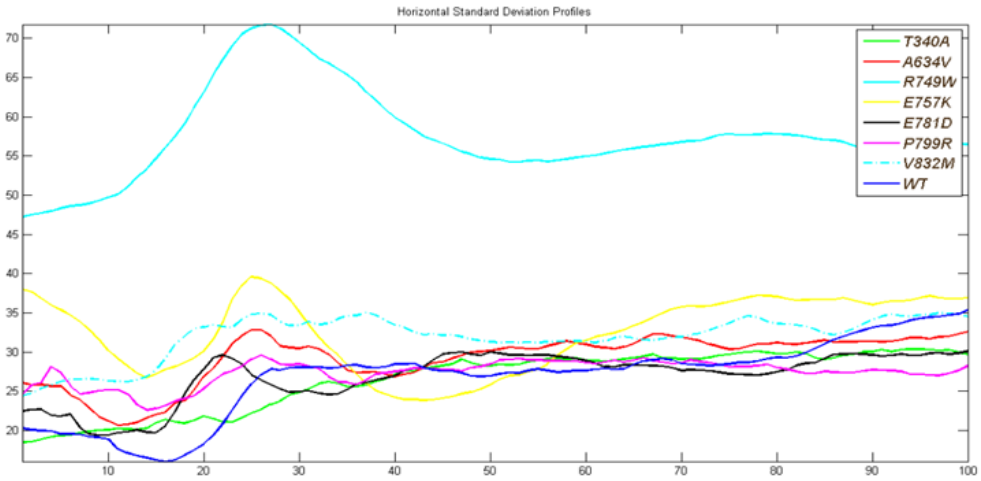
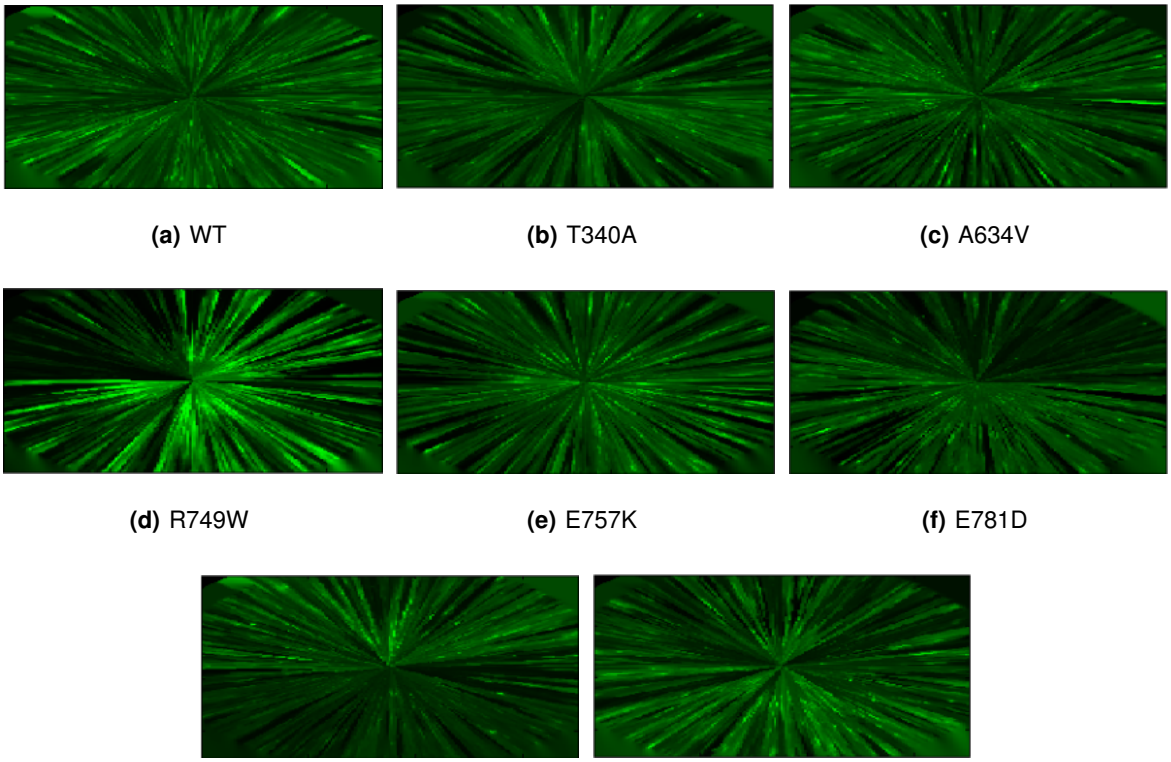


Figure 3.19: Comparison of the RD horizontal standard deviation profiles of each form of E-cadherin map

Regarding the average profiles of the Radial GEFMs is possible to observe that: R749W has the highest maximum intensity; R749W, E757K, A634V and E781D have the maximum value in similar position (approximately 0.25) ; T340A,V832M, P799R and WT display steadier profiles. Regarding the STD, is possible to observe that WT, T340A, A634V, E757K, E781D, P799R, V832M STD are similar in intensity and shape,whereas R749W STD is much higher. The creation of the cell profile from the respective RDM, described in 3.1.5, results are displayed in Fig. 3.20.



(a) WT **(b)** T340A **(c)** A634V

(d) R749W **(e)** E757K **(f)** E781D

(g) P799R **(h)** V832M

Figure 3.20: Cell profile reconstruction.

Fig. 3.20 allow to perceive, through a visual analysis, the cell membrane heterogeneities within

cells of the same culture. To correct this heterogeneities not related with trafficking dynamics, these maps were geometrically compensated.

3.2 Geometric Compensation

In this work, a new approach is proposed for the quantification and distribution of molecules within cells by characterization of average internuclear and radial profiles. These average profiles are difficult to obtain mainly due to heterogeneous cell geometric shape and size variability. This Bayesian algorithm compensate, in the INMs and RDMs, size and shape heterogeneities of the cells. This algorithm models each columns as a finite dimension continuous field, estimated from the initial observations, and imposing similarity between adjacent columns, allow the adjustment of each column observations, smoothing the solution. The adjustment of the observations position is shown in Fig. 3.21.

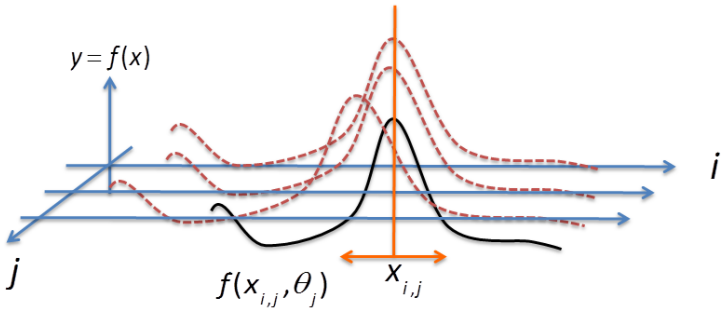


Figure 3.21: Iteration process of observation's adjustment.

This method based on the estimation of a continuous field in R^2, is divided in three major steps: **Initialization**, **Similarity Interpolation function**, **Observation's position adjustment**. In initialization, a function $y = f(x_j, c_j)$, depending on a set of parameters $\mathbf{c_j}$, is estimated from the set of intensities, y_j and corresponding locations x_j. Afterwards the compensation step is performed into two sub-steps where a energy function is minimized simultaneously with respect to the set of parameters, $\mathbf{c_j}$, defining the ideal profile and the compensated locations of the points of the intensity profile, $\mathbf{x_j}$:

$$\mathbf{c}_j^t = \arg\min_{\mathbf{c}} E(\mathbf{y}_j, \mathbf{x}_j^t, \mathbf{c}_j) \tag{3.1}$$

$$\mathbf{x}_j^{t+1} = \arg\min_{\mathbf{x}} E(\mathbf{y}_j, \mathbf{x}_j, \mathbf{c}_j^t) \tag{3.2}$$

The compensation step aims at estimating the vector of coefficients, $\mathbf{c_j}$, and simultaneously the new locations for the observations that regularize the map of profiles, $\mathrm{x_j}$. The regularization occur through the optimization of an energy function where the similarity of adjacent columns is imposed. In the end of this algorithm the final image is built moving the set of intensities, $\mathbf{y_i}$ to the new corresponding locations $\mathbf{x_j}$. The empty pixels are obtained by an interpolation process.

3.2.1 Mathematical Formulation

Let $Y = \{y_{i,j}\}$ be the $\mathrm{N} \times \mathrm{M}$ map of intensity profiles (RD or IN) obtained from M different N length normalized profiles, and $X = \{x_{i,j}\}$ the corresponding locations along the j^{th} profile. The initial locations, in the non-compensated map of profiles, after dimension normalization, are evenly distributed in the interval $[0, 1]$, meaning that $x^0_{i,j} = \frac{i}{(N-1)}$.

Let $f_j(\mathbf{x}_j, \mathbf{c}_j) = \sum_{k=0}^{L-1} c_{k,j}\phi_{k,j}(x)$ be a 1D finite continuous function, depending on a column vector of parameters, $\mathbf{c}_j = [c_{0,j}, c_{1,j}, ..., c_{L-1,j}]^T$, to describe the underlying ideal j^{th} column intensity profile of the image map of profiles. $y_{i,j}$ is the i^{th} intensity observation from the j^{th} profile taken at the $x_{i,j}$ location and $\phi_{k,j}(x)$ is the k^{th} basis function. These functions are evenly distributed in the interval $[0, 1]$, with $0 \leq k \leq L - 1$. They are shifted versions of a mother basis function according with

$$\phi_{k,j}(x) = \phi(\frac{x}{\Delta} - k) \tag{3.3}$$

where $\Delta = \frac{1}{(L-1)}$. The locations of the observations, $x_{i,j}$, are assumed to be geometrically distorted which means that they are corrupted by position noise,

$$x_{i,j} = x^*_{i,j} + \epsilon_{i,j} \tag{3.4}$$

Let also $C = \{c_{i,j}\}$ be a $L \times M$ matrix of coefficients, where $\mathbf{c}_j$ are the coefficients that define each j^{th} column of the continuous map profile . Each ideal profile function may be written as $f_j(x) = \Phi^T(x)\mathbf{c}_j$ where $\Phi(x) = [\phi_0(x)\phi_1(x), ..., \phi_{L-1}(x)]^T$ is a column vector containing the values of the L basis functions computed at location x. The optimal coefficients $\mathbf{c}_j$ and observations position $\mathbf{x}_j$ are estimated solving the following optimization problem:

$$[c_j, x_j]^* = \arg\min_{\mathbf{c}_j, \mathbf{x}_j} E(\mathbf{x}_j, \mathbf{y}_j, \mathbf{c}_j) \tag{3.5}$$

where the energy function minimized is:

$$E(\mathbf{x}_j, \mathbf{y}_j, \mathbf{c}_j) = E_y(\mathbf{x}_j, \mathbf{y}_j, \mathbf{c}_j) + E_p(\mathbf{c}_j) + E_c(\mathbf{c}_j) + E_x(\mathbf{x}_j). \tag{3.6}$$

In this equation the energy function is composed by one *data fidelity term* and three *prior terms*. The *data fidelity term*,

$$E_y(\mathbf{x}_j, \mathbf{y}_j, \mathbf{c}_j) = \sum_{i}^{N-1} (f_j(x_{i,j}) - y_{i,j})^2 \tag{3.7}$$

that pushes the solution towards the data. The first *prior term*,

$$E_p(\mathbf{c}_k) = \alpha \sum_{i} (c_{i,j} - c_{i-1,j})^2 \tag{3.8}$$

is used to stabilize the iterative process smoothing the solution imposing similarity of the coefficients describing each j^{th} profile , $c_{i,j} - c_{i-1,j}$. The second *prior term*,

$$E_c(\mathbf{c}_j) = \sum_{i,j=0}^{L-1,M-1} \beta(c_{i,j} - c_{i,j-1})^2, \tag{3.9}$$

smooths the solution imposing similarity between homologous coefficients on neighboring columns, $c_{i,j} - c_{i,j-1}$, in order to force the similarity of the ideal profiles, $f_j(x)$. The third prior term,

$$E_x(\mathbf{x}_j) = \gamma \sum_{i,j}^{N-1,M-1} (x_{i,j} - x_{i,j-1})^2, \tag{3.10}$$

is a prior function to keep the displacement compensation adjustment of the intensity locations at each profile under control and prevent degenerated solutions. The end locations, $x_{0,j}$ and $x_{N-1,j}$ are fixed with values 0 and 1 respectively. α, β e γ are prior hyper parameters.

Using matrix notation the equations (3.7) , (3.8), (3.9) and (3.10) can be written as follows

$$E_y(\mathbf{x}_j, \mathbf{y}_j, \mathbf{c}_j) = \sum_j (\Phi_j^T(\mathbf{x}_j)\mathbf{c}_j - \mathbf{y}_j)^T(\Phi_j^T(\mathbf{x}_j)\mathbf{c}_j - \mathbf{y}_j) \tag{3.11}$$

$$E_p(\mathbf{c}_k) = \alpha(\theta\mathbf{c}_j)^T(\theta\mathbf{c}_j) \tag{3.12}$$

$$E_c(\mathbf{c}_j) = \sum_j \left(\beta(\mathbf{c}_j - \mathbf{c}_{j-1})^T(\mathbf{c}_j - \mathbf{c}_{j-1})\right) \tag{3.13}$$

$$E_x(\mathbf{x}_j) = \gamma \sum_j (\theta\mathbf{x}_j)^T(\theta\mathbf{x}_j) \tag{3.14}$$

where $\Phi_j(\mathbf{x}_j) = [\Phi(x_{0,j}), \Phi(x_{1,j}), ...\Phi(x_{N-1,j})]$ is a $N \times L$ matrix computed for each j^{th} column profile and θ is the following difference operator:

$$\theta = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & -1 & 1 \end{bmatrix} \tag{3.15}$$

The energy function defined in (3.6) is minimized in three steps, according with:

$$\mathbf{c}_j^0 = \arg\min_{\mathbf{c_j}} E_y(\mathbf{x}_j, \mathbf{y}_j, \mathbf{c}_j)) + E_p(\mathbf{c}_j), 0 \leq j \leq M-1 \tag{3.16}$$

$$\mathbf{c}_j^t = \arg\min_{\mathbf{c_j}} E_y(\mathbf{x}_j^{t-1}, \mathbf{y}_j, \mathbf{c}_j) + E_c(\mathbf{c}_j), 0 \leq j \leq M-1 \tag{3.17}$$

$$\mathbf{x}_j^t = \arg\min_{\mathbf{x}_j} E_y(\mathbf{x}_j, \mathbf{y}_j, \mathbf{c}_j^t) + E_x(\mathbf{x}_j), 0 \leq j \leq M-1 \tag{3.18}$$

where t is the iteration index of the iterative optimization process where (3.17) and (3.18) steps alternate until convergence is achieved.

3.2.1.A Optimization

TThe minimization step (3.16) (Initialization), is performed finding the stationary point with respect to $\mathbf{c}_j$ of $\nabla_{\mathbf{c}_j} E(\mathbf{x}_j, \mathbf{y}_j, \mathbf{c}_j) = 0$, leading to:

$$(\Phi\mathbf{c}_j - \mathbf{y}_j)^T(\Phi\mathbf{c}_j - \mathbf{y}_j) + \alpha(\theta\mathbf{c}_j)^T(\theta\mathbf{c}_j) = 0, \tag{3.19}$$

with the following solution,

$$\mathbf{c}_k^0 = (\Phi^T\Phi + \alpha\Theta)^{-1}\Phi^T\mathbf{y}_k \tag{3.20}$$

where $\Theta = \theta^T\theta$.

The minimization step (3.17) (Similarity Interpolation), is performed by finding the stationary point with respect to $\mathbf{c}_j$, of $\nabla_{\mathbf{c}_j} E(\mathbf{x}_j^{t-1}, \mathbf{y}_j, \mathbf{c}_j) = 0$, that leads to:

$$\Phi_j(\Phi_j^T \mathbf{c}_j - \mathbf{y}_j) + \alpha\Theta^T \mathbf{c}_j + \beta\left[2\mathbf{c}_j - \mathbf{c}_{j-1} - \mathbf{c}_{j+1}\right] = 0, \tag{3.21}$$

with the following solution,

$$\mathbf{c}_j^t = \left(\Phi_j \Phi_j^T + \alpha\Theta^T + \beta I_L\right)^{-1}(2\beta\bar{\mathbf{c}}_j^{t-1} + \Phi_j \mathbf{y}_j) \tag{3.22}$$

where $\Theta = \theta^T\theta$, I_L is an L dimension identity matrix and $\bar{\mathbf{c}_j^{t-1}} = (\mathbf{c}_{j-1}^{t-1} + \mathbf{c}_{j+1}^{t-1})/2$ is the average of the neighboring columns.

The minimization step (3.18) (Observation's position adjustment) is obtained by computing the derivative with respect to the i^{th} element of the j^{th} profile of intensities $\mathbf{x}_j$ of $\frac{\partial E}{\partial x_{i,j}} = 0$, resulting in:

$$\frac{\partial E}{\partial x_{i,j}} = z_{i,j} + 2(x_j(i) - \bar{x}_{i,j}) = 0, \tag{3.23}$$

where $z_{i,j} = \frac{1}{\beta}(f_j(x_{i,j} - y_{i,j})\dot{f}_j(x_{i,j})$ and $\bar{\mathbf{x}}_{i,j} = \left(x_{i-1,j} + x_{i+1,j}\right)/2$ is the average values of the neighboring intensity locations. The resulting equation to compute $\mathbf{x}_j$ is:

$$x_{i,j} = \frac{1}{2}\left(\frac{z_{i,j}}{\gamma} + (\mathbf{x}_j^T \Upsilon_i^T)\right) \tag{3.24}$$

where Υ_i is the j^{th} line of a shift matrix operator:

$$\Upsilon = \begin{bmatrix} 2 & 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 2 \end{bmatrix} \tag{3.25}$$

3.2.2 Experimental Results

In this section results with synthetic and real data are presented. The algorithm was implemented, and data processed, in $Matlab$®software.

3.2.2.A Synthetic Data

The synthetic data, displayed in Fig.3.22, 3.23, 3.24, 3.25.a), to illustrate the application of the proposed method, are 256×256 gray scale black images with a white line representing the distorted fluorescent cellular membrane. This images simulate internuclear and radial profiles of non ideal circular cells. In this experiment four synthetic image were processed:

- An oblique straight line (see fig. 3.22 a))
- Several oblique straight lines with different slopes. (see fig. 3.23 a))
- An half arc of cosine (see fig. 3.24 a))
- A complete arc of cosine (see fig. 3.25 a))

The first two images intend to simulate the information distortions obtained in the INPs, and the last two images aim at simulating the distorted cellular membrane characteristic in RDPs. The result of the GC of this synthetic profiles is displayed in Fig. 3.22, Fig 3.23, Fig. 3.24 and 3.25.

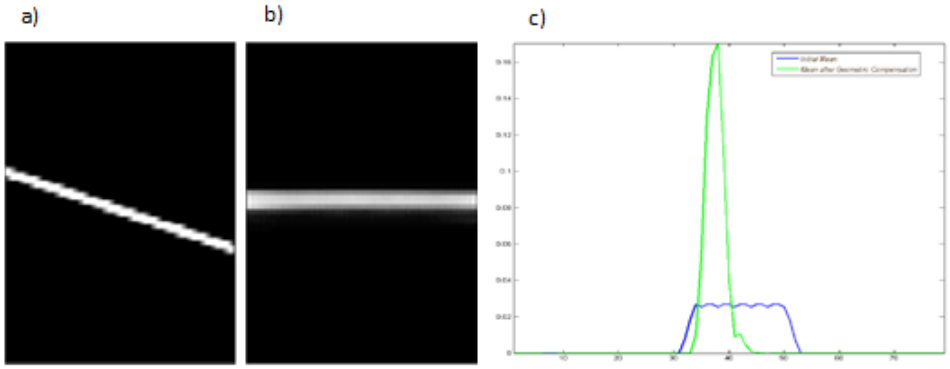


Figure 3.22: Synthetic data. a) An oblique straight line (initial), b) Geometric compensated data , c) Mean comparison before and after GC

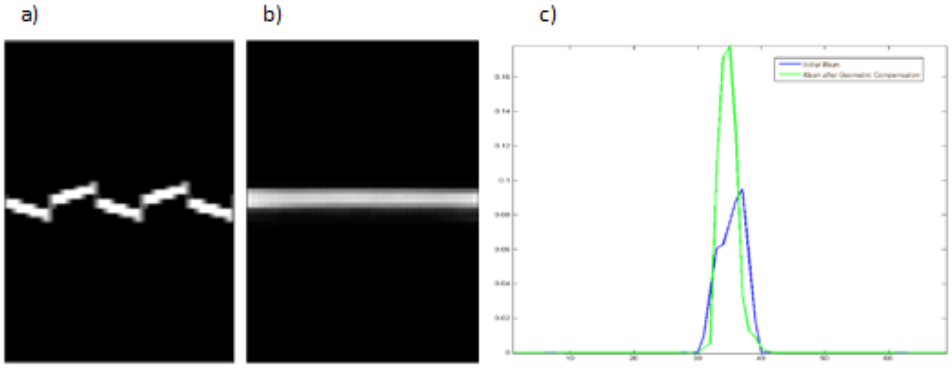


Figure 3.23: Synthetic data. a) Several oblique straight lines (initial), b) Geometric compensated data , c) Mean comparison before and after GC

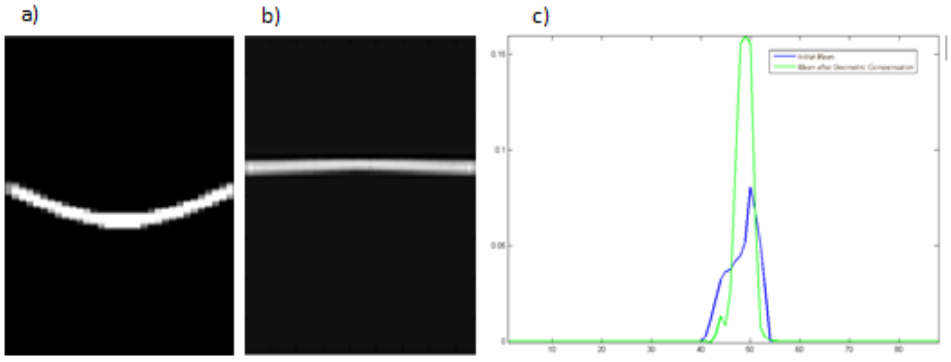


Figure 3.24: An half arc of cosine (initial),b) Geometric compensated data , c) Mean comparison before and after GC

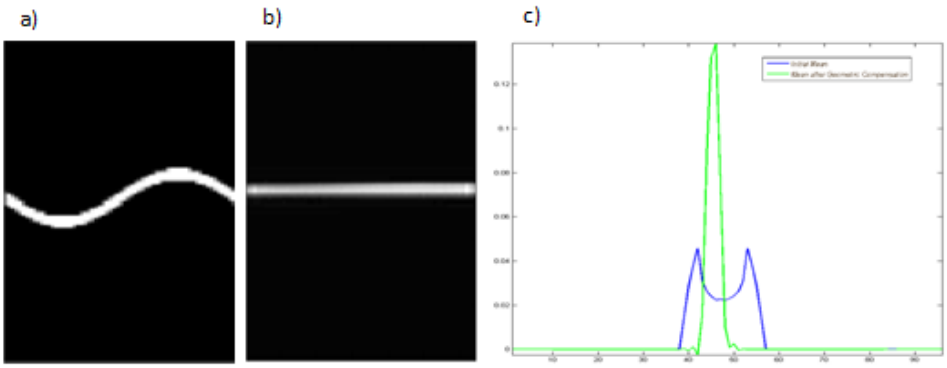


Figure 3.25: A complete arc of cosine (initial), b) Geometric compensated data , c) Mean comparison before and after GC

This results show that the profiles were correctly compensated, displaying the stability of the algorithm for different initial data geometries. All the images show a columns smoothness effect due to the alignment of high intensity observations in adjacent columns. In coherence with what was expected, the observations align in one straight line which results in a higher maximum intensity and lower dispersion in the horizontal mean vector. Also the standard deviation diminishes (see fig. 3.22,3.23,3.24,3.25 c)).

3.2.2.B Real FM Data

A real IF image with WT E-cadherin tagged with Green Fluorescent Proteins (GFPs) is used here to illustrate the application of this algorithm to real data profiles collected from fluorescence microscopy images. TheIF image is shown in Fig.3.26. Two types of Maps were geometrically compensated:

- IN Map (fig.3.27 b)).
- RD Map (fig.3.27 c)).

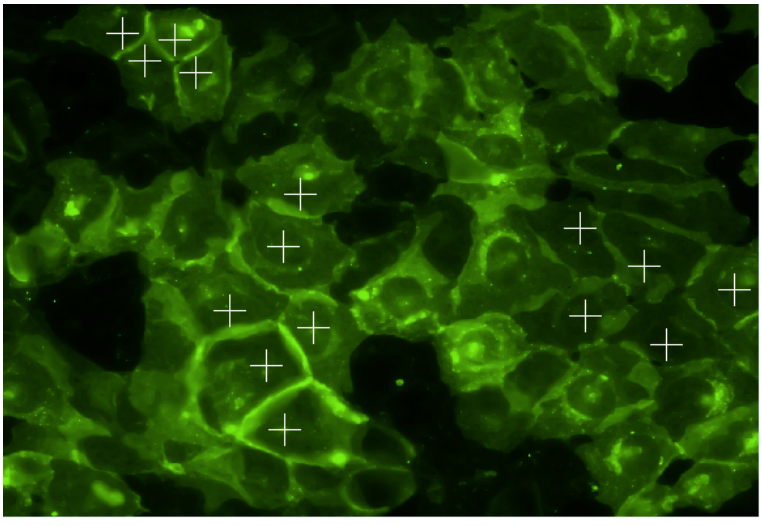
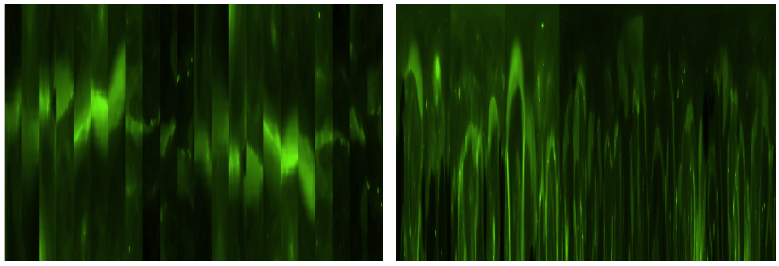


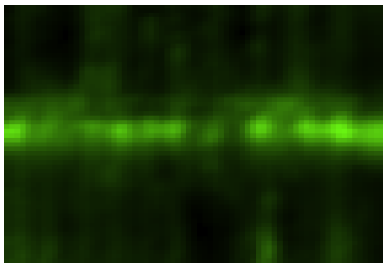
Figure 3.26: Real immunofluorescence image with tagged E-Cadtherin.

Both of the maps processed using this algorithm came from the preprocessing and profiles collection of the same FM image containing WT E-cadherin (Nº 38), and the results are shown in Fig. 3.27. The images b) and d) of Fig. 3.27 were built moving the initial observations to the new calculated positions, $\hat{x}_k$ and interpolating the empty remaining pixels.

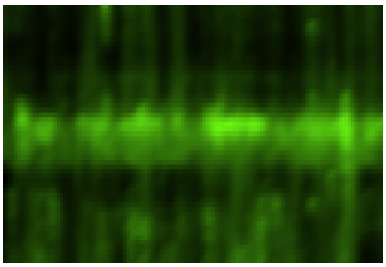


(a) Non comp. IN profile map

(b) Non comp. Radial profile map



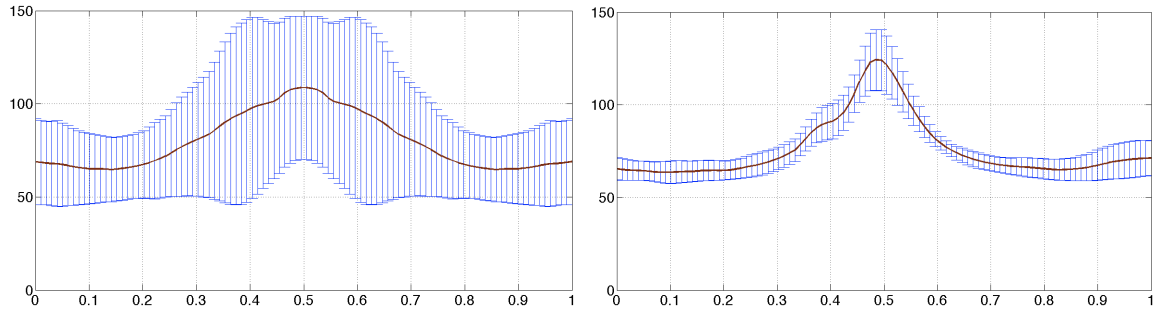
(c) Compensated IN profile map



(d) Compensated Radial profile map

Figure 3.27: Real immunofluorescent images of tagged E-Cadtherin. a) Original plaque, b) IN profiles from selected cells, c) Radial profiles from selected cells, d) geometric compensated IN map of profiles, e) d) geometric compensated Radial map of profiles.

The INMs contain information corrupted with distortions due to the process of collection and biologic heterogeneities in shape and size of cellular membranes. Not only the profiles are collected in different directions but also the membrane is not a straight well defined line due to biologic shape variability. To correct the variations described before, INMs were geometrically compensated and the results are displayed in Fig. 3.27 a,b). To quantitatively compare the INMs before and after geometric compensation both the mean and STD profiles were computed, in column direction, and are displayed in Fig. 3.28 a) and b), respectively.



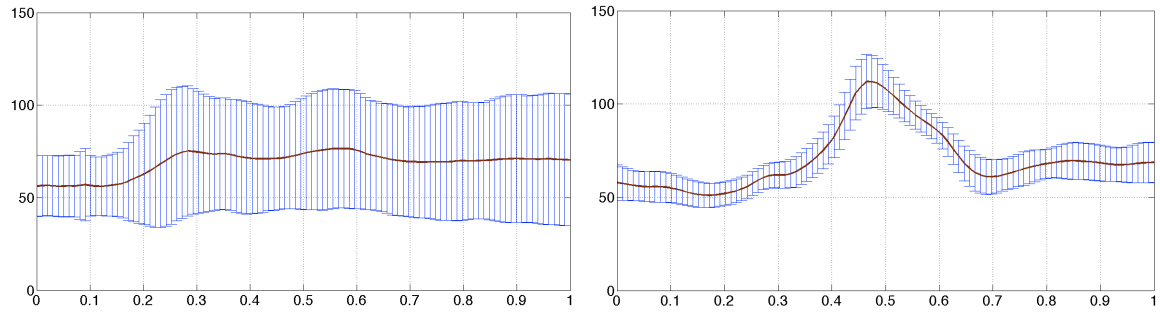
(a) Non comp. IN profile map)

(b) Compensated IN profile map

Figure 3.28: IN horizontal mean profile combined with respective STD in each point, a) before and b) after GC.

In Fig. 3.27 b) the alignment of similar intensity observations in adjacent columns lead to the appearance of an almost straight line in the horizontal center of the image. This high intensity line represents the cell membrane, where E-cadherin performs its adhesive function. As can be observed in 3.28 both E-cadherin's membrane position and Membrane's intensity are better defined after GC. This can be seen by the increase in the intensity peak value of the horizontal mean vector obtained after the geometric compensation in comparison with the initial one. This is reflected in the increase of mean's norm in position $x = 0.5$ (membrane), and the overall decrease in data dispersion. The STD significantly decreases after GC, specially near the cellular membrane position.

The RDMs suffer from variations due to the membrane biological shape and size heterogeneity. Because cell's nuclei and membranes are not perfectly spheric, the polar collection of data nuclei centered will not be a straight line (can be seen in fig. 3.27) c). To compensate the diversity described, radial profiles were geometrically compensate using the algorithm described. The results are displayed in Fig. 3.27 c,d). To quantitatively compare the RDMs before and after geometric compensation both the mean and STD profiles were computed, in column direction, and are displayed in Fig. 3.28 a) and b), respectively.



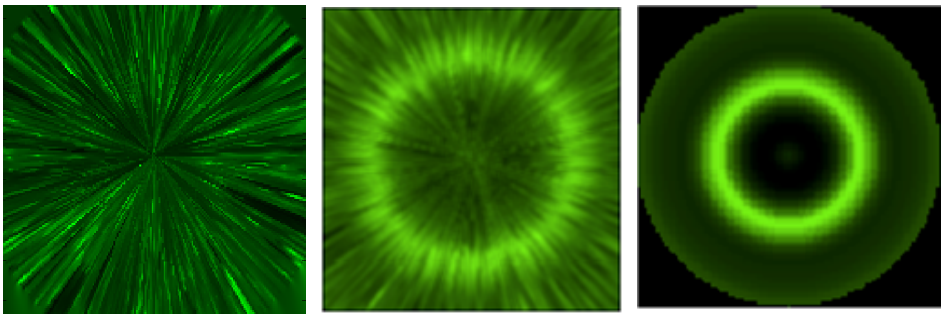
(a) Non comp. RD profile map

(b) Compensated RD profile map

Figure 3.29: RD horizontal mean profile combined with respective STD in each point, a) before and b) after GC.

Fig. 3.27 c) shows that there is actually a movement of alignment between higher intensity observations as expected. In Fig. 3.29, that compares the horizontal mean vectors and STD of the RDM before and after GC, is shown that STD decreases substantially after GC. This represents that

the columns are similar, displaying the high and low intensities in same position leading to a better definition of both the membrane's actual position and intensity.The geometrically compensated map mean profile display a well defined centered high intensity peak . This increase in intensity represents an increase in E-cadherin molecules. As this peak position is near $\mathrm{x}_{0.5}$, is expected that this peak represents the cell membrane, where E-cadherin performs its adhesion role. The creation of the cell profile from the respective RDM, described in 3.1.5, results are displayed in Fig. 3.30.



(a) Before GC **(b)** After GC **(c)** Virtual

Figure 3.30: Cell profiles

Fig. 3.30 a) allow to perceive, through a visual analysis, the cell membrane heterogeneities within cells of the same culture. This differences are attenuated after GC, since the alignment of the intensities in the geometrically compensated RDM lead to the urge of a well defined spherical line in the compensated and virtual cell profiles.

4

Feature Collection and Statistical Analysis

Contents

4.1 Features

The expression pattern of membrane proteins at inter-cellular level, namely the cell membrane, give information that reflects the functional activity of E-cadherin protein, leading to a better understanding of the functional results of some structure alterations. To characterize the distribution both at membrane and cytoplasmic level, several quantitative objective features that reflect E-cadherin expression level are extracted. The features selected intend to discriminate the distribution of functional, healthy E-cadherin from the expression of dysfunctional, mutated E-cadherin.

The features were collected from the horizontal mean and STD profiles of the geometrically compensated INMs and RDMs. Mean and STD profiles reflect not only E-cadherin molecules concentration distribution but also molecules relative position in cell. The features collected from the mean profile are described on Table 4.1. These features reflect the distribution of E-cadherin whithin the cell, through the quantification of: high concentrations of E-cadherin intensity and relative position; E-cadherin concentration in the membrane; abnormal high levels of E-cadherin in the cytoplasm.

Table 4.1: Features Description (Mean profile)

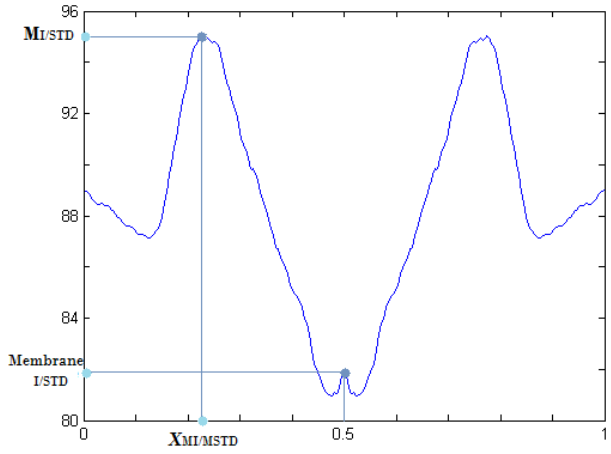
Maximum Intensity (MI)	Maximum value observed in the mean profile.
Position of MI	Maximum Intensity (MI) position, x.
Membrane's Intensity	Intensity value at cell membrane, $f(x_{0.5})$.
Mean Intensity	Profile's mean intensity, $\hat{x}$.
IntensitySTD	Measure of mean profile variation/dispersion concerning the average,$\hat{x}$.
Maximum Mean Ratio (MMR)	Ratio between maximum and mean intensity.
Total Variation (TV)	Measure of mean profile behavior in adjacent observations.

The features collected from the STD profile are described on Table 4.2.

Table 4.2: Features Description (STD profile)

Maximum Standard Deviation (MSTD)	Maximum value observed in STD profile.
Position of the MSTD	MSTD position, x.
Membrane's STD	STD value at cell membrane, $f(x_{0.5})$.
Mean STD	Profile's mean STD, $\hat{x}$.
Maximum Mean STD Ratio	Ratio between maximum and mean STD.
Total Variation (TV)	Measure of STD profile behavior in adjacent observations.

Fig. 4.1 illustrates the collection of: MI or MSTD; Position of MI/MSTD and Membrane's intensity/STD. Also, presentes the mathematical formulation of: mean $\hat{x}$, STD, MMR and TV value.



Mean

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

Standard Deviation

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2}$$

Maximum Mean Ratio

$$\text{MMR} = \frac{\text{M}}{\bar{x}}$$

Total Variation

$$V_b^a(f) = \sup_P \sum_{i=0}^{n_P - 1} |f(x_{i+1}) - f(x_i)|$$

Figure 4.1: Features collection and mathematical formulation.

4.2 Data Analysis

To characterize and discriminate among WT E-cadherin and remaining mutant forms both a brief clustering analysis and a final statistical analysis were developed. The clustering method chosen was the k-means algorithm. This algorithm minimizes the sum, over all clusters, of the within-cluster sums of point-to-cluster-centroid Euclidean distances, grouping the data in terms of resemblance. This algorithm was implemented to group the feature collection results, to study for each feature the resemblances between WT E-cadherin and mutant forms. Afterwards, a statistical analysis was developed to study the discriminative potential of the features collected from the mean profiles. In this work, Mann-Whitney U test were developed to compare for each feature selected the null hypothesis that WT E-cadherin and each one of the mutant forms are similar. This statistical analysis provides a dissimilarity analysis in terms of the features collected between WT E-cadherin and remaining HDGC related mutant forms.

4.2.1 Clustering Analysis

K-means clustering (MacQueen, 1967) is a method commonly used to automatically partition N-dimensional data into k sets. It proceeds by selecting k initial cluster centers and then iteratively refining them as follows:

1. Each instance d_i is assigned to the closest cluster center.

2. Each cluster center C_j is updated to be the mean of its constituent instances.

The algorithm converges when there is no further change in assignment of instances to clusters [53]. In *Matlab*®the function that implements k-means clustering is k-means. kmeans(X,k), partitions the points in the n -by -p data matrix X into k clusters. In this case since we are working with the results of the feature collection which means that n=10 (number of maps) and p=1 as features were analyzed separately. The number initial clusters were $k = \frac{n}{2} - 1 = 4$. This iterative partitioning minimizes

the sum, over all clusters, of the within-cluster sums of point-to-cluster-centroid Euclidean distances. *kmeans* returns an n-by-1 vector IDX containing the cluster indices of each point. [54] In Tables 5.2, 5.4, 5.7 and 5.9 each value display the color of the respective cluster to which it belongs. This mean that values displayind the same color were assigned the same cluster and on the other hand values displaying different colors were assigned to different clusters. This algorithm provides an assessment of resemblance.

4.2.2 Statistical Analysis

The Mann-Whitney U test (also called the Mann-Whitney-Wilcoxon (MWW)) is an overall comparison of distributions in terms of both shape and location. Usually, this test is used to test the null hypothesis that two samples come from the same population, i.e. have the same median. The alternative hypothesis is that observations in one sample tend to be larger than observations in the other. Although it is a non-parametric test, it assumes the two distributions are similar in shape. Initially, Mann-Whitney U test assumes the existence of a population x ,$x_1, x_2, ..., x_n$, and another population y, $y_1, y_2, ...y_m$. Afterwards, the method compares each observation x_i with every observation y_j obtaining a total number of pairwise comparisons of $n \times m$. If the populations have the same meanl then each x_i has an equal chance , $P = \frac{1}{2}$ of being greater or smaller than each y_j. Consequently, the the null hypothesis is $H_0 : P(x_i > y_j) = \frac{1}{2}$ and the alternative hypothesis $H_1 : P(x_i > y_j) \neq \frac{1}{2}$. With this hypothesis, the algorithm estimates the relation between all the observations x_i and y_j. The number of times x_i is greater than y_j is denoted as U_x and the number of times x_i is smaller than y_j is denoted as U_y. The null hypothesis assumes $U_x \approx U_y$. [55].

The test equivalent to a Mann-Whitney U-test in *Matlab@* is *ranksum*. *ranksum* tests the null hypothesis that data in x and y are samples from continuous distributions with equal medians, against the alternative that they are not. The function $[p, h] = ranksum(x, y)$ returns the $p - value$ of a two-sided Wilcoxon rank sum test and a logical value indicating the test decision, h. If $h = 1$ indicates a rejection of the null hypothesis, and $h = 0$ indicates a failure to reject the null hypothesis at the 5% significance level. [56]

5

Characterization of E-cadherin distribution: Experimental Results

Contents

The characterization of the E-cadherin distribution is divided in two different analysis: INMs analysis, described in section 5.1 and RDMs analysis described in section 5.2. In this chapter, several quantitative objective features that reflect E-cadherin expression level are extracted to characterize the distribution both at membrane and cytoplasmic level. Also, a statistical analysis is developed to test the discriminative potential of the features selected to identify WT E-cadherin among the HDGC related selected mutant forms.

5.1 Internuclear Profiles Characterization

The geometrically compensated maps were decomposed into two profiles: mean intensity profile and STD profile. Both these profiles were collected in lines direction, to study E-cadherin molecules distribution concerning the *nuclei* involved in the process of collection. Also, to establish a phenotype-mutation relation, all the maps containing the same kind of E-cadherin were displayed together. Finally, GEFM combining INMs of the same form of E-cadherin were developed and are marked with the * symbol. These profiles are displayed in Fig. 5.1 and Fig.5.2.

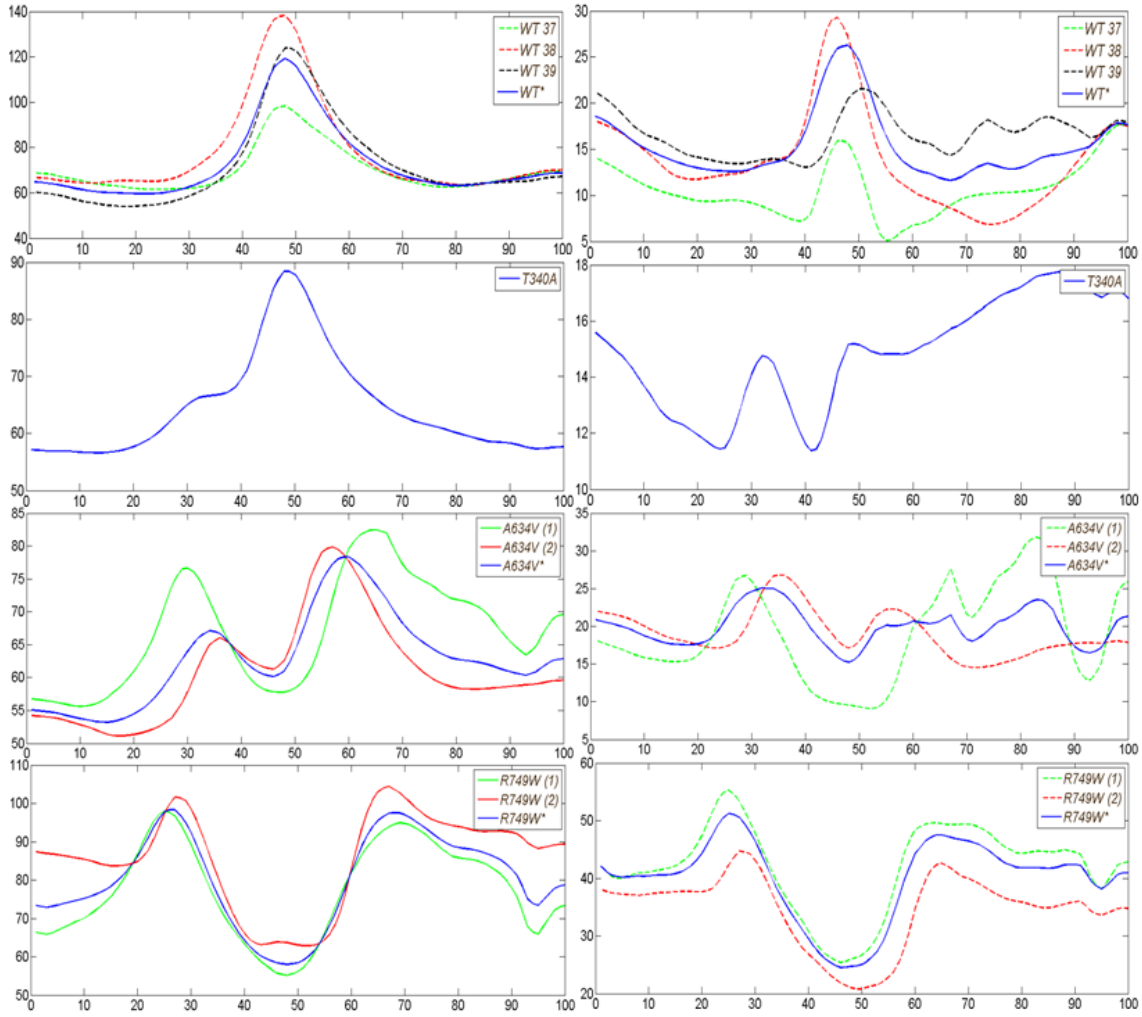


Figure 5.1: Comparison of the INMs concerning each form of E-cadherin, in this case WT, T340A, A634V, R749W. (left) horizontal mean profiles, (right) horizontal standard deviation

INMs of WT, A634V and R749W, among the same kind, have similar horizontal mean profiles both in maximum intensity as in shape. Also, in these cases, STD profiles magnitude is similar. Therefore, only the GEFM of each kind of E-cadherin is considered on further analysis. In P797R particular case, although the intensity of the plaques are different both the mean profile shape and STD are similar. This suggests that the intensity variation is due mainly to the process of acquisition of data. Thus, from now on, the GEFM is used to represent this form of E-cadherin.

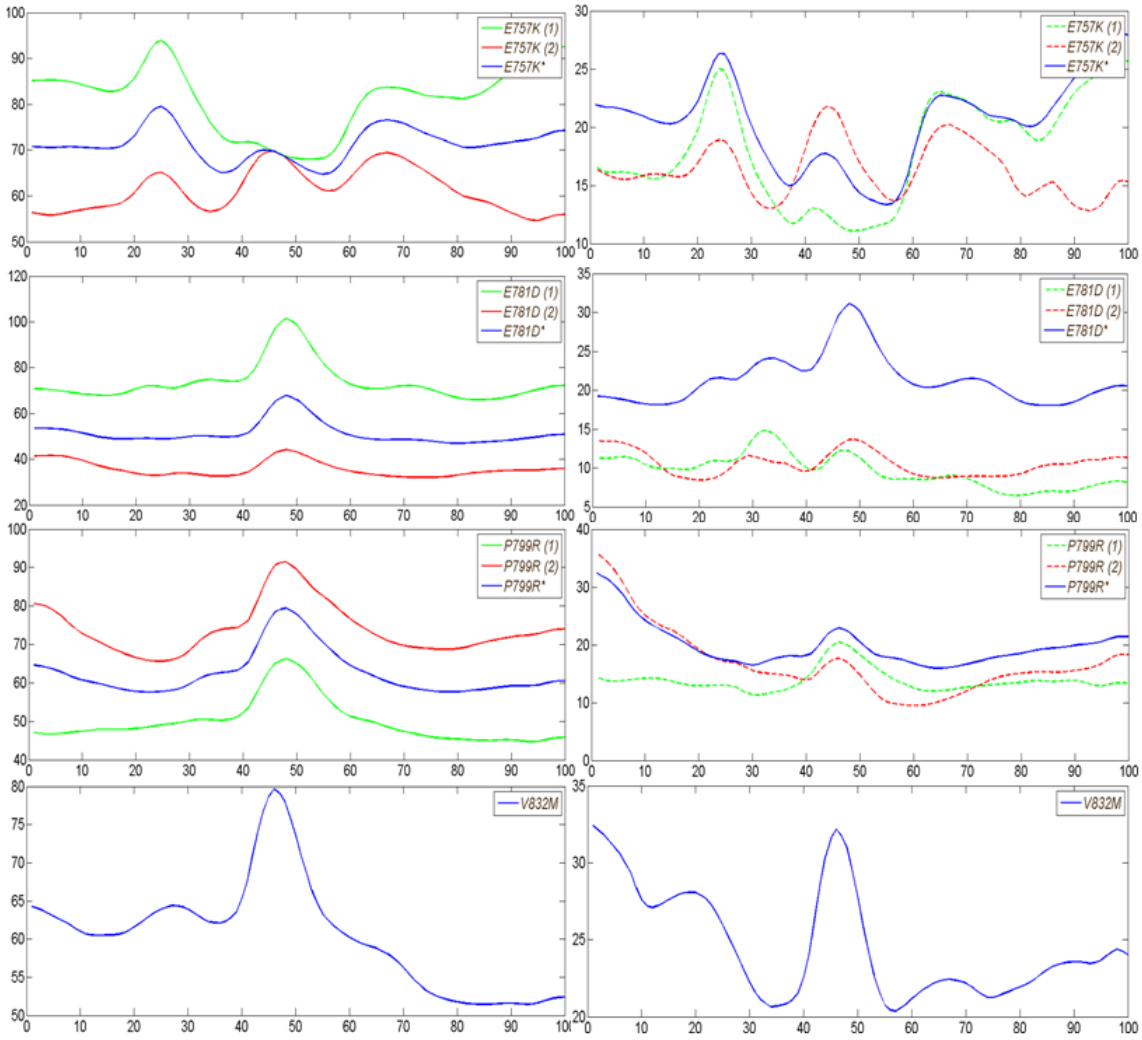


Figure 5.2: Comparison of the INMs concerning each form of E-cadherin, in this case E757K, E781D, P799R, V832M. (left) horizontal mean profiles, (right) horizontal standard deviatio.n

In contrast E757K, E781D INMs, display either a dissimilar horizontal mean profile or an increase in STD profile magnitude in the GEFM. Accordingly, to avoid inaccurate judgments both these maps will be evaluated separatly as: E757K (1) or (57), E757K (2) or (59), E781D (1) or (63) and E781D (2) or (64). Finally, T340A and V832M only INM is threated as GEFM. The maps selected after this mutation behavior analysis are displayed in Fig. 5.3.

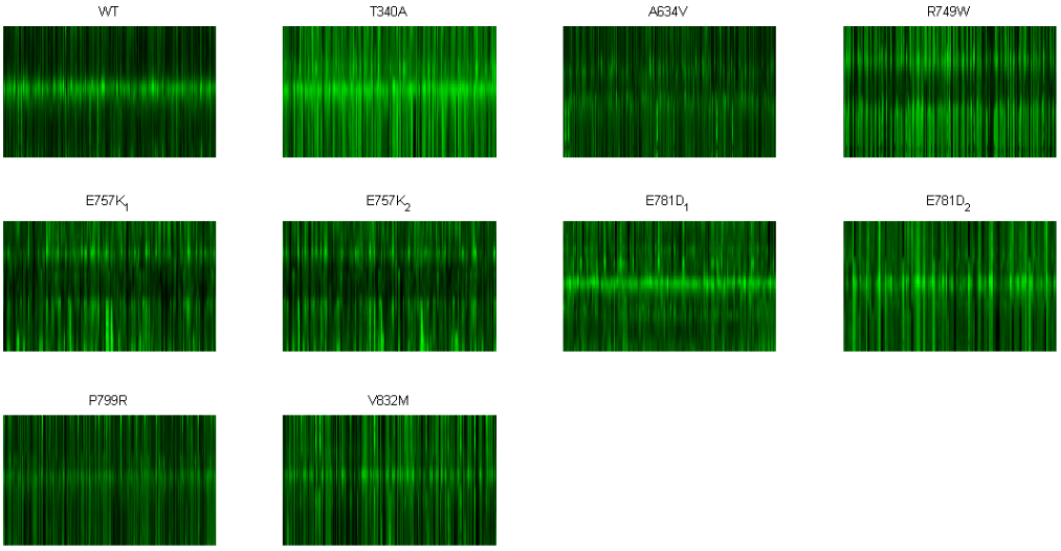


Figure 5.3: Geometrically compensated INMs selected.

The molecules distribution, in each GEFM, is described through the mean projection and respective STD interval in each point in Fig.5.4.

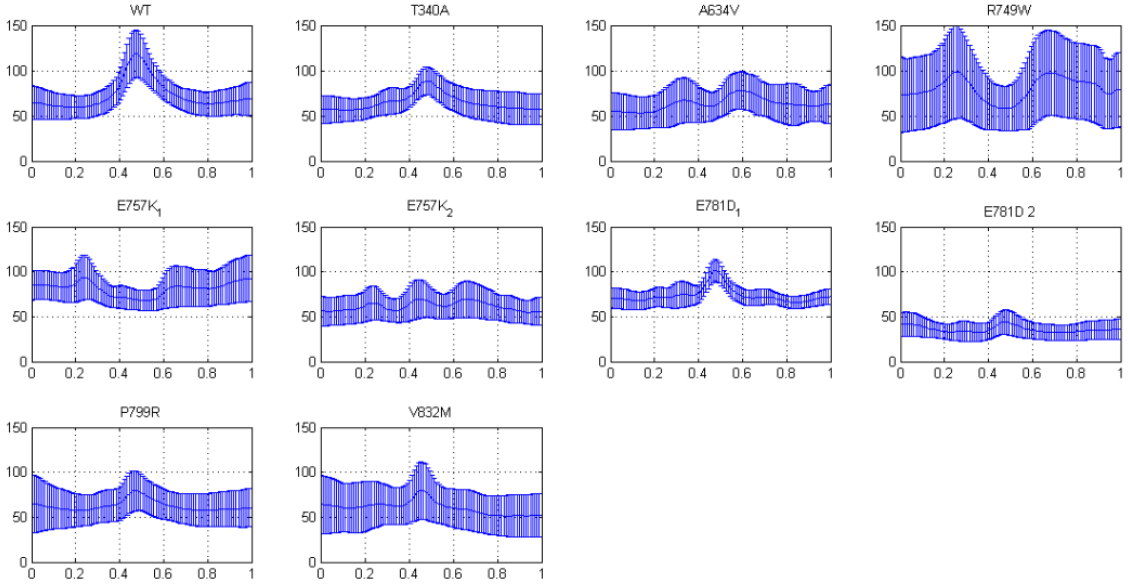


Figure 5.4: IN Horizontal mean profile combined with respective STD interval in each point.

As can be seen in Fig.5.4, WT E-cadherin has a quite different phenotype either qualitatively or quantitatively when compared with E-cadherin mutant forms. To perceive qualitative differences between GEFMs, the mean and STD profiles of the maps selected were computed and are displayed in Fig. 5.5 and Fig.5.6.

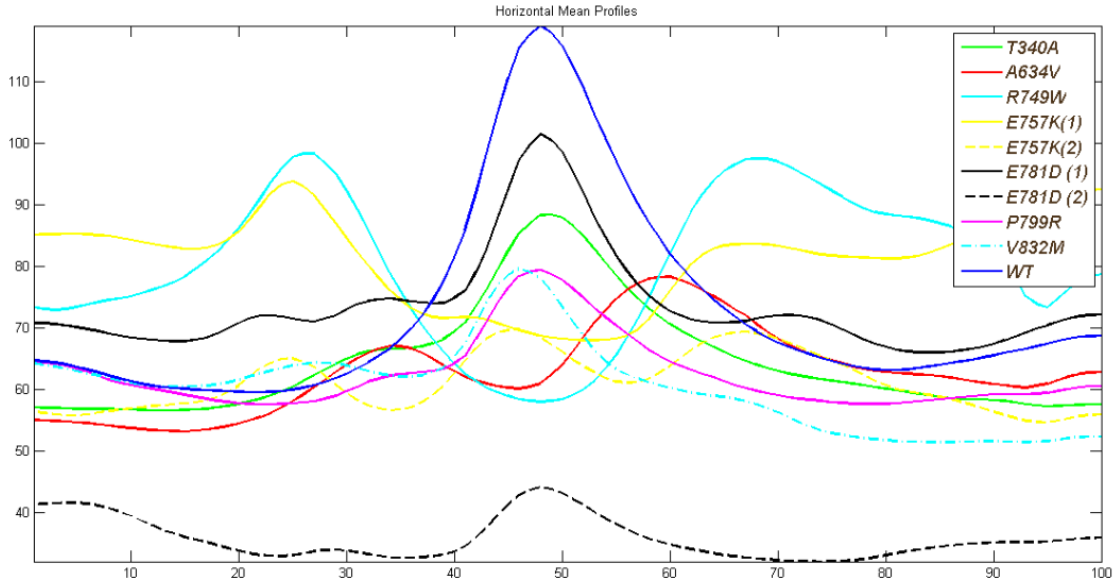


Figure 5.5: Distinct horizontal mean profiles after GC.

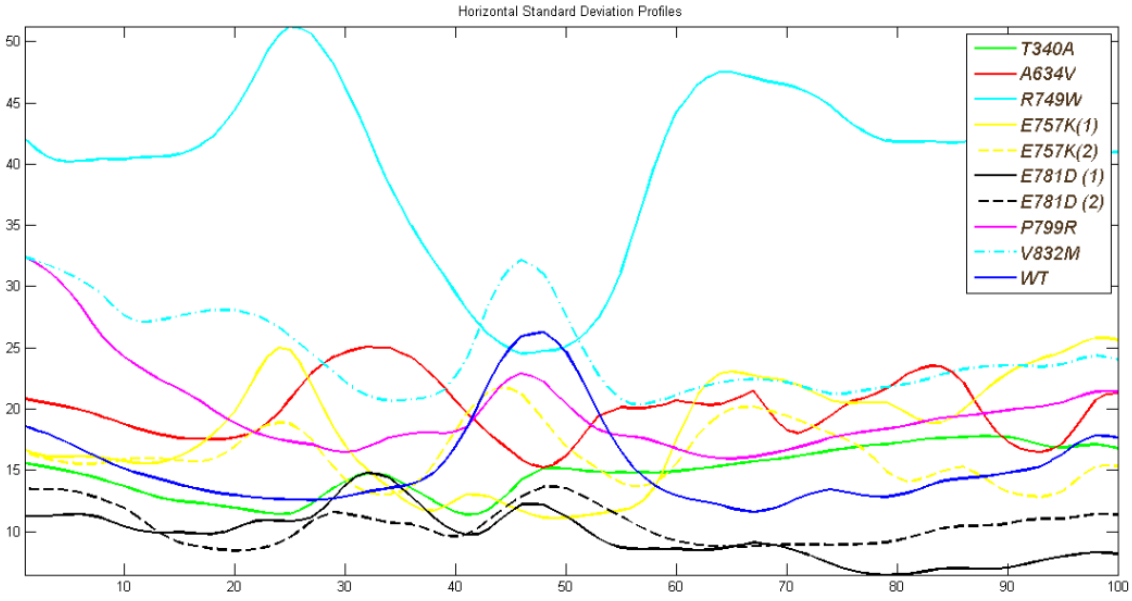


Figure 5.6: Distinct Horizontal STD profiles after GC.

Fig.5.5 and 5.6 show that qualitatively WT profiles are quite different either in shape or in intensity distribution. To quantitatively characterize the distribution both at membrane and cytoplasmic level, and distinguish the E-cadherin expression level for the different types of E-cadherin, objective features were extracted.

5.1.1 Feature Collection

To distinguish E-cadherin expression level in the final GEFMs, objective features were extracted from mean and STD profiles. The features collected from mean profiles were: Maximum Intensity (MI), Position of the MI, Membrane's intensity, Mean Intensity, Intensity STD, Maximum Mean Ratio (MMR) and Total Variation (TV). The quantitative results of mean profile feature collection are displayed in

the Table 5.1

Table 5.1: Features collected from GEFMs' mean profiles.

Mean Profile		Maximum Intensity MI	Position of the MI	Membrane's intensity	Mean Intensity	Intensity STD	MMR	Total Variation (TV)
WT	1	119.15	0.48	115.88	72.77	16.29	1.64	20.02
T340A	2	88.41	0.49	87.92	64.58	8.94	1.37	9.84
A634V	3	78.33	0.61	64.01	62.96	6.87	1.24	8.13
R749W	4	98.35	0.26	58.44	80.11	11.92	1.23	17.16
E757K (1)	5	93.87	0.25	68.22	81.28	7.31	1.15	9.98
E757K (2)	6	69.78	0.46	66.09	61.26	4.63	1.14	8.02
E781D(1)	7	101.52	0.48	98.66	73.65	8.18	1.38	14.09
E781D(2)	8	44.17	0.48	43.18	35.75	3.36	1.24	5.57
P799R	9	79.44	0.48	77.79	62.50	5.78	1.27	8.17
V832M	10	79.66	0.46	73.53	60.44	7.12	1.32	9.28

To understand quantitative reasons of similarity in each one of these features a k-means clustering analysis was performed. Each element display one color, representing the cluster it belongs. Therefore, elements contained in the same cluster display the same colors and consequently, elements of different cluster display different colors.The results of this evaluation is displayed in Table 5.2.

Table 5.2: Features collected from GEFMs' mean profiles after a kmeans analysis.

Mean Profile		Maximum Intensity MI	Position of the MI	Membrane's intensity	Mean Intensity	Intensity STD	MMR	Total Variation (TV)
WT	1	119.15	0.48	115.88	72.77	16.29	1.64	20.02
T340A	2	88.41	0.49	87.92	64.58	8.94	1.37	9.84
A634V	3	78.33	0.61	64.01	62.96	6.87	1.24	8.13
R749W	4	98.35	0.26	58.44	80.11	11.92	1.23	17.16
E757K (1)	5	93.87	0.25	68.22	81.28	7.31	1.15	9.98
E757K (2)	6	69.78	0.46	66.09	61.26	4.63	1.14	8.02
E781D(1)	7	101.52	0.48	98.66	73.65	8.18	1.38	14.09
E781D(2)	8	44.17	0.48	43.18	35.75	3.36	1.24	5.57
P799R	9	79.44	0.48	77.79	62.50	5.78	1.27	8.17
V832M	10	79.66	0.46	73.53	60.44	7.12	1.32	9.28

According with Table 5.2, the most discriminant features are Membrane's Intensity and the MMR. In these features WT E-cadherin cluster has 1 element. In Intensity STD, Mean Intensity and TV the cluster assigned to WT E-cadherin have 2 elements. In the remaining features, WT E-cadherin cluster has at least 4 elements. This suggest that neither MI or Position of the MI are good discriminating between WT E-cadherin and selected mutant forms. A further statistical analysis (Mann-Whitney U test) is performed to these features to understand the feature behavior in each profile constituent of these GEFMs. This analysis is present in section 5.1.2. To visually corroborate the k-means analysis, Fig. 5.7 displays simultaneously the WT mean profile and mutant forms mean profile.

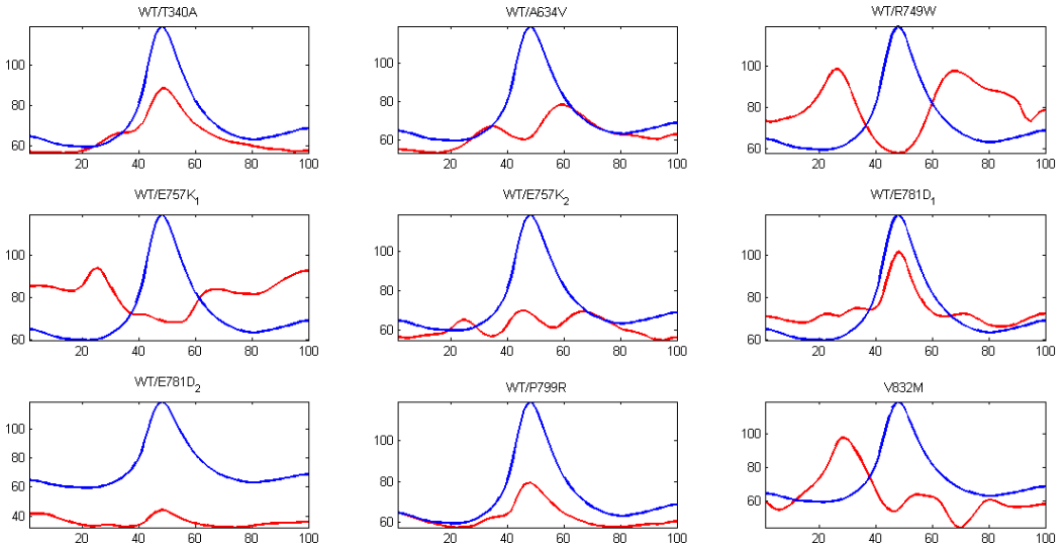


Figure 5.7: Mean profile comparison between WT E-cadherin and mutant forms.

Finally, the features collected from STD profiles were: Maximum Standard Deviation (MSTD), position of the MSTD, STD at the membrane, mean STD, Maximum Mean STD Ratio, Total Variation (TV). The results of the feature collection are shown in the Table 5.3

Table 5.3: Features collected from GEFMs' STD profiles.

Variance Profile							
		MSTD	Position MSTD	Membrane's STD	Mean STD	Maximum Mean STD Ratio	Total Variation (TV)
WT	1	26.28	0.48	24.68	15.41	1.70	6.30
T340A	2	17.78	0.89	15.16	14.84	1.20	2.69
A634V	3	25.071	0.32	16.16	19.95	1.26	6.16
R749W	4	51.24	0.25	25.07	39.93	1.28	10.57
E757K	5	25.82	0.99	11.12	18.11	1.43	7.21
E781D (1)	6	21.77	0.44	16.98	16.32	1.33	5.93
E781D (2)	7	14.81	0.32	11.29	9.64	1.54	3.42
P799R (1)	8	13.67	0.49	13.55	10.53	1.30	2.99
P799R(2)	9	32.41	0.01	20.46	19.83	1.63	4.65
V832M	10	32.40	0.01	27.75	24.64	1.32	6.67

A similar cluster analysis to study the quantitative profiles resemblances was performed and the results are displayed in the Table 5.4.

Table 5.4: Features collected from GEFMs' STD profiles after a k-means analysis

Variance Profile		MSTD	Position MSTD	Membrane's STD	Mean STD	Maximum Mean STD Ratio	Total Variation (TV)
WT	1	26.28	0.48	24.68	15.41	1.70	6.30
T340A	2	17.78	0.89	15.16	14.84	1.20	2.69
A634V	3	25.071	0.32	16.16	19.95	1.26	6.16
R749W	4	51.24	0.25	25.07	39.93	1.28	10.57
E757K	5	25.82	0.99	11.12	18.11	1.43	7.21
E781D (1)	6	21.77	0.44	16.98	16.32	1.33	5.93
E781D (2)	7	14.81	0.32	11.29	9.64	1.54	3.42
P799R (1)	8	13.67	0.49	13.55	10.53	1.30	2.99
P799R(2)	9	32.41	0.01	20.46	19.83	1.63	4.65
V832M	10	32.40	0.01	27.75	24.64	1.32	6.67

The k-means analysis suggest that the feature with higher discriminant potential is Maximum Mean STD Ratio, where WT E-cadherin actually show the highest value. In these features WT E-cadherin cluster has 2 element. In Membrane's MSTD and Position of MSTDthe cluster assigned to WT E-cadherin have 3 elements. However, the feature Position of MSTD discriminative potential may be reduced by the similarity of the values of P799R(2) and WT E-cadherin. Concerning the feature collection from STD profiles, Total Variation, MSTD and Mean MSTD are not good features as WT E-cadherin cluster has more than 4 elements. To visually corroborate the k-means analysis, Fig. 5.8 displays simultaneously the WT STD profile and mutant forms STD profile.

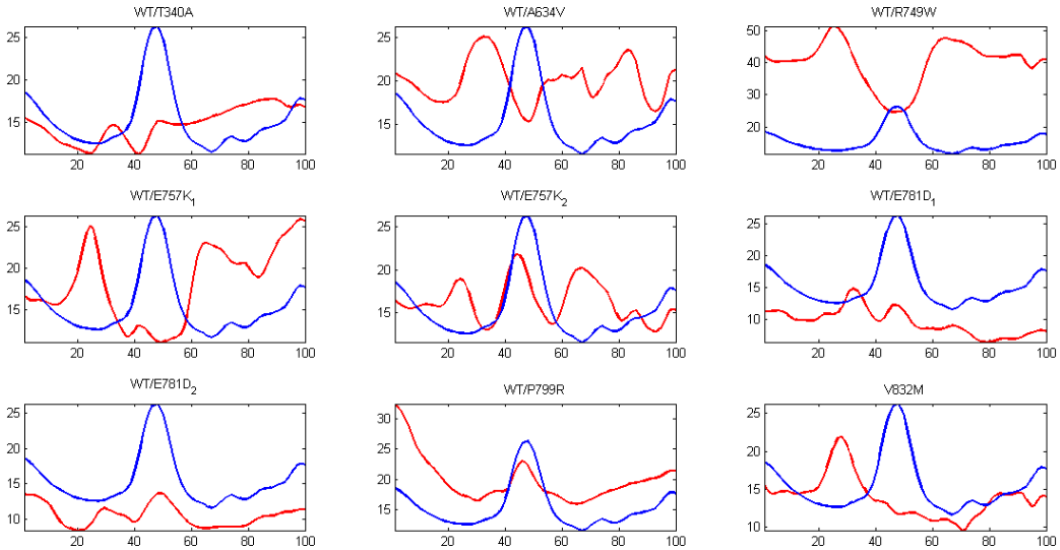


Figure 5.8: STD profile comparison between WT E-cadherin and mutant forms.

5.1.2 Statistical analysis

In this section, Mann-Whitney U test are performed to the features selected to characterize GEFM mean profiles. The purpose of this statistical analysis is to understand the ability of these features to discriminate between WT E-cadherin and selected HDGC related mutant forms. Therefore, Mann-Whitney U tests were performed comparing E-cadherin mutant forms results withWT. The results of

the tests performed are displayed in Table 5.5

Table 5.5: Mean Profile Statistical Analysis - Mann-Whitney U test (p-value)

	Membrane's Intensity	MMR	Mean Intensity	Intensity STD	Total Variation	MI	Position of MI
	WT						
T340A	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.016 (*)
A634V	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
R749W	< 0.0001	< 0.0001	0.022(*)	0.114	0.090	0.849	0.442
E757K (1)	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.208
E757K (2)	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.0061 (*)
E781D (1)	< 0.0001	< 0.0001	0.096	< 0.0001	< 0.0001	< 0.0001	0.072
E781D (2)	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
P799R	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
V832M	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

(*)This p-value supports the alternative hypothesis ($p < 0.05$).

The statistical analysis show that both Membrane's intensity and MMR discriminate perfectly between mutant forms and WT E-cadherin. On the other hand, Intensity STD, Total Variation and MI features are unable to correctly discriminate between WT and R749W E-cadherin. Furthermore, mean intensity feature not only fails distinguishing WT and E781D (2) but also have a high p-value comparing WT and R749W E-cadherin. Finally, the less discriminative feature, in coherence with the kmeans analysis, is Position of the MI. In this feature the U-test not only fails in three tests (R749W, E757K(1) and 781D(1)) but also obtains a high $p - value$ in two tests (T340A and E781D(1)).

To visually corroborate the statistical analysis, Fig. 5.9 displays simultaneously the WT mean profile and mutant forms mean profile.

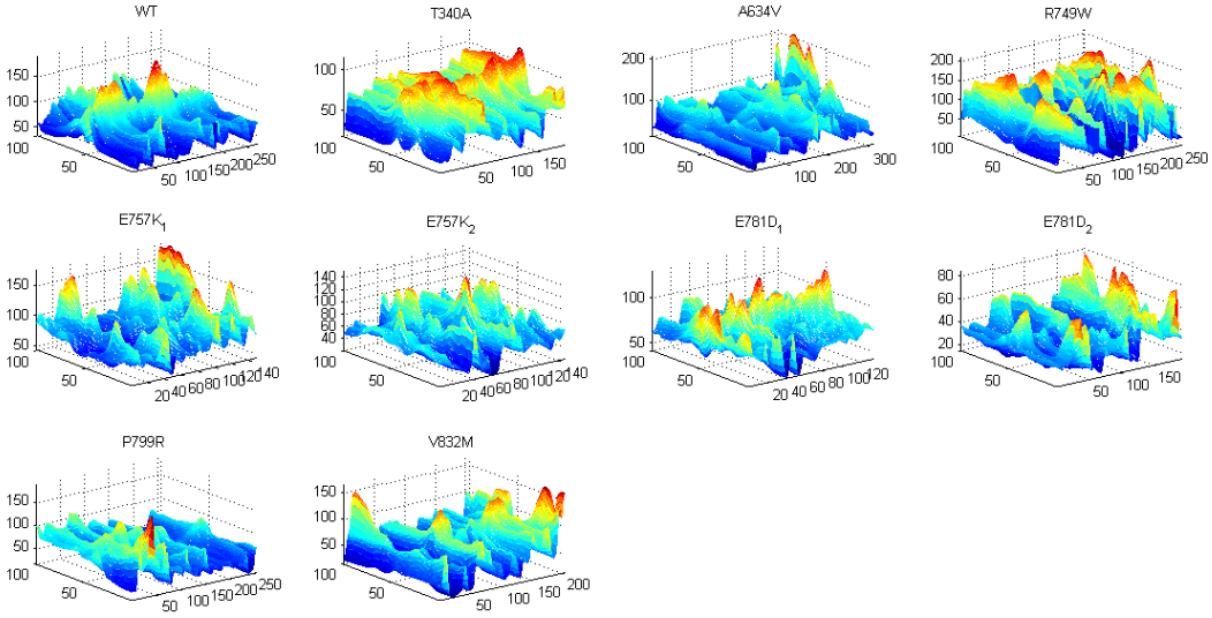


Figure 5.9: WT E-cadherin and mutant forms map of mean profiles.

5.2 Radial Profiles Characterization

The intensity behavior of the maps resulting after the geometric compensation is described through mean intensity and STD profiles. Both these profiles were collected in lines direction, to study E-cadherin molecules distribution concerning the *nucleus* center involved in the process of collection. To study a possible phenotype-mutation relation: all the maps containing the same kind of E-cadherin were displayed together and GEFM combining INMs of the same form of E-cadherin were developed. These profiles are displayed in Fig. 5.1 and Fig.5.2.

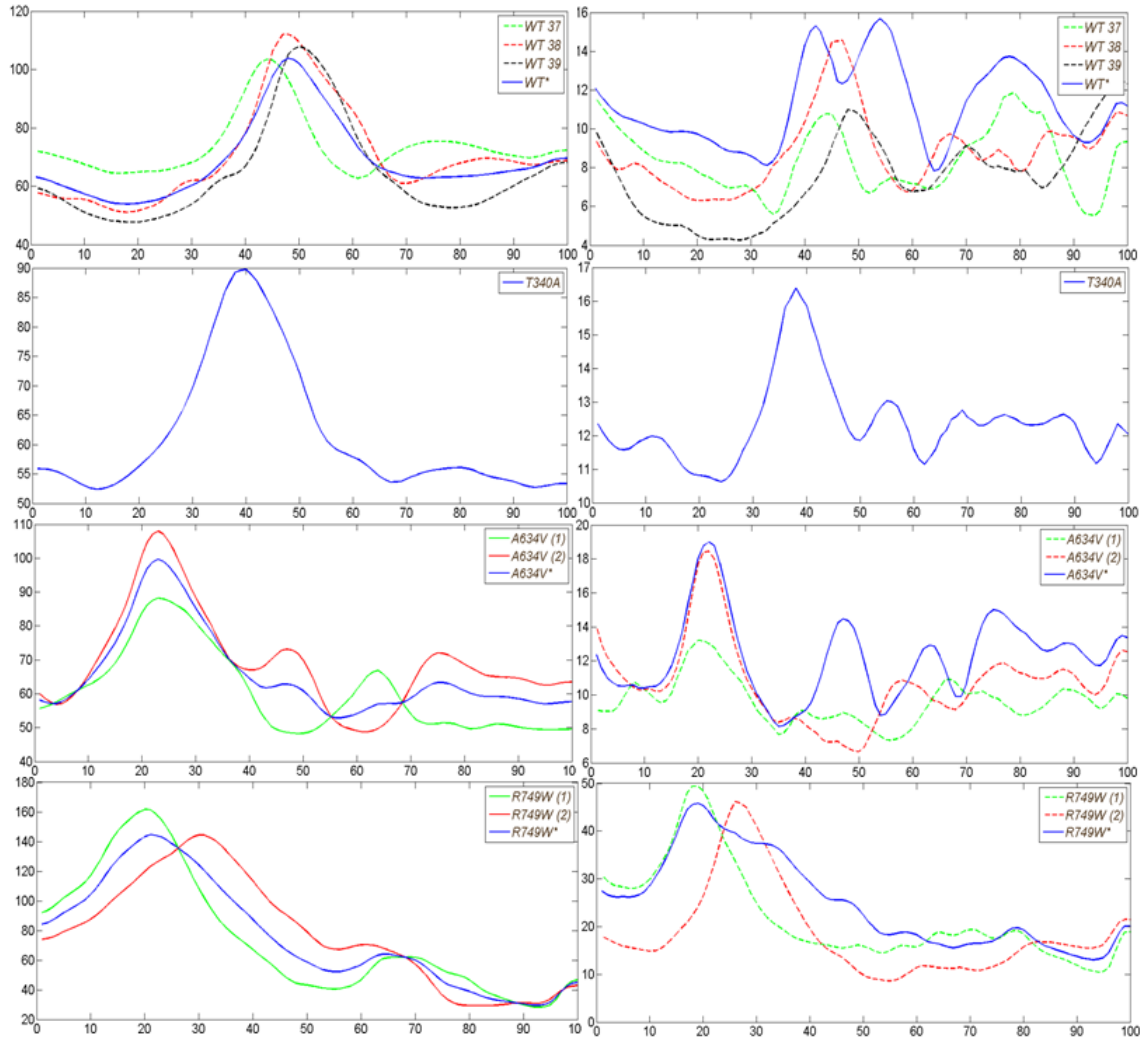


Figure 5.10: Comparison of the RDMs concerning each form of E-cadherin, in this case WT, T340A, A634V, R749W. (left) horizontal mean profiles, (right) horizontal standard deviation

The RDMs of WT, A634V, R749W, E757K display, among them, mean profiles with similar maximum intensity and shape. In addition, not only these maps STD are similar but also the GEFM STD has the same order of magnitude. Thus, to simplify further analysis, in these cases, only the GEFM be considered in analysis from henceforth. On the other hand, E781D, P799R RDMs display not only a dissimilar horizontal mean profile but also an increase in magnitude on GEFM STD profile. To avoid inaccurate judgments both these maps will be evaluated separatly as: E781D (1) or (63), E781D (2) or (64), P799R (1) or (65) and P799R (2) or (67). Finally, T340A and V832M only INMs is threated as

GEFM.

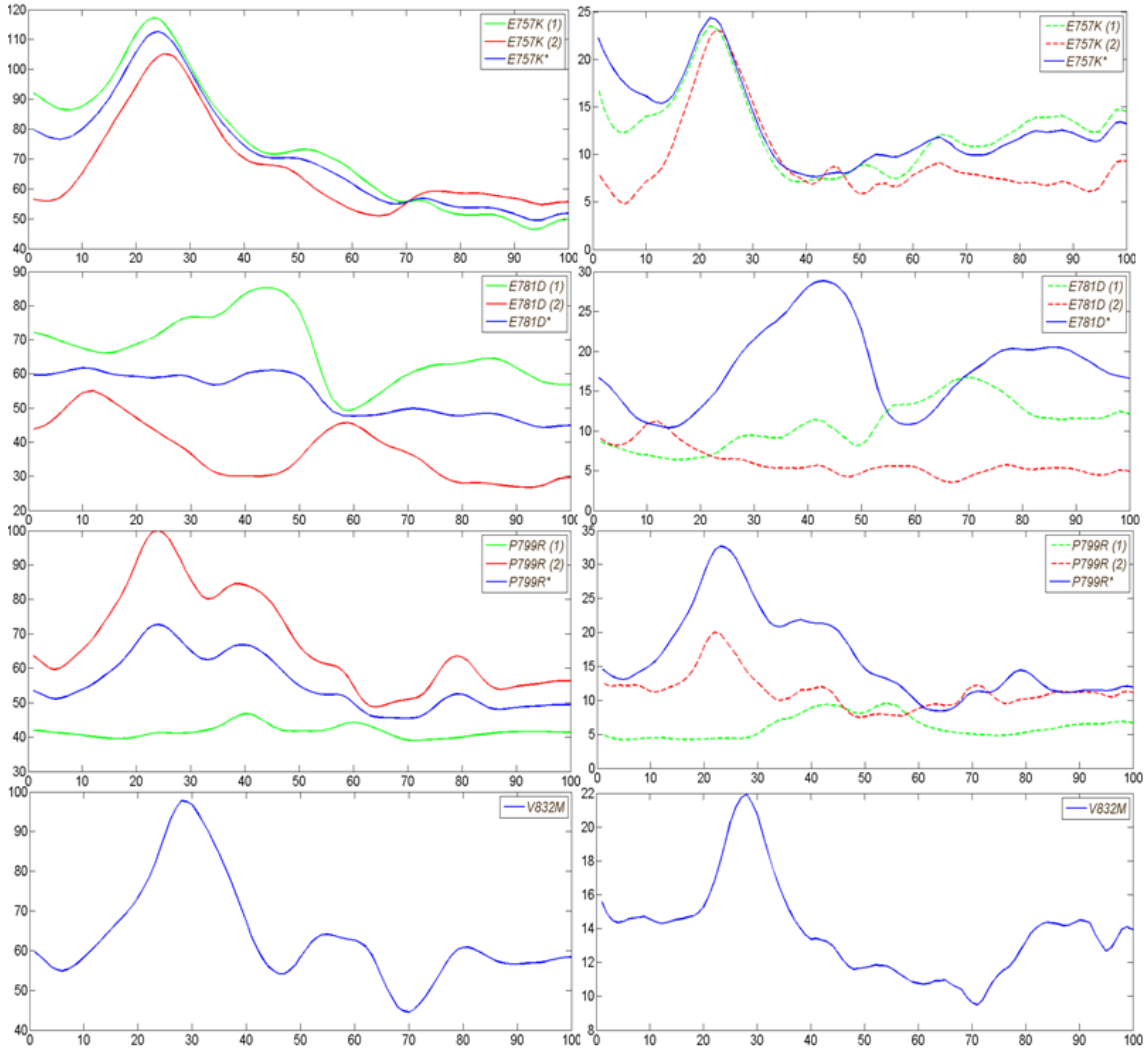


Figure 5.11: Comparison of the RDMs concerning each form of E-cadherin, in this case E757K, E781D, P799R, V832M. (left) horizontal mean profiles, (right) horizontal standard deviation

The maps selected after this mutation behavior analysis are displayed in Fig.5.12.

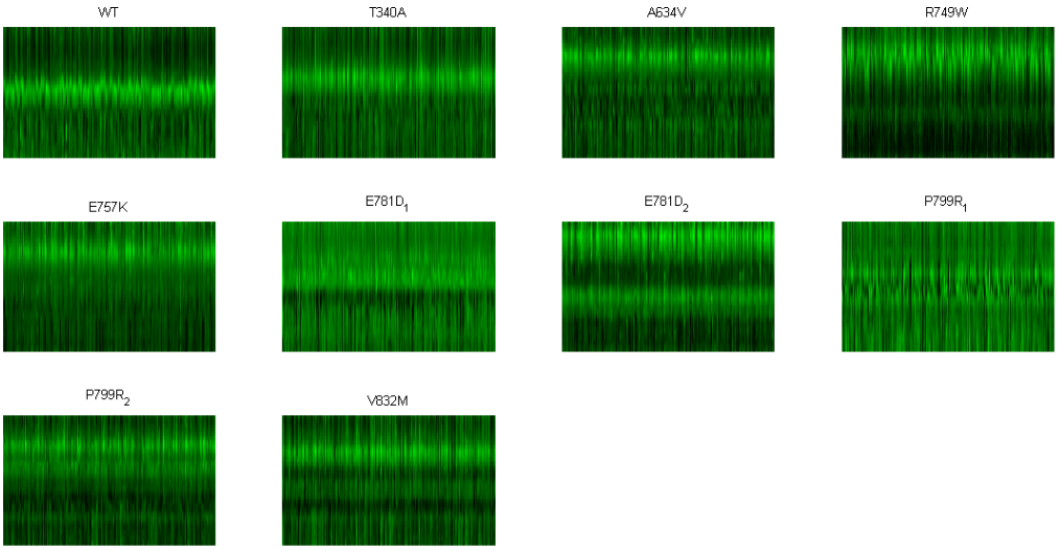


Figure 5.12: Geometrically compensated RDMs selected.

The molecules distribution, in each GEFM, is described through the mean projection and respective STD interval in each point in Fig.5.13.

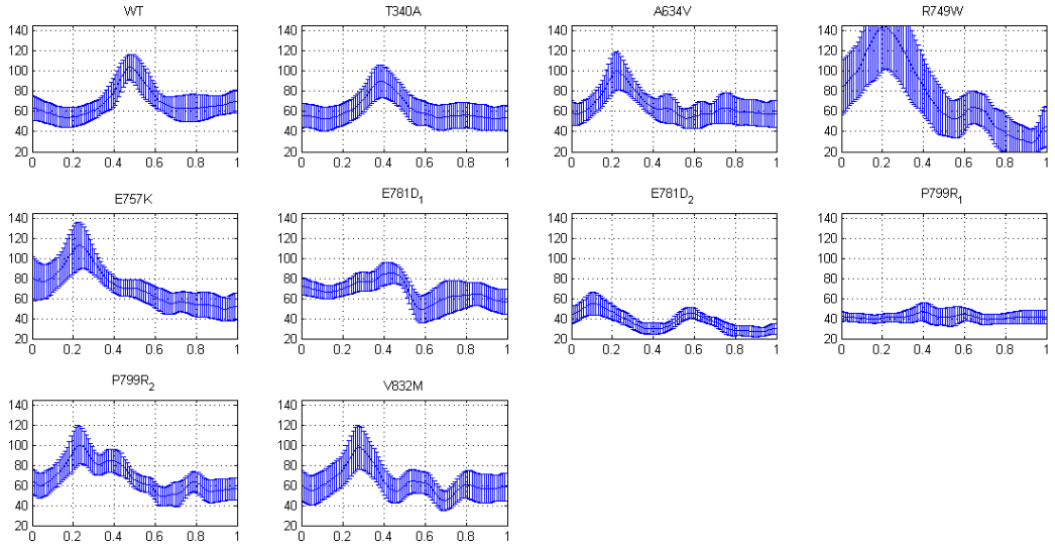


Figure 5.13: Horizontal mean profile combined with respective standard deviation interval in each point.

As can be seen in Fig.5.13, WT E-cadherin has a quite different phenotype either qualitatively or quantitatively when compared with remaining mutant forms, except for T340A. In the particular case of T340A, the profiles are indeed similar, however with further attention is possible to see differences in both maximum intensity and high intensity relative position. For a better understanding of qualitative resemblances and dissimilarities between GEFMs, both mean and STD profiles of the maps selected were computed and are displayed in Fig. 5.14 and Fig.5.15.

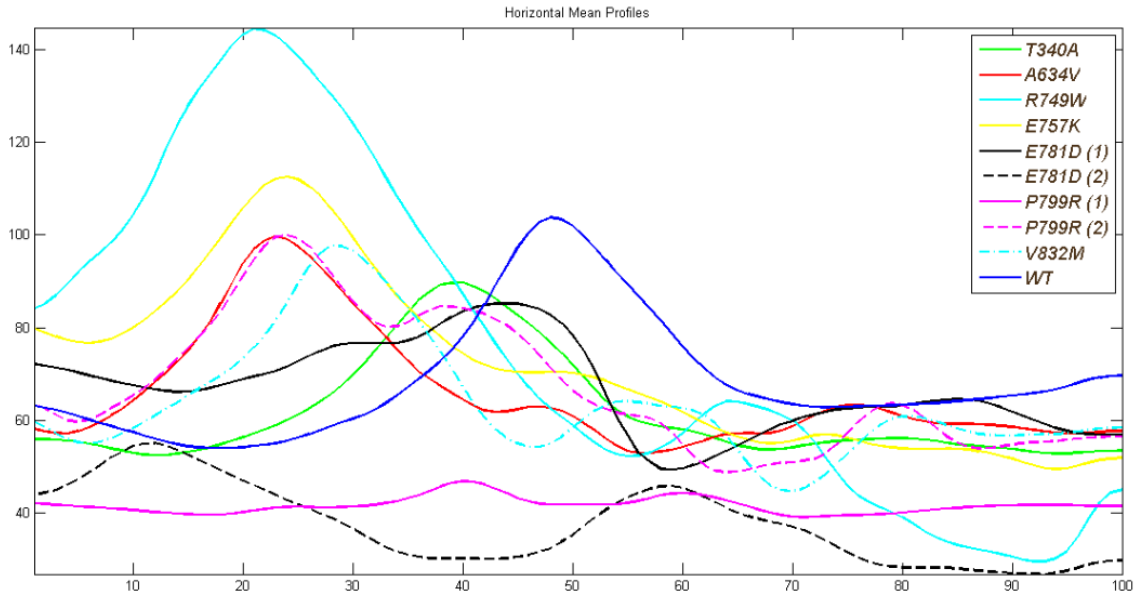


Figure 5.14: Distinct horizontal mean profiles in the WT and mutant cells after GC.

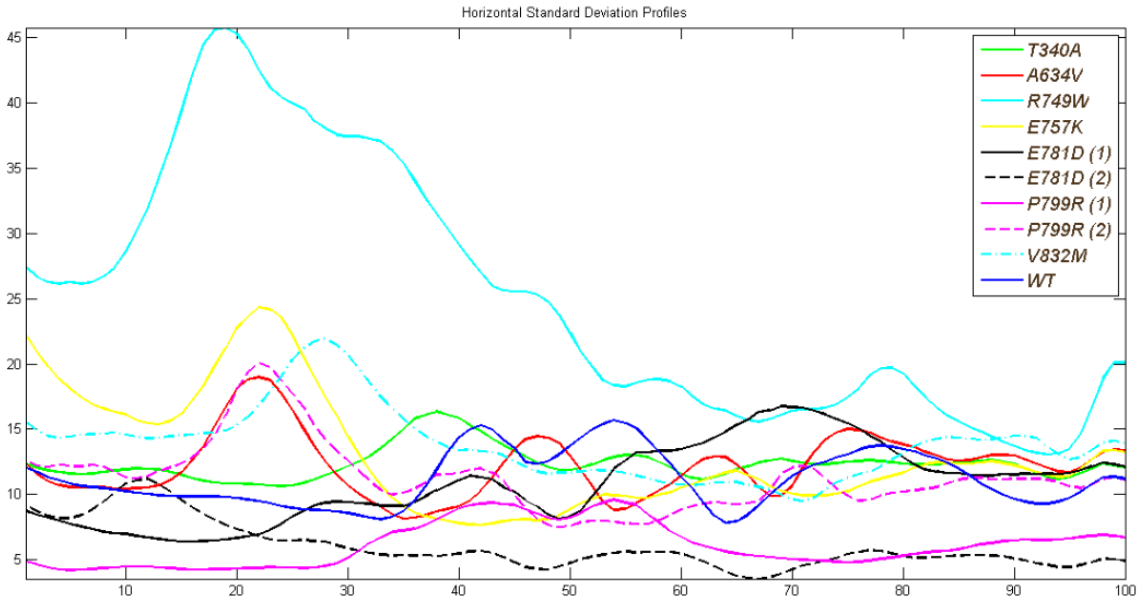


Figure 5.15: Distinct Horizontal Standard Deviation profiles in the WT and mutant cells after GC.

Fig.5.5 show that WT profile is not easy to distinguish among the remaining mutant forms, neither in intensity magnitude or is shape. However, when considering the position of maximum concentration of molecules, WT displays a value considerably different. In order to quantitatively characterize the distribution both at membrane and cytoplasmic level, and distinguish the E-cadherin expression level for the different types of E-cadherin, objective features were extracted. This results are shown in the next subsection.

A last visual analysis was performed. Cell maps were created from the geometrically compensated RDMs selected, simulating for each type of E-cadherin the cell environment. The cell profiles, reconstructed using the process described in Fig. 3.14, are displayed in Fig. 5.16.

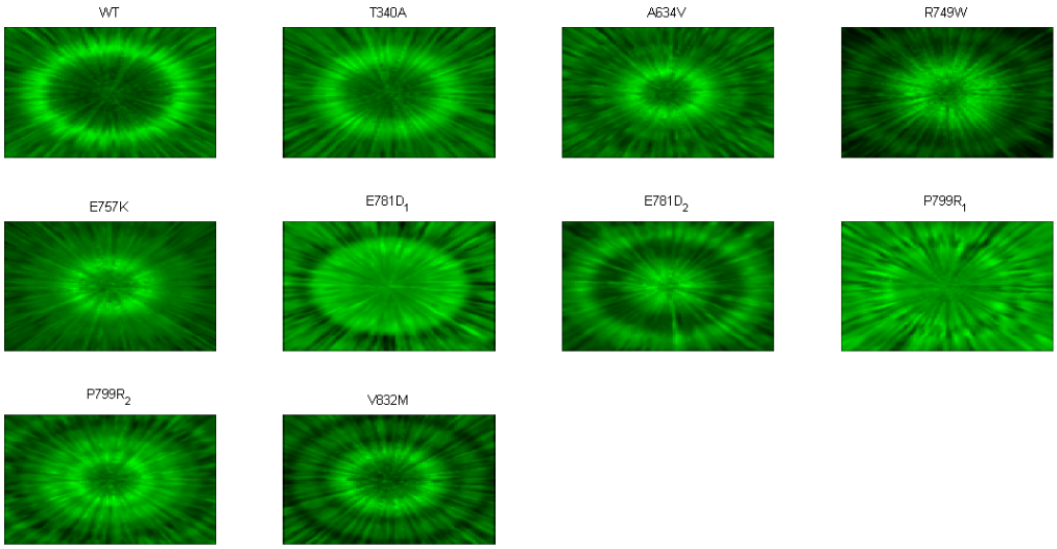


Figure 5.16: Cell profile reconstruction obtained from each one of the GEFM.

These cell profiles suggest significant qualitative and quantitative differences between WT E-cadherin and remaining mutant forms phenotype.

5.2.1 Feature Collection

To characterize and distinguish E-cadherin expression levels in final RD GEFMs, objective features were extracted from mean and STD profiles. The features collected from these mean vector were: Maximum Intensity (MI), position of the MI, intensity at the membrane, mean intensity, intensity standard deviation, Maximum Mean Ratio (MMR) and Total Variation (TV). The quantitative results of feature collection are displayed in the Table 5.6

Table 5.6: Features collected from RDGEFMs' mean profiles.

Mean Profile								
		Maximum Intensity (MI)	Position of the MI	Membrane's intensity	Mean Intensity	Intensity STD	Maximum Mean Ratio (MMR)	Total Variation (TV)
WT	1	103.82	0.48	101.92	68.62	13.243	1.513	15.18
T340A	2	89.73	0.40	71.99	61.53	11.27	1.4583	11.81
A634V	3	99.73	0.23	60.80	65.70	12.51	1.52	15.08
R749W	4	144.64	0.21	58.81	76.2	36.09	1.90	25.05
E757K	5	112.64	0.24	70.18	71.61	18.58	1.573	14.693
E781D(1)	6	85.18	0.45	78.42	66.96	9.57	1.27	12.99
E781D(2)	7	55.02	0.12	35.18	37.42	8.56	1.47	8.95
P799R(1)	8	46.73	0.40	41.77	41.57	1.72	1.12	3.6354
P799R(2)	9	99.92	0.24	66.32	67.40	14.42	1.4824	16.524
V832M	10	97.73	0.28	58.31	63.59	12.80	1.54	18.10

For a better understanding of quantitative reasons of similarity and dissimilarity within these features, a k-means clustering analysis was performed. In this analysis each element display one color representing the cluster it belongs. Therefore, elements contained in the same cluster display the same colors and consequently, elements of different cluster display different colors. The results of this evaluation are displayed in Table 5.7.

Table 5.7: Features collected from RDGEFMs' mean profiles after a kmeans (4 clusters) analysis.

Mean Profile								
		Maximum Intensity (MI)	Position of the MI	Membrane's intensity	Mean Intensity	Intensity STD	Maximum Mean Ratio (MMR)	Total Variation (TV)
WT	1	103.82	0.48	101.92	68.62	13.243	1.513	15.18
T340A	2	89.73	0.40	71.99	61.53	11.27	1.4583	11.81
A634V	3	99.73	0.23	60.80	65.70	12.51	1.52	15.08
R749W	4	144.64	0.21	58.81	76.2	36.09	1.90	25.05
E757K	5	112.64	0.24	70.18	71.61	18.58	1.573	14.693
E781D (1)	6	85.18	0.45	78.42	66.96	9.57	1.27	12.99
E781D (2)	7	55.02	0.12	35.18	37.42	8.56	1.47	8.95
P799R(1)	8	46.73	0.40	41.77	41.57	1.72	1.12	3.6354
P799R(2)	9	99.92	0.24	66.32	67.40	14.42	1.4824	16.524
V832M	10	97.73	0.28	58.31	63.59	12.80	1.54	18.10

According with Table 5.7, the most discriminant feature is Membrane's Intensity, where WT E-cadherin cluster has only 1 element. In Position of the MI and mean intensity features, WT E-cadherin cluster has 4 elements. In remaining features- MI,Intensity STD, MMR and TV- WT E-cadherin cluster has at least 5 elements. A further statistical analysis (Mann-Whitney U test) is performed to these features to understand the feature behavior in each profile constituent of these RD GEFMs. This analysis is present in Section 5.2.1.A. To visually corroborate the k-means analysis, Fig. 5.17 displays simultaneously the WT mean profile and mutant forms mean profile.

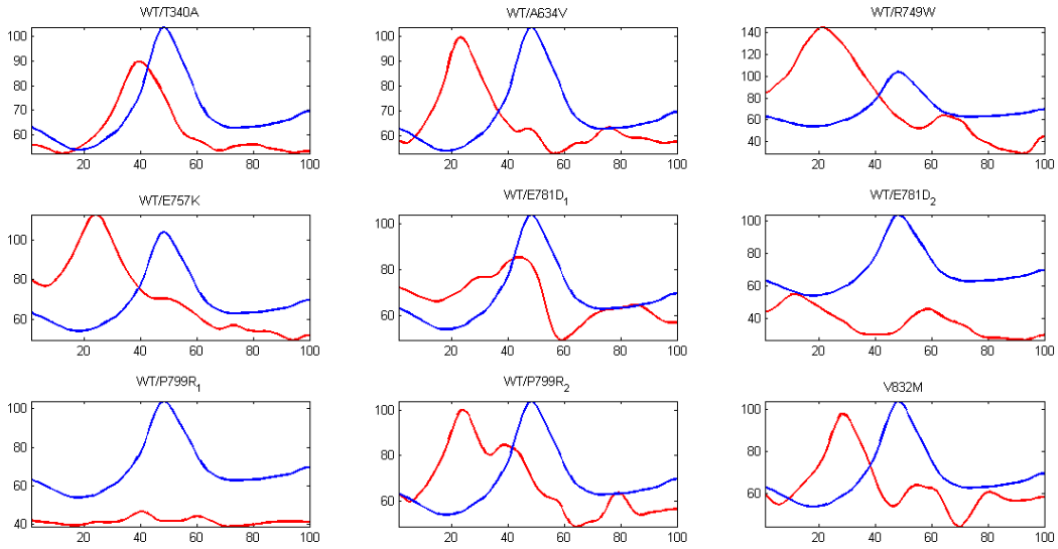


Figure 5.17: STD profile comparison between WT E-cadherin and mutant forms.

Finally, the features collected from STD profiles were: Maximum Standard Deviation (MSTD), position of the MSTD, STD at the membrane, mean STD, Maximum Mean STD Ratio, Total Variation (TV). The results of the feature collection are shown in the Table 5.8

Table 5.8: Features collected from GEFMs' STD profiles.

Variance Profile							
		MSTD	Position MSTD	Membrane's STD	Mean STD	Maximum Mean STD Ratio	Total Variation (TV) (TV)
WT	1	15.68	0.54	13.72	11.20	1.40	5.01
T340A	2	16.36	0.38	11.85	12.36	1.32	2.79
A634V	3	18.99	0.22	12.99	12.21	1.55	6.65
R749W	4	45.78	0.19	22.36	25.03	1.83	9.52
E757K	5	24.37	0.22	8.85	13.03	1.87	6.44
E781D (1)	6	16.75	0.70	8.26	10.81	1.55	3.55
E781D (2)	7	11.21	0.12	4.75	6.00	1.87	2.60
P799R (1)	8	9.61	0.54	8.21	6.11	1.57	2.33
P799R(2)	9	20.07	0.22	7.59	11.32	1.77	5.55
V832M	10	21.93	0.28	11.70	13.84	1.58	4.43

A similar cluster analysis to study the quantitative profiles resemblances was performed and the results are displayed in the Table 5.9.

Table 5.9: Features collected from GEFMs' STD profiles after a k-means analysis

Variance Profile							
		MSTD	Position MSTD	Membrane's STD	Mean STD	Maximum Mean STD Ratio	Total Variation (TV)
WT	1	15.68	0.54	13.72	11.20	1.40	5.01
T340A	2	16.36	0.38	11.85	12.36	1.32	2.79
A634V	3	18.99	0.22	12.99	12.21	1.55	6.65
R749W	4	45.78	0.19	22.36	25.03	1.83	9.52
E757K	5	24.37	0.22	8.85	13.03	1.87	6.44
E781D(1)	6	16.75	0.70	8.26	10.81	1.55	3.55
E781D(2)	7	11.21	0.12	4.75	6.00	1.87	2.60
P799R (1)	8	9.61	0.54	8.21	6.11	1.57	2.33
P799R (2)	9	20.07	0.22	7.59	11.32	1.77	5.55
V832M	10	21.93	0.28	11.70	13.84	1.58	4.43

Table 5.9, show that for k-means algorithm the most discriminant feature is Maximum Mean STD Ratio , where WT E-cadherin cluster has 2 element. Followed by Total Variation (TV) and Position of MSTD. In these features WT E-cadherin cluster has exactly 3 element. However a close analysis of the Position of MSTD shows that WT E-cadherin and the mutant form P799R (1) display the same value suggesting this feature inability to discriminate between WT and P799R (1). In the remaining features - MSTD, Membrane's STD and Mean STD- WT E-cadherin cluster has at least 4 elements, suggesting they are not good discriminative features. To visually corroborate the k-means analysis, Fig. 5.18 displays simultaneously the WT STD profile and mutant forms STD profile.

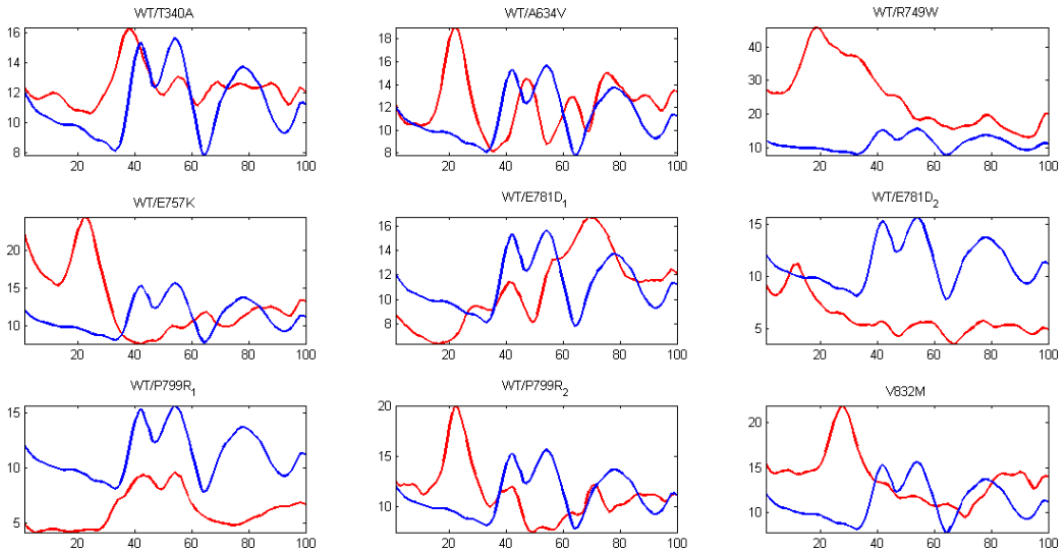


Figure 5.18: Mean profile comparison between WT E-cadherin and mutant forms.

5.2.1.A Statistical analysis

In this section, Mann-Whitney U test are performed to the features selected to characterize RDGEFM mean profiles. Since the purpose of this statistical analysis is to understand the ability of these features to discriminate between WT E-cadherin and selected HDGC related mutant forms, the test were developed concerning WT maps. The Mann-Whitney U tests were performed comparing E-cadherin mutant forms results withWT.The results of the tests performed are displayed in Table 5.10.

Table 5.10: Mean Profile Statistical Analysis - Mann-Whitney U test (p-value)

	Membrane's Intensity	Mean Intensity	Position of MI	MI	Total Variation (TV)	Intensity STD	Maximum Mean Ratio (MMR)
	WT						
T340A	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
A634V	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.019(*)	< 0.0001
R749W	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
E757K	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.055	0.008(*)	0.079
E781D (1)	< 0.0001	0.054	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
E781D (2)	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
P799R(1)	< 0.0001	< 0.0001	0.057	< 0.0001	< 0.0001	< 0.0001	< 0.0001
P799R (2)	< 0.0001	0.043(*)	< 0.0001	< 0.0001	< 0.0001	0.397	< 0.0001
V832M	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.130	0.048 (*)

(*)This p-value still supports the alternative hypothesis ($p < 0.05$).

The statistical analysis show that both Membrane's intensity and MI discriminate perfectly between mutant forms and WT E-cadherin. In addition, Position of the MI only fails discriminating between WT and P799R(1). Therefore if P799R(1) is an abnormal image and is proven that it should be discarded, then Position of MI is a candidate for a outstanding discriminative feature. Also, TV ,Mean Intensity and MMR all fail in one test. TV fails to distinguish between WT E-cadherin and E757K. Mean intensity and MMR fail when comparing WT/E781D (1) E-cadherin and acWT/E757K, respectively. Each one of this last features also obtain a high p-values in one U-test. Mean intensity is when testing WT/ P799R(2) and MMR is in test WT/V832M. Finnally, the less discriminative feature, in coherence with the k-means is Intensity STD. In this feature the U-test not only fails in two tests (P799R(2) and

V832M) but also obtains a high $p - value$ in two tests (A634V and E757K).

To visually corroborate the statistical analysis, Fig. 5.19 displays simultaneously the WT mean profile and mutant forms mean profile.

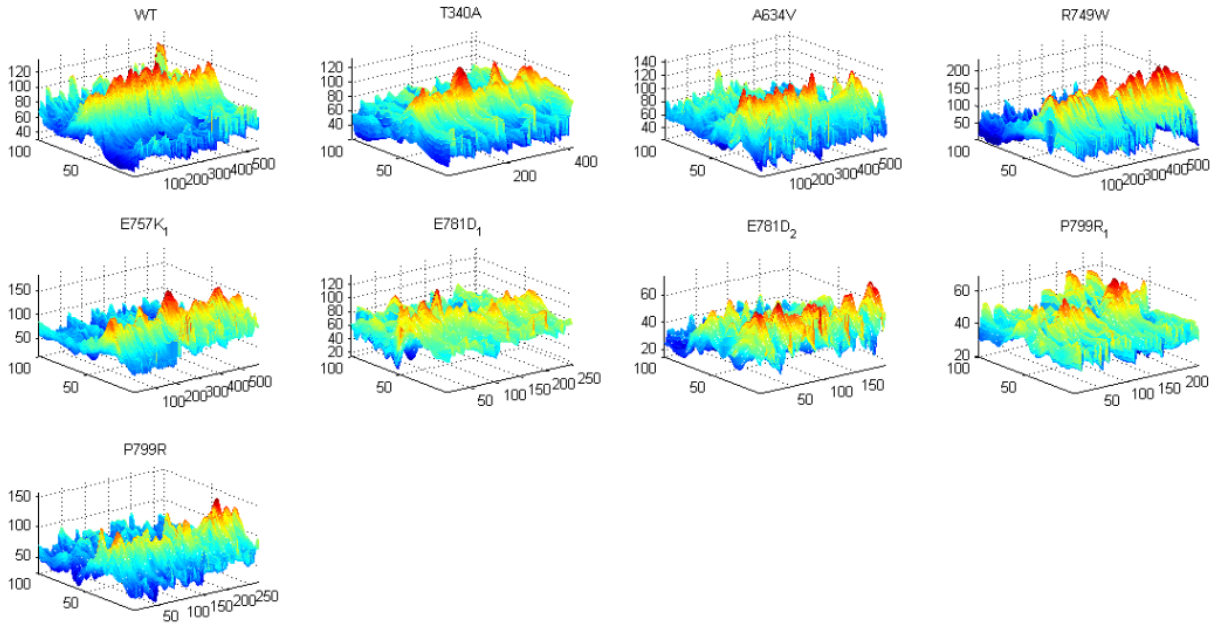


Figure 5.19: WT E-cadherin and mutant forms map of mean profiles.

6

Conclusions and Future Work

Contents

6.1 Conclusions

Throughout this dissertation several conclusions were drawn, namely about E-cadherin molecular distribution within the cell and GC algorithm stability and efficiency.

Initially, *in vitro* assays were developed to study the expression patterns of WT E-cadherin and Hereditary Diffuse Gastric Cancer (HDGC) related mutant forms of E-cadherin, namely: T340A, A643V, R749W, E757K, E781D, P799R and V832M. To characterize the molecular distribution within the cell, Internuclear Maps (INMs) and Radial Maps (RDMs) were extracted from these images. These initial INMs show that WT E-cadherin has the highest concentration of molecules in the cell membrane, $x = 0.5$. E-cadherin mutant forms display profiles different from WT either in intensity or in distribution pattern. Actually, in some mutant forms, the molecules distribution behavior is quite different, namely in the number and position of high intensity peaks. Furthermore, in the initial RDMs, WT E-cadherin membrane heterogeneity is more noticeable. Fig. 3.15 evidence cell membrane fluctuations in shape and size, through the increase and decrease of membrane distance to the center of the nucleus. Consequently, WT E-cadherin radial distribution profile is more difficult to distinguish.

Afterwards, the profiles were processed by a Geometric Compensation (GC) algorithm to correct in intensity abnormalities associated with cell membrane heterogeneities. Synthetic data were developed to simulate the distortions of the membrane, expected in IN and RD maps. The GC algorithm synthetic results show, not only the stability of the algorithm for different geometric conformations, but also that the algorithm correctly aligns the synthetic observations. Furthermore, the GC algorithm real data analysis shows the stability of the method for FM images. The results of the real data show an alignment of the observations, reflected in STD significant decrease after GC, specially of the high intensity observations. This enhancements is greater in RDMs, as suggested by Fig. 3.23 and 3.25 that better simulate IN and RD maps, respectively. Specially, in RDMs there is a rearrangement of the information contained in the map evidencing the geometrically corrected cell membrane. This suggest that the GC algorithm, improves not only the quantification of E-cadherin present in cell membrane but also the exact position of E-cadherin molecules distribution, which can be in the membrane or in the cytoplasm near the nucleus.

The last part of this dissertation consist on the characterization, after the geometric compensation, of the molecular distribution within the cell. The characterization of E-cadherin distribution was done through the collection of objective quantitative features from the mean and STD profiles of the IN and RDs maps built. To study the quantitative differences among profiles a statistical analysis was developed.

From the feature collection and statistical analysis of INMs mean profiles, it can be concluded that: the better discriminative features are Membrane's Intensity and MMR, with an outstanding performance; intensity STD, TV and MI are able to distinguish WT E-cadherin from the remaining mutant forms, with the exception of R749W; WT E-cadherin has the higher MI; mean intensity feature, although presenting a high p-value when comparing WT with R749W, discriminate between WT and remaining mutant forms except for E781D(1); the position of the MI is a poor discriminative feature,

unable to distinguish WT E-cadherin from R749W, E757K(1) and E781D(1).The feature collection of INMsSTD profiles and k-means clustering analysis, suggest Maximum Mean STD Ratio, Membrane's STD and Position of MSTD as potential interesting discriminative features. The results of RDMs mean profiles feature collection and statistical analysis of mean profiles show that: Membrane's intensity and MI have a perfect performance discriminating WT E-cadherin among mutant forms; position of the MI has potential to be a good feature, if image P799R(1) is proven to be ruled out; TV and MMR are able to distinguish WT E-cadherin from the remaining mutant forms, with the exception of E757K; MMR display a high p-value when testing WT and V832M E-cadherin; mean intensity feature, although presenting a high p-value when comparing WT with P799R(2), discriminate between WT and remaining mutant forms except for E781D(1); the intensity STD is a poor discriminative feature, unable to distinguish WT E-cadherin from P799R(2) and V832M. The feature collection of RDMsSTD profiles and k-means clustering analysis, suggest Maximum Mean STD Ratio, Position of MSTD and TV of the STD profile as potential discriminative features.

An overall analysis of the features show that the very best features discriminating WT from the selected HDGC related mutant forms are: Membrane's Intensity, in both IN and RD maps; MMR, in INMs and MI, in RDMs.

This dissertation presents a way to characterize E-cadherin distribution within the cell and suggest some features to be used in further classification systems. In some of the selected HDGC related E-cadherin forms, the phenotype expected is not well know yet. The cases were the functional impact of the mutation is known, like WT E-cadherin, R749W, E757K and finally V832M, validate the results of this work. Functional E-Cadherin is synthesized at the Golgi, transported and stabilized at the membrane and continuously recycled. Therefore, an equidistant membrane together with an increase in intensity values are traits of a healthy linkage between cells which mean presence of WT E-cadherin. Also, the constant E-cadherin distribution in the cytoplasm, suggest the absence of abnormal E-cadherin in degradation. This work results support the literature as WT E-cadherin profiles, both in IN and RD analysis, display a single well define high intensity peak in position $x = 0.5$ and constant values elsewhere. Also, as described in section 2.3, R749W and E757K mutant forms result from a juxta-membrane region mutation that prevent the binding with p120 -catenin (p120ctn). This means that not only the E-cadherin is unable to stabilize in the membrane but also that these mutant forms are recognized by Endoplasmic Reticulum Associated Degradation (ERAD). In these images both the intensity at the membrane is lower than WT and some high intensity peaks appear near the nucleus, due to the increase in E-cadherin molecules in degradation path. Finally in V832M particular case, the gene mutation interferes with E-cadherin link to $\beta - catenin$, decreasing the number of E-cadherin molecules in the membrane, since its transport is compromised. These facts are depicted in the mean profiles of both the IN and RD profiles of these mutations and quantified in the features collection results.

6.2 Future Work

The work developed in this dissertation can still be improved, namely in the correct characterize of some E-cadherin mutant forms distribution. In both analysis (IN and RD), due to the insufficient number of images processed, some E-cadherin mutant forms distribution profiles were not clear. In this thesis, the number of images processed was limited by the number of *in vitro* assays developed. To improve and validate the biological results obtained, the number of images processed by this method has to be increased.

Furthermore, to create a Computer Aid Diagnosis (CAD) framework for semi-automatic detection of dysfunctional proteins in clinical practice, for both screening and diagnosis, a classification system has to be studied in further detail. This classification system developed should not only distinguish WT E-cadherin and these HDGC related mutant forms, but also discriminate among mutant forms. An useful approach would be to improve the features selection adding some morphological features, using image local decomposition e.g. wavelet decomposition.

In respect to the geometrical compensation algorithm, although in this study it was used exclusively in FM images, there is no mathematical evidence that this algorithm is not stable in other kinds of imaging images. A further study of this algorithm behavior and stability in different types of images can be done.

Finally, in some cases like R749W and E757K E-cadherin mutant forms, recent results suggest that a chaperon treatment would be able to restore the functionality of the E-cadherin molecules. A qualitative and quantitative comparison between WT E-cadherin and threated R749W and E757K E-cadherin IN and RD distribution profiles, would be interesting to understand of the validity of this method.

Bibliography

[1] J. Paredes, J. Figueiredo, A. Albergaria, P. Oliveira, J. Carvalho, A. S. Ribeiro, J. Caldeira, A. M. Costa, J. Simões Correia, J. O. Oliveira, H. Pinheiro, S. S. Pinho, R. Mateus, C. A. Reis, M. Leite, M. S. Fernandes, F. Schmitt, F. Carneiro, C. Figueiredo, C. Oliveira, and R. Seruca, "Epithelial e- and p-cadherins: Role and clinical significance in cancer," *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, vol. 1826, no. 2, pp. 297–311, December 2012.

[2] B. M. Gumbiner, "Regulation of cadherin-mediated adhesion in morphogenesis," *Nature reviews Molecular cell biology*, vol. 6, no. 8, pp. 622–634, 2005.

[3] Q. Wu, F. Merchant, and K. Castleman, *Microscope image processing*. Academic press, 2010.

[4] F. V. Roy, Ed., *Progress in Molecular Biology and Translational Science : The Molecular Biology of Cadherins*. Academic Press, 2013, vol. 116.

[5] R. M. Cisco, J. M. Ford, and J. A. Norton, "Hereditary diffuse gastric cancer," *Cancer*, vol. 113, no. S7, pp. 1850–1856, 2008.

[6] G. Suriano, S. Yew, P. Ferreira, J. Senz, P. Kaurah, J. M. Ford, T. A. Longacre, J. A. Norton, N. Chun, S. Young *et al.*, "Characterization of a recurrent germ line mutation of the e-cadherin gene: implications for genetic testing and clinical management," *Clinical Cancer Research*, vol. 11, no. 15, pp. 5401–5409, 2005.

[7] P. Guilford, J. Hopkins, J. Harraway, M. McLeod, N. McLeod, P. Harawira, H. Taite, R. Scoular, A. Miller, and A. E. Reeve, "E-cadherin germline mutations in familial gastric cancer," *Nature*, vol. 392, no. 6674, pp. 402–405, 1998.

[8] M. Takeichi, "Functional correlation between cell adhesive properties and some cell surface proteins." *The Journal of Cell Biology*, vol. 75, no. 2, pp. 464–474, 1977.

[9] A. Nagafuchi and M. Takeichi, "Cell binding function of e-cadherin is regulated by the cytoplasmic domain." *The EMBO Journal*, vol. 7, no. 12, p. 3679, 1988.

[10] K. Vleminckx, L. Vakaet Jr, M. Mareel, W. Fiers, and F. Van Roy, "Genetic manipulation of e-cadherin expression by epithelial tumor cells reveals an invasion suppressor role," *cell*, vol. 66, no. 1, pp. 107–119, 1991.

[11] G. Handschuh, S. Candidus, B. Luber, U. Reich, C. Schott, S. Oswald, H. Becke, P. Hutzler, W. Birchmeier, H. Hoefler *et al.*, "Tumour-associated e-cadherin mutations alter cellular morphology, decrease cellular adhesion and increase cellular motility." *Oncogene*, vol. 18, no. 30, p. 4301, 1999.

[12] B. Mayer, J. P. Johnson, F. Leitl, K. W. Jauch, M. M. Heiss, F. W. Schildberg, W. Birchmeier, and I. Funke, "E-cadherin expression in primary and metastatic gastric cancer: down-regulation correlates with cellular dedifferentiation and glandular disintegration," *Cancer research*, vol. 53, no. 7, pp. 1690–1695, 1993.

[13] P. D. Pharoah, P. Guilford, C. Caldas *et al.*, "Incidence of gastric cancer and breast cancer in cdh1 (e-cadherin) mutation carriers from hereditary diffuse gastric cancer families," *Gastroenterology*, vol. 121, no. 6, pp. 1348–1353, 2001.

[14] A. Brooks-Wilson, P. Kaurah, G. Suriano, S. Leach, J. Senz, N. Grehan, Y. Butterfield, J. Jeyes, J. Schinas, J. Bacani *et al.*, "Germline e-cadherin mutations in hereditary diffuse gastric cancer: assessment of 42 new families and review of genetic screening criteria," *Journal of medical genetics*, vol. 41, no. 7, pp. 508–517, 2004.

[15] G. Suriano, M. J. Oliveira, D. Huntsman, A. R. Mateus, P. Ferreira, F. Casares, C. Oliveira, F. Carneiro, J. C. Machado, M. Mareel *et al.*, "E-cadherin germline missense mutations and cell phenotype: evidence for the independence of cell invasion on the motile capabilities of the cells," *Human molecular genetics*, vol. 12, no. 22, pp. 3007–3016, 2003.

[16] G. Suriano, S. Seixas, J. Rocha, and R. Seruca, "A model to infer the pathogenic significance of cdh1 germline missense variants," *Journal of Molecular Medicine*, vol. 84, no. 12, pp. 1023–1031, 2006.

[17] J. Simões-Correia, J. Figueiredo, R. Lopes, F. Stricher, C. Oliveira, L. Serrano, and R. Seruca, "E-cadherin destabilization accounts for the pathogenicity of missense mutations in hereditary diffuse gastric cancer," *PLoS One*, vol. 7, no. 3, p. e33783, 2012.

[18] G. Suriano, C. Oliveira, P. Ferreira, J. C. Machado, M. C. Bordin, O. De Wever, E. A. Bruyneel, N. Moguilevsky, N. Grehan, T. R. Porter *et al.*, "Identification of cdh1 germline missense mutations associated with functional inactivation of the e-cadherin protein in young gastric cancer probands," *Human molecular genetics*, vol. 12, no. 5, pp. 575–582, 2003.

[19] A. Tang, M. S. Eller, M. Hara, M. Yaar, S. Hirohashi, and B. A. Gilchrest, "E-cadherin is the major mediator of human melanocyte adhesion to keratinocytes in vitro," *Journal of cell science*, vol. 107, no. 4, pp. 983–992, 1994.

[20] H. L. Spencer, A. M. Eastham, C. L. Merry, T. D. Southgate, F. Perez-Campo, F. Soncin, S. Ritson, R. Kemler, P. L. Stern, and C. M. Ward, "E-cadherin inhibits cell surface localization of the pro-migratory 5t4 oncofetal antigen in mouse embryonic stem cells," *Molecular biology of the cell*, vol. 18, no. 8, pp. 2838–2851, 2007.

[21] P. Ferreira, M. J. Oliveira, E. Beraldi, A. R. Mateus, T. Nakajima, M. Gleave, J. Yokota, F. Carneiro, D. Huntsman, R. Seruca *et al.*, "Loss of functional e-cadherin renders cells more resistant to the apoptotic agent taxol in vitro," *Experimental cell research*, vol. 310, no. 1, pp. 99–104, 2005.

[22] J. Figueiredo, J. Simões-Correia, O. Söderberg, G. Suriano, and R. Seruca, "Adp-ribosylation factor 6 mediates e-cadherin recovery by chemical chaperones," *PloS one*, vol. 6, no. 8, p. e23188, 2011.

[23] P. Dey, "Role of ancillary techniques in diagnosing and subclassifying non-hodgkin's lymphomas on fine needle aspiration cytology," *Cytopathology*, vol. 17, no. 5, pp. 275–287, 2006.

[24] E. K. Zuba-Surma, M. Kucia, A. Abdel-Latif, J. W. Lillard, and M. Z. Ratajczak, "The imagestream system: a key step to a new era in imaging." *Folia Histochemica et Cytobiologica*, vol. 45, no. 4, pp. 279–278, 2008.

[25] A. Giangreco, K. B. Jensen, Y. Takai, J. Miyoshi, and F. M. Watt, "Necl2 regulates epidermal adhesion and wound repair," *Development*, vol. 136, no. 20, pp. 3505–3514, 2009.

[26] O. Heimstädt, "The fluorescence microscope (das fluoreszenzmikroskop)," *Z Wiss Mikrosk*, vol. 28, pp. 330–337, 1911.

[27] A. Coons, H. Creech, and R. Jones, "Immunological properties of an antibody containing a vol. 47, pp. 200–202, 1941.

[28] O. Shimomura, F. H. Johnson, and Y. Saiga, "Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, aequorea," *Journal of cellular and comparative physiology*, vol. 59, no. 3, pp. 223–239, 1962.

[29] M. Chalfie, Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher, "Green fluorescent protein as a marker for gene expression," *Science*, vol. 263, no. 5148, pp. 802–805, 1994.

[30] R. Y. Tsien, "The green fluorescent protein," *Annual review of biochemistry*, vol. 67, no. 1, pp. 509–544, 1998.

[31] M. Fonseca and C. Manjunath, "Registration techniques for multisensor remotely sensed imagery," *Photogrammetric Engineering and Remote Sensing*, vol. 14, no. 1049-1056, 1996.

[32] B. Zitova and J. Flusser, "Image registration methods: a survey." *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[33] J. M. Sanches and J. S. Marques, "Joint image registration and volume reconstruction for 3d ultrasound," *Pattern Recognition Letters*, vol. 24, no. 4-5, pp. 791–800, 2003.

[34] S. Li, J. Wakefield, and J. Noble, "Automated segmentation and alignment of mitotic nuclei for kymograph visualisation," in *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, 30 2011-april 2 2011, pp. 622 –625.

[35] O. Dzyubachyk, J. Essers, W. A. v. Cappellen, C. Baldeyron, A. Inagaki, W. J. Niessen, and E. Meijering, “Automated analysis of time-lapse fluorescence microscopy images: from live cell images to intracellular foci,” *Bioinformatics*, vol. 26, no. 19, pp. 2424–2430, 2010.

[36] P. Carneiro, M. S. Fernandes, J. Figueiredo, J. Caldeira, J. Carvalho, H. Pinheiro, M. Leite, S. Melo, P. Oliveira, S. oes Correia, M. J. Oliveira, F. Carneiro, C. Figueiredo, J. Paredes, C. Oliveira, and R. Seruca, “E-cadherin dysfunction in gastric cancer - cellular consequences, clinical applications and open questions,” *FEBSLetters*, vol. 586, no. 18, pp. 2981–2989, August 2012.

[37] G. Berx, K.-F. Becker, H. Höfler, and F. Van Roy, “Mutations of the human e-cadherin (cdh1) gene,” *Human mutation*, vol. 12, no. 4, pp. 226–237, 1998.

[38] M. Takeichi, “Morphogenetic roles of classic cadherins.” *Curr Opin Cell Biol*, vol. 7, pp. 619–627, 1995.

[39] B. M. Gumbiner *et al.*, “Cell adhesion: the molecular basis of tissue architecture and morphogenesis.” *Cell*, vol. 84, no. 3, pp. 345–357, 1996.

[40] H. R. Petty, “Fluorescence microscopy: established and emerging methods, experimental strategies, and application in immunology,” *Microscopy Research and Technique*, vol. 70, pp. 687–709, 2007.

[41] H. C. Ishikawa-Ankerhold, R. Ankerhold, and G. P. Drummen, “Advanced fluorescence microscopy techniques - frap, flip, flap, fret and flim,” *Molecules*, vol. 17, no. 4, pp. 4047–4132, 2012.

[42] I. D. Odell and D. Cook, “Immunofluorescence techniques,” *Journal of Investigative Dermatology*, vol. 133, no. 1, p. e4, 2013.

[43] P. J. Robinson, J. Sturgis BS, and G. L. Kumar, “IHC Staining Methods (5th Ed), Chapter 10: Immunofluorescence, FIFtH edition,” http://www.dako.com/08002_03aug09_ihc_guidebook_5th_edition_chapter_10.pdf, pp. 62–65.

[44] L. T. Corporation. (2013, May) Alexa fluorÂ® 488 goat anti-mouse igg (h+l). [Online]. Available: http://www.invitrogen.com/1/1/10887-alexa-fluor-488-goat-anti-mouse-igg-h-l.html

[45] D. Tschumperlé. (2013, May) The cimg library. [Online]. Available: http://cimg.sourceforge.net

[46] E. Bengtsson, C. Wahlby, and J. Lindblad, “Robust cell image segmentation methods,” *PATTERN RECOGNITION AND IMAGE ANALYSIS C/C OF RASPOZNAVANIYE OBRAZOV I ANALIZ IZOBRAZHENII.*, vol. 14, no. 2, 2004.

[47] E. Meijering, “Cell segmentation: 50 years down the road [life sciences],” *Signal Processing Magazine, IEEE*, vol. 29, no. 5, pp. 140–145, 2012.

[48] L. P. Coelho, A. Shariff, and R. F. Murphy, “Nuclear segmentation in microscope cell images: a hand-segmented dataset and comparison of algorithms,” in *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*. IEEE, 2009, pp. 518–521.

[49] M. Sezgin and B. Sankur, “Survey over image thresholding techniques and quantitative performance evaluation,” *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 146–165, 2004.

[50] N. Otsu, “A threshold selection method from gray level histograms,” *IEEE Trans. Systems, Man and Cybernetics*, vol. 9, pp. 62–66, Mar. 1979.

[51] H. Digabel and C. Lantuéjoul, “Iterative algorithms,” in *Actes du Second Symposium Europée en d'Analyse Quantitative des Microstructures en Sciences des Matériaux, Biologie et Médecine, Caen, 4-7 October 1977*, E. J.-L. Chermant, Ed. Stuggart: Riderer Verlag, 1978, pp. 85–99.

[52] S. Beucher and C. Lantuéjoul, “Use of watersheds in contour detection,” *International Workshop on Image Processing, Real-Time Edge and Motion Detection/Estimation, Rennes,*, september 1979.

[53] J. MacQueen *et al.*, “Some methods for classification and analysis of multivariate observations,” in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 281-297. California, USA, 1967, p. 14.

[54] MathWorks. (2013, May) kmeans. [Online]. Available: http://www.mathworks.com/help/stats/kmeans.html

[55] H. B. Mann and D. R. Whitney, “On a test of whether one of two random variables is stochastically larger than the other,” *The annals of mathematical statistics*, vol. 18, no. 1, pp. 50–60, 1947.

[56] MathWorks. (2013, May) ranksum. [Online]. Available: http://www.mathworks.com/help/stats/ranksum.html