TÉCNICO LISBOA

The Dynamics of the Spontaneous Epithelial-to-Mesenchymal Transition Process

Duarte Manuel Catrola Palma Ferreira

Thesis to obtain the Master of Science Degree in

Biological Engineering

Supervisor(s): Doctor Denis Gaston Wirtz and Professor Cláudia Alexandra Martins Lobato da Silva

Examination Committee

Chairperson:	Professor Duarte Miguel de França Teixeira dos Prazeres
Supervisor:	Professor Cláudia Alexandra Martins Lobato da Silva
Member of the Committee:	Professor Gabriel António Amaro Monteiro

November 2015

Acknowledgments

To begin I would like to thank Dr. Denis Wirtz for providing me with the opportunity to work in his lab at Johns Hopkins University, and for all his assistance and ideas throughout the length of the project.

To Dr. Pei-Hsun Wu for guiding me through this project, for all his assistance during experimental and analysis work, and most of all for his feedback.

To Dr. Daniele Gilkes for all her assistance during the experimental work and all the technical support.

To everyone at Wirtz Lab: to Dr. Jude Phillips for all his help, both in and out of the lab, his patience, to Hasini Jayatilaka for her teachings and her assistance; to Tânia Perestrelo for all her help in lab and adjusting to a new environment; to Nicolás Perez for being a great help and an even bigger friend; and to Dr. Meng-Horng Lee, Dr. Yu-Tai Chang, Dr. Lijuan He and Jacob Sarnecki for everything they did for me. And most of all, for all the good times we had, in and out of the lab.

A great many thanks to all the friends I made while there, for everything everyone did for me, Inês and Alexandre; and for all the good times we had.

In Portugal, many thanks to Leonor Guedes da Silva for paving the way and keeping open the communication; to Inês Godet for including me in the amazing opportunity, all her company and her help during this stage; to Alexandre Carnet for his company and help as well; to my supervisor Professor Cláudia Lobato da Silva for her assistance, Professor Miguel Prazeres for keeping in touch and checking up on me every so often; and most of all to my friends with whom I kept in touch during these months abroad.

To my family, for always being there, keeping in contacting and listening to me every once in a while, for all their help. And most of all, to my mother and father for all the support and encouragement. Thank you very much.

Resumo

A Transição Epitelial-Mesenquimal (EMT) é um importante processo que ocorre naturalmente no corpo humano. Foi identificada a sua presença no desenvolvimento embrionário, no processo de cura de ferimentos, e suspeita-se fortemente que seja uma das principais causas na metástase cancerígena. Estudos anteriores levaram à formação da hipótese que, ao contrário do pensamento geral, a EMT é algo que ocorre constante e espontâneamente no corpo humano, com o intuito de manter, num tecido, um equilíbrio de células epiteliais e mesenquimais. Estudou-se esta possibilidade, analisando-se não só a linha celular murina nMuMG, mas também duas outras derivadas por expansão clonal de células epiteliais e mesenquimais (E e M clones respetivamente). Observou-se inicialmente que tanto a linha mãe como as linhas derivadas dela mantinham percentagens diferentes dos dois tipos de células. Com microscopia time-lapse observou-se células com mudanças morfológicas, assumindo-se assim mudança no fenótipo. As linhas filha demonstraram variações ao nível proteico e de mRNA quando comparadas com a linha nMuMG, sendo que M clone mostrou semelhanças a amostras de nMuMG expostas a TGF-β excepto ao nível do mRNA. Isto pode ser indicador que os correntes estudos do processo EMT que recorrem a TGF-β correm o risco de ser incapazes de descrever com exactidão o que na realidade acontece. Também foi observado que, ao nível proteico, a densidade da cultura celular afectava de modo diferente as três linhas estudadas, embora se verificasse no mRNA que as tendências se mantinham independentes.

Palavras-chave: EMT espontâneo, epitelial, mesenquimal, proteína, mRNA

Abstract

The Epithelial-to-Mesenchymal Transition (EMT) is an important naturally occurring process in the human body. It has been identified to be present in embryonic development, wound healing, and there is strong suspicion that it is one of the leading causes of cancer metastasis. Previous studies had led to the theory that, unlike commonly thought, EMT is a constantly occurring spontaneous process, that it serves to, in the environment, maintain a balance of epithelial and mesenchymal cells. In this work this possibility is further studied, analyzing not only the murine epithelial cell line nMuMG, but also two others obtained by clonal expansion of an epithelial and a mesenchymal cells (E and M clone respectively). We observe that both the parental cell line show signs of both cell types, but so do the clonal lines. With time-lapse microscopy some cells were observed changing morphology, assumed as a phenotype switch. We show that the M clone and E clone cell lines show differences in protein and mRNA levels when compared to the nMuMG line, despite originating from it; and that for the most cases the M clone samples behave similarly to a TGF-β exposed nMuMG sample, except in regards to the mRNA levels. This is seen as an indication that current EMT studies using TGF-β may not accurately describe what occurs in the human body. Also noted was the fact that all three cell lines had different behaviors regarding cell crowding at the protein level while still keeping the same trend in mRNA transcription.

Keywords: spontaneous EMT, epithelial, mesenchymal, protein, mRNA

Contents

List of Tables

List of Figures

Glossary

2D Two dimensions

3D Three dimensions

ADGE Amplified Differential Gene Expression

Akt protein kinase B, associated with AKT8 virus oncogene cellular homolog

AT Adenine-thymine

BMP Bone morphogenic protein

BMPR Bone morphogenic protein receptor

Bp Base pair

BSA Bovine Serum Albumine

β-TRCP β-transducin repeat-containing protein

cDNA complementary DNA, obtained by reverse transcription of mRNA

CSC Cancer Stem Cells

C_T Cycle threshold

CTC Circulating Tumor Cell

Dhh Desert hedgehog

DNA Deoxyribonucleic acid

dRFU/dT Derivative of RFU as function of T

E Epithelial

E (%) Primer efficiency (in percentage)

ECM Extracellular matrix

EGF Epidermal growth factor

EGFR epidermal growth factor receptor

ELISA Enzyme-linked immunosorbent assay

EMT Epithelial-to-Mesenchymal Transition

EndMTEndothelial-to-Mesenchymal Transition

ERK Extracellular signal-regulated kinase

FGF Fibroblast growth factor

FOX forkhead box protein family

GAPDH Glyceraldehyde-3-Phosphate Dehydrogenase

GC Guanine-cytosine

GFP Green Fluorescent Protein

Gli gliblastoma-isolated protein family

GSK-3β Glycogen synthase kinase 3 β

HIF Hypoxia inducible factor

HRP Horseradish peroxidase

Ihh Indian hedgehog

IL-6 Interleukin-6

ILK Integrin-linked kinase

JAG2 Protein jagged-2

JNK Jun N-terminal kinase

LEF Lymphoid enhancer-binding factor

LOX lysil oxidase

LRP LDL-receptor-related protein

M Mesenchymal

MAPK Mitogen-activated protein kinase

METMesenchymal-to-Epithelial Transition

miRNA micro RNA

MMP Matrix metalloproteinase enzymes

mRNA messenger RNA

mTORC mammalian target of rapamycin complex

MUC1 mucin 1, cell surface associated

NET Neuroendocrine tumor

NF-κB Nuclear factor κ B

nMuMG normal murine mammary gland cell line

P16INK4a cyclin-dependent kinase inhibitor 2A

p21 cyclin-dependent kinase inhibitor 1

PDGF Platelet-derived growth factor

PDK1 phosphoinositide dependent kinase 1

PI3K Phosphoinositide 3 kinase

PIAS Protein inhibitors of activated Stats

PIP$_2$ Phosphatidylinositol 3,4-bisphosphate

PIP$_3$ Phosphatidylinositol 3,4,5-trisphosphate

PTCH Patched protein family

PVDF Polyvinylidene fluoride

RFP Red Fluorescent Protein

RFU Relative fluorescence units

RNA Ribonucleic acid

RPA Ribonuclease Protection Assay

Rpm Rotations per minute

RTK Receptor tyrosine kinase

RT-qPCR Reverse Transcription – quantitative Polimerase Chain Reaction

SAGE Serial Analysis of Gene Expression

SD Standard deviation of the distribution

SDS-PAGE Sodium Dodecyl Sulfate – Polyacrilamide Gel Electrophoresis

SEM Standard error of the mean

SFRP1 Secreted Frizzled-related protein 1

Shh Sonic hedgehog

SIP1 Survival of motor neuron protein-interacting protein 1

SMAD Contraction of Sma (gene Small body size) and Mad (Mothers against decapentaplegic)

Smo Smoothened protein

T Temperature

TACE (Tumor necrosis factor)TNFα-converting enzyme

TBS Tris Buffered Saline

TCF T cell factor (Transcription factor)

TGF-β Transforming growth factor β

TGF-βR Transforming growth factor β receptor

TLE1 Transducer-like enhancer protein 1

TTF-1 thyroid transcription factor-1

UV Ultraviolet

ZEB1 Zinc finger E-box-binding homeobox 1

ZO-1 zona occludens 1

Chapter 1

Introduction

1.1 Cell: the beginning

The world of microscopic life became visible around the beginning of the 17th century, with the invention of the microscope [1]. Following several observations by microscopists of the time, Robert Hooke is generally attributed with being the first person to provide a description of the cell, in *Micrographia*, a work published in 1665 [1]. At the time, the cell was defined as the microscopic units that made up the structure of a slice of cork, but nonetheless, the modern meaning of the word cell comes from this book.

In this period, the old Aristotelian doctrine of spontaneous generation was known and accepted as the prevalent theory [1]. This stated that the environment had the ability to spontaneously create different life forms.

However, starting with Lazzaro Spallanzani, this line of thought started being put in question, as experiments showed that organisms were derived from previously existing ones, and proved the existence of a gap between life and inanimate matter [1]. However, it is considered that it wasn't until *ca.* 150 years ago that the theory of spontaneous generation was completely disproved by Louis Pasteur.

Around this time, there were several individuals performing research on this topic, and Matthias Schleiden, Theodor Schwann and Rudolf Virchow created what is now known to be the Classical Cell Theory. It consisted of three primary points, which later on became five and this new version was referred to as Modern Cell Theory [2]. The five tenets are:

1. All living organisms are composed of one or more cells.
2. Cells are the basic units of structure, function and physiology in living things.
3. All cells arise from pre-existing, living cells, by biogenesis.
4. Cells contain and pass on hereditary information during cell division.
5. All cells are relatively the same in relation to chemical composition and metabolic activity.

Along with this, arose the concept of cell division (a way to increase the number of cells through the existing ones). This implied that all the cells in an organism have origin from pre-existing ones, and the

ultimate conclusion that every multicellular organism originated from a single cell, the fertilized egg [3].

1.2 The discoveries of the cell types

Through biological research, analyzing increasingly more complex organisms, it was found that from the initial cell, the zygote, resulted several different cell forms, carrying morphological and phenotypical differences, earning the name: cell types.

The rise of the different cell types is only possible because of the zygote, not only the initial cell, but also the primary stem cell. A stem cell is a cell that has two main characteristics: it is capable of self renewal through division, giving rise to more stem cells, and it is capable of turning from an unspecialized cell to a tissue/organ-specific cell with specialized functions. This change from stem cell to function-specific cell is referred to as differentiation, and is the process through which a single self renewing cell can develop into a fully grown multicellular organism [3, 4].

A complex multicellular organism may contain many varied cell types, each group with its specific function. While these types may present visual or functional differences between them, their origin is common, they are all descendants of the original zygote. This means they are genetically identical, they carry the same genotype, but through regulation mechanisms, certain parts are expressed and others repressed, leading to the final observable differences between cell types.

It was in the late 19th century that, based on shape and organization during the embryonic development, two of the main types of cells in the human bodies were recognized: epithelial and mesenchymal [5]. These observations were substantiated based on the observable shape and organizational differences during the embryonic development.

1.3 Epithelial and Mesenchymal cells: what are they?

Epithelial cells are the main cell type present in epithelial tissue. In this tissue, they are tightly bound (tight and gap junctions, desmossomes, *etc.*) in sheets called epithelia. Epithelial tissue lines the cavities and the surfaces of structures throughout the body, and is also usually composed of neuron endings (for sensory perception). However, this type of tissue has no capillaries, so obtaining nutrients and the disposal of waste is done by diffusion through the basement membrane (which separates epithelial tissue from connective tissue) [6].

Epithelial cells are organized in continuous polarized sheets, which is one of the most basic forms of organization in multi-cellular organisms. It can be seen in many organisms besides humans, and the epithelial blastoderm cells (epiblast) are the basis for initiating the embryo development [3, 7]. These sheets can be further organized in different ways (continguous simple or stratified sheets).

When talking about simple epithelia, the cells that create it are polarized along an apical-basal axis (front of the cell-extremity attached to basal membrane), that is perpendicular to the plan of the sheet of cells; are firmly connected with their neighbor cells to ensure mechanical integrity. They also form a permeability barrier separating tissues and organs from one another (even the embryo from the rest of the

environment) [7]. In order to further maintain polarity and stabilize epithelial architecture, a membrane of extracellular matrix proteins is formed on the basal side (thus basal membrane).

On a side note, this barrier also restrains the epithelial cells, preventing changes in shape, polarity and motility, besides being a physical barrier for cellular transformations, like the ones explained further on in section 1.5 [7].

Mesenchymal cells are mostly present in the mesenchyme, from where they can develop into the connective tissue and the lymphatic and circulatory systems. Contrary to epithelial cells, these prefer to grow surrounded by ECM instead of similar cells and are also rather mobile.

Table 1.1: A compilation of the most common characteristics of both epithelial and mesenchymal cells

Epithelial phenotype:	**Mesenchymal phenotype:**
Growth in colonies	Increased production of ECM compounds [3]
Apical-basal polarity [5]	Resistance to apoptosis [3]
Round morphology [8]	Invasiveness [3]
Strong intercellular adhesion [9, 10]	Spindle morphology [9]
	Cell motility [9]
	Isolated growth [11]
	Resistance to anoikis [11]

The main phenotypic differences are listed in Table 1.1, while most of the molecular markers of these two types of cells can be seen in Figure 1.1.

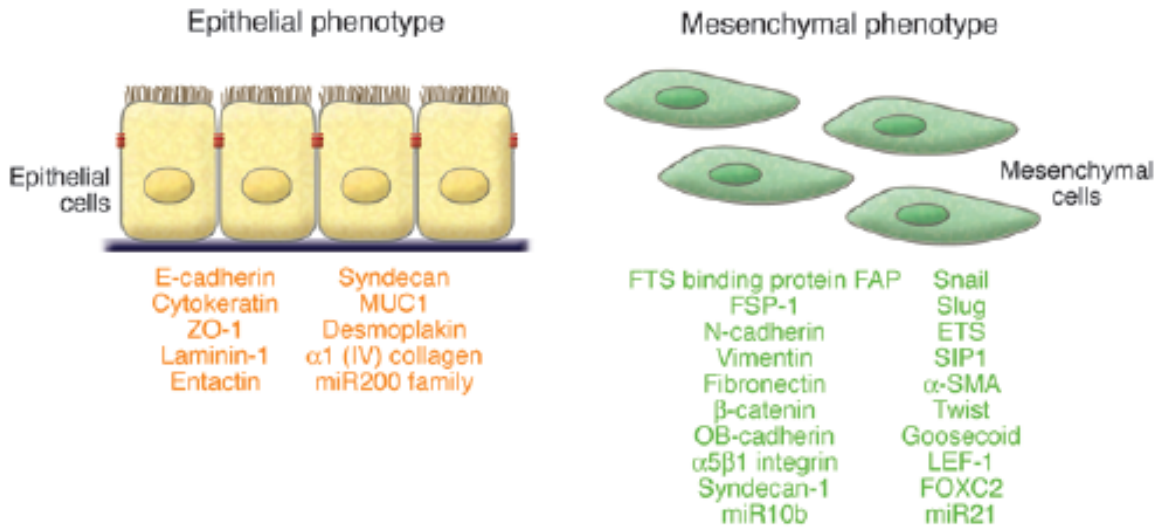

Figure 1.1: Compilation of some of the most known protein markers for identification of epithelial and mesenchymal markers. ZO-1: zona occludens 1; MUC1: mucin 1, cell surface associated; miRNA200: micro RNA 200; SIP1: Survival of motor neuron protein-interacting protein 1; FOXC2: forkhead box c2. Original image obtained from [3]

In epithelial tissues, epithelial cells all have a specific function to perform, while mesenchymal cells maintain a support role [3]. This observation supported the notions that a cell needed to reach its final stage of differentiation to perform specialized functions, and that once achieved, said cell would be maintained in this stage [12].

Further studies later on provided evidence that it was not so, some cell types, epithelial and mes-

enchymal among them, proved to contain some sort of plasticity, being able to switch between different types despite the fact that they had reached their final differentiation stage [3, 7, 13]. Also related was the discovery that, although it is yet unknown whether this can occur naturally in humans (it is known to occur in some highly regenerative species [14]), terminally differentiated cells can be induced to become stem cells once again, another example of plasticity [15].

1.4 Cellular transformations and the Epithelial-to-Mesenchymal Transition (EMT)

Frank Lillie, in 1908, was the first researcher to mention a transition between cellular states [16]. However, it wasn't until the late 1960s that, with Elisabeth Hay's detailed description of the formation/development of the primitive streak in the chick that there was another mention of "Epithelial-to-Mesenchymal Transformation" [17, 18]. Still, it took a while (1982) for this cellular event to be recognized as a cell process [19]. Afterwards, focus on EMT started increasing, gaining more momentum with Stocker and Perryman's research that indicated that under certain conditions, this process could be induced [20, 21, 22].

Nowadays, this process is called "Epithelial-to-Mesenchymal Transition", in order to reflect the reversible nature of the changes observed in the cells (this correlates with the discovery of the "Mesenchymal-to-Epithelial Transition") [3, 23, 24, 25].

1.5 EMT, its opposite MET and even EndMT

EMT is a process that leads an epithelial cell, through biochemical modifications, to obtain a mesenchymal phenotype. It ends with a mesenchymal cell (with migratory potential) and the degradation of the basement membrane to which the previously epithelial cell was attached [3].

Situations that EMT has been observed in range from development (dispersing cells in embryogenesis) [5] to tissue repair (presence of some of the fibroblasts in injured areas) and pathological stress (invasion and metastasis of some epithelial cancers) [3, 10, 26].

MET on the other hand is the reverse of the previously mentioned process. Here, mesenchymal cells are modified towards obtaining an epithelial phenotype.

MET has been found in most of the situations EMT is known to occur, chief among them the developmental period, as some of the cells dispersed through EMT are turned back to epithelial cells once again, but also in pathological cases, it is thought that epithelial cancer metastasis starts with EMT to obtain mobile cells, but ends with MET to turn these cells into secondary tumors [3, 27, 28, 29].

A simple schematic representation of the mentioned EMT and MET processes can be found in Figure 1.2.

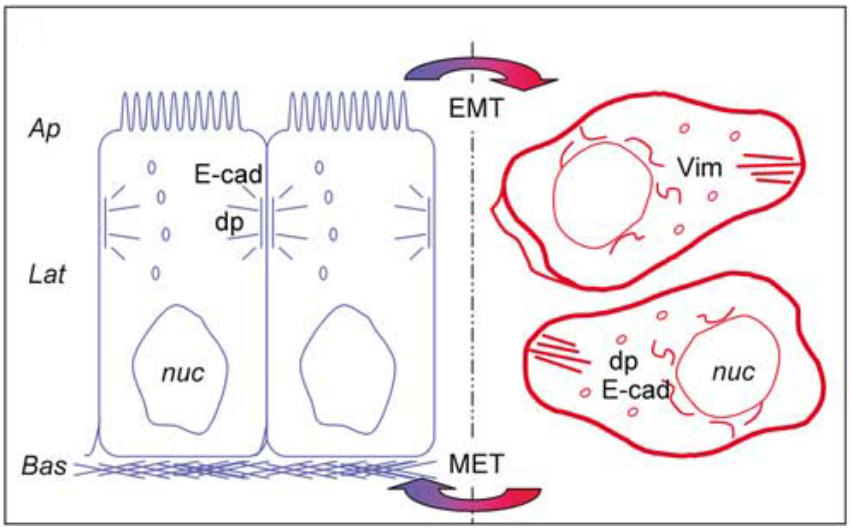

Figure 1.2: Simplified diagram describing the EMT/MET process. Ap: apical side of the membrane; Lat: lateral side of the membrane; Bas: basal side of the membrane/basal membrane; E-cad: E-cadherin; dp: desmoplakin; nuc: nucleus; Vim: vimentin. Image obtained from [10].

It is important to note that although the EMT/MET processes are usually thought of as the transition between two states, that is not entirely accurate. It has been observed in studies that the transition may be halted at nominally any stage, therefore leading to a variety of different possible combinations of factors expressed or repressed, resulting in cells that may have characteristics of both epithelial and mesenchymal phenotypes [3].

A particular case of EMT discovered was the EndMT, upon which endothelial cells undergo a change towards obtaining a mesenchymal phenotype [3, 30]. This process was first identified in heart formation studies [30], but it has since been identified in many other occasions, like fibrosis (it is suspected that many fibroblast in cases of fibrosis are generated through EndMT [3]) and even cancer [30].

1.6 The different types of EMT

There are three main situations in which one can observe EMT (which is not to say that there are only these three, but so far these three scenarios compile the majority of reported EMT observations), and though it is essentially the same program occurring, they all end up producing mesenchymal cells with different functions [3]. Figure 1.3 provides a visual representation of the three situations in which EMT has been identified.

The EMT referred to as type 1 are the ones involved in implantation of the embryo, placenta formation [3] and a variety of tissue remodeling events, including mesoderm formation, neural crest development, heart valve development, secondary palate formation, among others [5, 26, 31]. There is no fibrosis involved and the resulting cells do not display an invasive phenotype.

Most of all the EMTs occurring at the embryonic development level begin with apical constriction, forcing the cytoplasm to converge on the basal region, inducing a shape change. This can aid the cells in breaking the basal lamina, but as mentioned previously, it does not occur on all EMT cases, marking this

change as not being an essential feature of the process. The mechanisms for rupturing the basal lamina in the EMTs verified at this stage are yet poorly understood. Along with this it is verified an increase in mesenchymal-like cell-cell and cell-matrix adhesion molecules, side-by-side with internalization of epithelial-like adhesion components (especially when it involves constriction). The combination of all these factors leads to detachment of the cell [32].

During the embryonic developmental stage, the EMT process is not activated as a one-way transition, committing cells to a final mesenchymal phenotype. In some cases, the mesenchymal cells, like the ones that formed the primary mesenchyme revert back to epithelial state through MET, originating secondary epithelia [3, 26, 31]. The best-studied example is the formation of the nephron epithelium during kidney development [26]. This ability of a mesenchymal cell to revert back to an epithelial phenotype demonstrates, once again, the possibility of substantial cell plasticity and also suggests this same interconversion may occur under certain pathological conditions.

Type 2 EMT refers to wound healing, tissue regeneration and organ fibrosis. Here, the transition occurs as part of the healing process, generating fibroblasts and other cells important in tissue reconstruction [12], following trauma and inflammatory response. In wound healing and tissue regeneration the EMT process is finished once inflammation is attenuated, but in the case of organ fibrosis, the fibroblasts contribute to perpetuate the effect, never ending the inflammatory response, leading up to organ destruction [3].

The origin of the tissue fibroblasts has long been debated, with the reigning concept being that they originate from embryonic mesenchymal cells that were left over. The other two possibilities held in

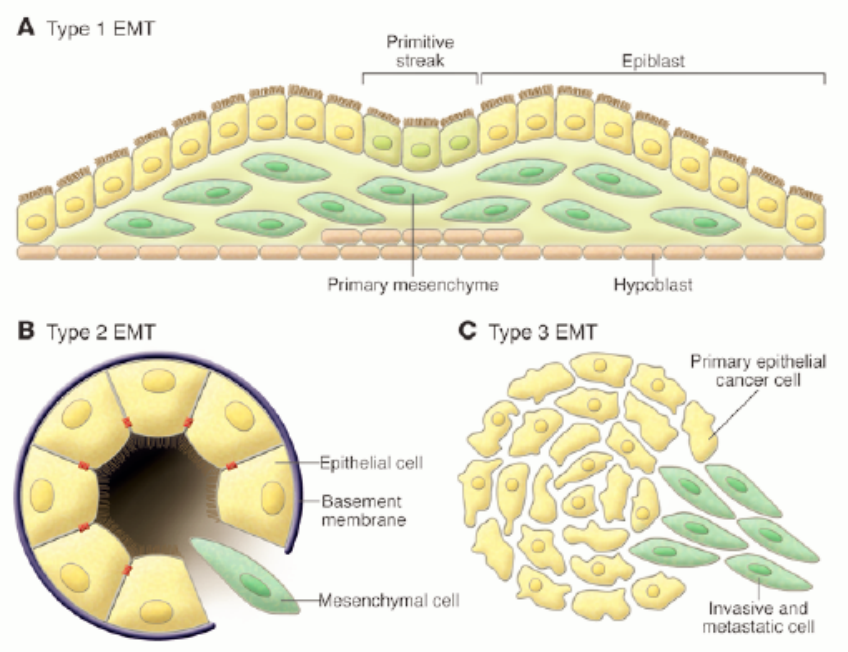

Figure 1.3: The three known EMT types [3]. A represents Type 1 EMT, which is associated with the formation of the primary mesenchyme, and eventually, the generation of the three germ layers. B showcases the Type 2 EMT, occurring in cases of inflammation. This type of EMT, unlike the previous one, occurs over longer periods of time, as long as the inflammation persists, which can lead to the destruction of the affected organ (fibrosis). C details the Type 3 EMT that is thought to occur in epithelial cancer cells, being suspected of initiating the metastatic process.

discussion were that they came through the blood stream once released from the bone marrow, and that they were locally derived from EMT occurring in tissues. These last two thoughts were found to be mechanistically identical, meaning that all of those fibroblasts are probably created through EMT. It was long accepted that fibroblasts were a cell type of low diversity, however there is now evidence, through the knowledge that some are formed through EMT and that they show subtle differences (biochemical, phenotypical, response to external factors) that disproves this. As it is, there is suspicion that fibroblasts created through EMT express a slightly different gene profile, containing still some traces of signaling pathways pertaining to those cells previous function, and as thus they can be a cell type as diverse as epithelial ones [12]. This also means it is possible that what occurs to create some of the fibroblasts observed in these cases is actually a partial-EMT, as some cells still maintain some of their epithelial markers [3].

Most injury studies that supply information towards Type 2 EMT knowledge pertain to the kidney, some to the lung, and in these cases, the injury leads to an abnormal amount of inflammatory cells that can easily incite EMT and fibroblast formation. These fibroblasts then secrete collagen and other matrix components on the wounded area, however, while this is helpful at moderate levels, if the inflammation process does not abate, then collagen will accumulate and eventually lead to organ failure through fibrosis [26, 33]. Also known is that in cardiac and renal fibrosis cases, EndMT is also verified [3, 26].

Type 3 EMT occurs in neoplastic tissue, in cells that have suffered genetic and epigenetic mutations (in oncogenes and tumor suppressing genes) leading to abnormal growth and formation of tumors. In carcinomas the EMT process leads to invasive mesenchymal cells, being a factor in cancer metastasis (the major reason of cancer death) [5]. Cancer cells undergoing EMT can do so to several extents, from maintaining some epithelial markers and gaining some mesenchymal ones up to losing all epithelial characteristics and becoming fully mesenchymal. So far, for carcinomas it is still unknown what is the definitive signaling inducing this process [3].

The initial markings of epithelial cancer growth are excess of epithelial cell proliferation and angiogenesis. The acquisition of invasion, marking the initiation of the metastatic process, begins by breaching the basal membrane. The genetic controls and biochemical mechanisms have been a source of focus for research [3], as preventing this initial stage could mean a massive breakthrough in cancer treatment.

One of the most convincing pieces of evidence that corroborate the interaction between EMT and cancer is the knowledge that multiple of the regulators involved in the EMT process can lead to enhanced tumors and increase metastasis [33].

Through some experiments, it was verified that some carcinoma cells can acquire mesenchymal phenotype, and that these cells are usually verified at the invasive front of primary tumors. One suggestion for the cause of EMT in cancer cells is that the genetic and epigenetic alterations undergone while forming the primary tumor are what makes these cells extremely susceptible to EMT-inducing signals [3].

The relevance of EMT in cancer metastasis has been long been a matter of debate, as the secondary tumors observed in clinical samples showed lack of mesenchymal phenotype and mesenchymal cells obtained after inducing EMT in epithelial tumor cells are almost indistinguishable from stromal or other

tumor-associated fibroblasts [26]. However, some studies have shown that the reversibility of EMT is why there is a lack of mesenchymal phenotype, the initial stimulus for the transition to occur from epithelial to mesenchymal is not present in the distant location the metastatic cells reach, and as such, transition from M to E occurs [8]. In other experiments, EMT characteristics, most notably the loss of some epithelial markers and expression of mesenchymal ones, has been associated with transformation of non-invasive to aggressive carcinoma [34] and also, features of EMT have been noticed in CTCs [35], evidence that this process is used in the dispersion of tumor cells.

The current thought on the actions of Type 3 EMT in cancer metastasis can be observed in Figure 1.4.

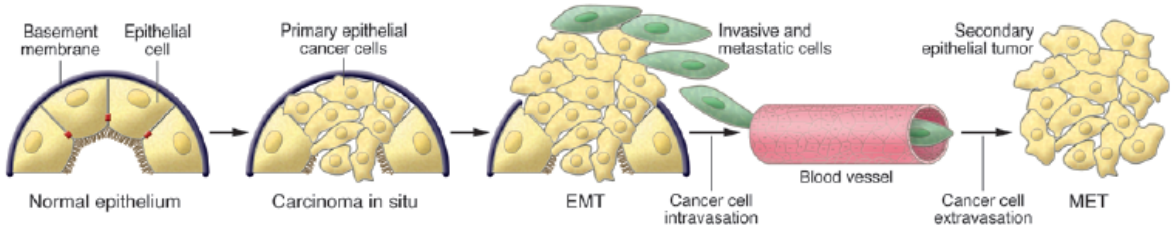

Figure 1.4: The supposed contribution of EMT/MET to cancer metastasis [3]. Transformation to invasive carcinoma takes several stages: epithelial cells begin to lose polarity and detaching from the membrane (EMT), invasive cells migrate to the bloodstream (intravasation), at a different location mesenchymal cells exit the blood vessels (extravasation), cells are converted back to epithelial state and then replicate, forming the secondary tumors (MET).

So far, the complete rundown of the signals involved in the EMT process is not yet clear, but there is enough knowledge (on the similarities and differences of the known signaling pathways registered to each EMT type) to be able to conclude on the existence of biochemical differences to accompany the functional ones verified between all three types of EMT [3].

1.7 EMT signaling

The EMT is a major cellular process, involving many different components, be it transcription factors, cell-surface proteins, cytoskeletal proteins, ECM-degrading enzymes, micro-RNAs, *etc*. A schematic representation of the known signal transduction pathways involved can be seen in Figure 1.5.

It has been assumed that the known molecules involved in these pathways can be used as markers to assess the existence and/or progression of the EMT process [3]. Table 1.2 lists some of the markers used nowadays.

One of the most important molecules in this process is E-cadherin, as it seems to be the one whose behavior is most conserved during the various different EMT observations. As such, most of the studies in EMT focus on this protein, and many of the discoveries of molecules involved in signaling pathways was through changes in this proteins' expression.

So far, many signaling pathways related to EMT activation have been discovered, the most notable ones involve the following molecules as activators: TGF-β, BMP, EGF, FGF, PDGF, Wnt, Notch receptor, Hedgehog and several integrins. Other notable pathways relate to ECM and even hypoxia signaling [25, 68].

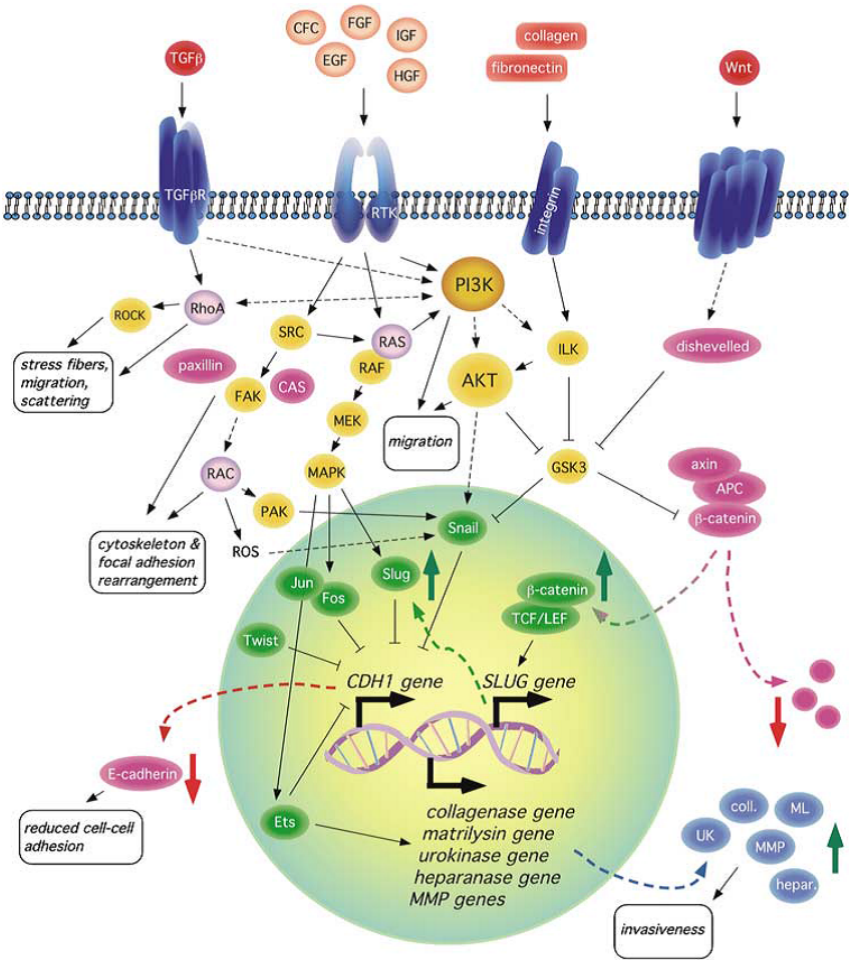

Figure 1.5: Representative summary of the more notable EMT signaling pathways, the molecules involved and their impact in the cell [10]. TGF-βR: TGF-β receptor; CFC: Cripto_Frl-1_Cryptic domain; IGF: Insulin-like growth factor; HGF: Hepatocyte growth factor; ROCK: Rho-associated, Coiled-coil-containing Protein Kinase; FAK: Focal Adhesion Kinase; CAS: cellular apoptosis susceptibility protein; RAC: Ras-related C3 botulinum toxin substrate; ROS: Reactive oxygen species; PAK: p21-activated kinase; MEK: Mitogen-activated protein kinase kinase (MAPKK); RAS: Ras kinase family; RAF: Raf kinase family; RhoA: Ras homolog gene family, member A; UK: Urokinase; ML: Matrilysin; CDH1: E-cadherin gene

1.7.1 TGF-β and BMP signaling

The most well characterized pathway is the one involving the TGF superfamily. The ligands can be any of the three TGF-β isoforms (1 through 3) or of the 6 BMP ones (2 through 7). The choice of ligand is usually associated with the situation in which EMT is being induced [25]. The receptor involved is TGF-βR (I and II) with TGF-β ligands and with BMP R (I and II). After the ligand attaches, the receptor II trans-phosphorylates receptor I, thus activating their kinase ability. There are several types of each receptor subunit, and different combinations respond slightly different to each isoform of the ligands, allowing for several cascades to be activated from there [25].

Something very important in this signaling pathway is that it branches out at this point, into SMAD dependent/independent signaling.

SMAD dependent signaling sees the recruitment of SMADs 2 and 3, which are phosphorylated and used in the formation of a complex with SMAD 4. This new complex has a specific region that facilitates

Table 1.2: Frequently used protein markers for studying EMT [34].

	Marker	Original function	Reference
Downregulated in EMT	α-catenin	Cell adhesion molecule	[36]
	β-catenin (membrane)[1]	Cell adhesion molecule	[37, 38]
	Claudin	Cell adhesion molecule	[39, 40]
	Cytokeratins	Cytoskeletal filament	[41, 42, 43]
	Desmoplakin	Cell adhesion molecule	[33]
	E-cadherin	Cell adhesion molecule	[41, 42, 44, 45, 46, 47, 48, 49, 50, 51]
	Occludin	Cell adhesion molecule	[52, 53]
Upregulated in EMT	Brachyury	Transcription factor	[54]
	β-catenin (cytoplasm/ nucleus)[1]	Transcription factor	[55]
	EGFR	Tyrosine kinase receptor	[51]
	Fibronectin	Cell adhesion molecule	[33]
	FOXC2	Transcription factor	[33]
	Integrin αvβ6	Cell adhesion molecule	[33]
	MMP-2/3/9	Enzyme(s)	[33]
	N-cadherin	Cell adhesion molecule	[49, 56]
	Notch-1	Transcription factor	[57]
	p16INK4a	Cell cycle regulator	[58, 59]
	Slug	Transcription factor	[60, 61]
	Snail	Transcription factor	[51, 60, 61]
	Sox10	Transcription factor	[33]
	TTF-1	Transcription factor	[42]
	Twist	Transcription factor	[62, 63]
	Vimentin	Cytoskeletal filament	[41, 43, 51, 64]
	ZEB1	Transcription factor	[49, 65, 66, 67]

[1]: Membranous depletion, but cytoplasmic accumulation/nuclear translocation.

DNA binding, and the steps that led to its formation expose the nuclear localization signal, which eases its transport in to the nuclear envelope [25]. On a side note, this pathway can also recruit SMADs 6 and 7, which bind to TGF-βRI and inhibit SMAD complex formation, allowing for an added level of control in case of unneeded activation of this pathway. BMP signaling goes much the same, with the only difference being that instead of SMAD2 and 3, the complex is formed with SMADs 1, 5 and 8.

Smurf 1 and 2 proteins, along with PIAS1 can be used to block this SMAD signaling, by binding SMADs and leading them to proteasome degradation. Smurf 1 and 2 are known to inhibit TGF-β signaling, while PIAS1 is known to be downregulated through this pathway [25].

Upon reaching the nucleus, SMAD complexes then bind to DNA and induce transcription of genes essential to EMT activation. Among these genes we find Snai1/2 and Twist, known to suppress expression of E-cadherin and Occludin; and also ZEB 1/2, whose functions are to facilitate the expression of the previously mentioned E-cadherin repressor genes [25]. Also important is the expression of LEF-1, that combines with the PI3K-Akt pathway directly activated by TGF-β, inhibits GSK-3β, which in turn enables β-catenin activation of LEF-1 so as to induce EMT. LEF-1 may also be activated without β-catenin by forming complex with SMAD 2 and 3.

The SMAD independent signaling involves many pathways common to the RTK signaling. The PI3K-Akt pathway phosphorylates PIP_2 to PIP_3, a phospholipid membrane protein that binds Akt and then recruits PDK1 to phosphorylate it. This phosphorylation can also be induced by integrin stimulation, which recruits ILK. Akt is known to inhibit GSK-3β, which as mentioned earlier activates β-catenin, but it also prevents Snai1 phosphorylation and thus degradation through β-TRCP. Additionally Akt also activates mTORC 1 and 2, which contribute in their own way to EMT [25].

TGF-β is also implicated in the activation of many other kinase pathways, such as ERK, JNK and

MAPK [25]. The pathways mentioned here can be activated either independent of SMAD signaling or through SMAD complexes.

Most of what was described can be observed in Figure 1.6.

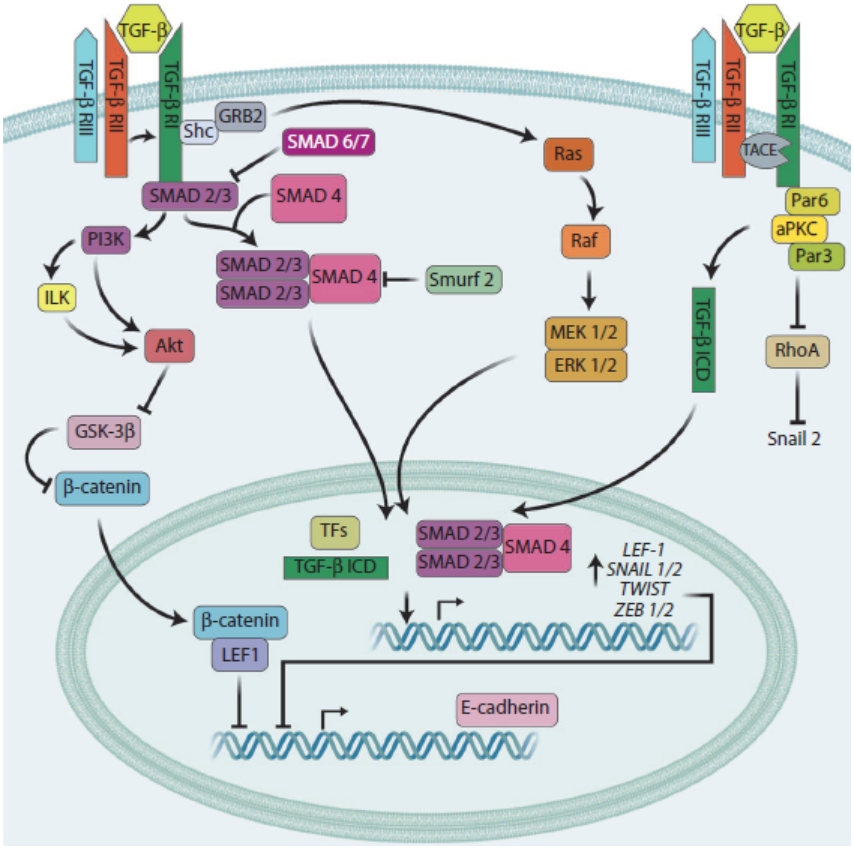

Figure 1.6: Representation of the TGF-β signaling pathways involved in EMT [25].

1.7.2 Wnt signaling

The signals generated by Wnt molecules are transferred across the cell membrane through Frizzled (a G-protein coupled receptor) and LRP receptors. The absence of Wnt leads to β-catenin phosphorylation through the action of GSK-3β and Axin, which causes its accumulation in the cytoplasm and eventual degradation. Through Wnt binding to Frizzled, GSK-3β ends up phosphorylating LRP6 and recruiting Axin and Dvl to the cell membrane. Without Axin to form a complex, GSK-3β does not act on β-catenin, allowing it to migrate to the nucleus. Maintaing GSK-3β in the cytoplasm also further prevents Snai1 from being phosphorylated and degraded [25].

As mentioned in subsection 1.7.1, nuclear β-catenin binds to LEF-1 or other members of the TCF/LEF family and promote EMT, most notably by inhibiting CDH1 (E-cadherin) gene expression [25].

As it was mentioned previously, this signaling pathway has at least one layer of crosstalk with the TGF-β one, specifically, in the β-catenin or SMAD complexes binding to LEF stage.

1.7.3 Notch signaling

The Notch receptor is a tans-membrane protein, containing both extra and intracellular domains. The interaction with another such receptor from a neighboring cell leads to a release of the intracellular domain (NICD), through cleavage by γ-secretase and TACE. As NICD contains a nuclear localization sequence, once detached, it translocates to the nucleus. Once there, it binds to a family of DNA-bound transcription repressor complexes (CSL), which transforms these repressors into activators, and leads to the expression of NF-κB, Akt and p21 (proteins more commonly known for their involvement in tumor development) [25].

Also known is that Notch signaling can regulate Snai1 expression, both directly and by inducing HIF-1α. This protein binds to the LOX promoter, facilitates transcription and further on mediates stabilization of Snai1 [25].

Snai2 can also interact with Notch, being essential for this signaling pathways' repression of E-cadherin and β-catenin. In endothelial cells, Notch represses VE-cadherin and induces EndMT. Through experimentation it has been confirmed that the absence of Notch leads to a partially reverse EMT (or partial MET) [25].

Notch can also, through NF-κB and β-catenin and various regulatory miRNAs, regulate EMT. JAG2 can also be used as a ligand for Notch receptor, promoting EMT through inhibition of the miRNA-200 family [25].

1.7.4 Hedgehog signaling

The Hedgehog (Hh) family includes many members, among them Shh, Dhh and Ihh. These ligands all bind to PTCH1 and 2, which in the absence of ligand inhibit Smo activity. The interaction of Hh with PTCH1/2 leads to the release of Smo, initiating a cascade that activates Gli family transcription factors. This ends up promoting the transcription of various genes, some of which PTCH, Wnt and Snai1 [25]; which are known to lead to EMT induction.

This signaling pathway leads to TGF-β1 secretion, crossing over to its own EMT pathway, and also leads to production of JAG2, leading another crossover, this time into Notch signaling. It also targets SFRP1 (EMT repressor [69]), which crosses into Wnt signaling and FOX1 and FOX2, transcription factors that regulate the accumulation of β-catenin in the cytoplasm [25].

Some studies show that Hh signaling appears to inhibit the Wnt signaling pathway through paracrine signaling, when in the presence of local mesenchymal stromal cells. It is possible that this is a cause for the epithelial phenotype verified in tumors, and that the cells observed at the invasive front, through being relatively free of such signaling, exhibit features consistent with EMT [25].

The Wnt and Notch signaling pathways, mentioned in subsection 1.7.2 and 1.7.3, along with the one described in this chapter are represented in Figure 1.7.

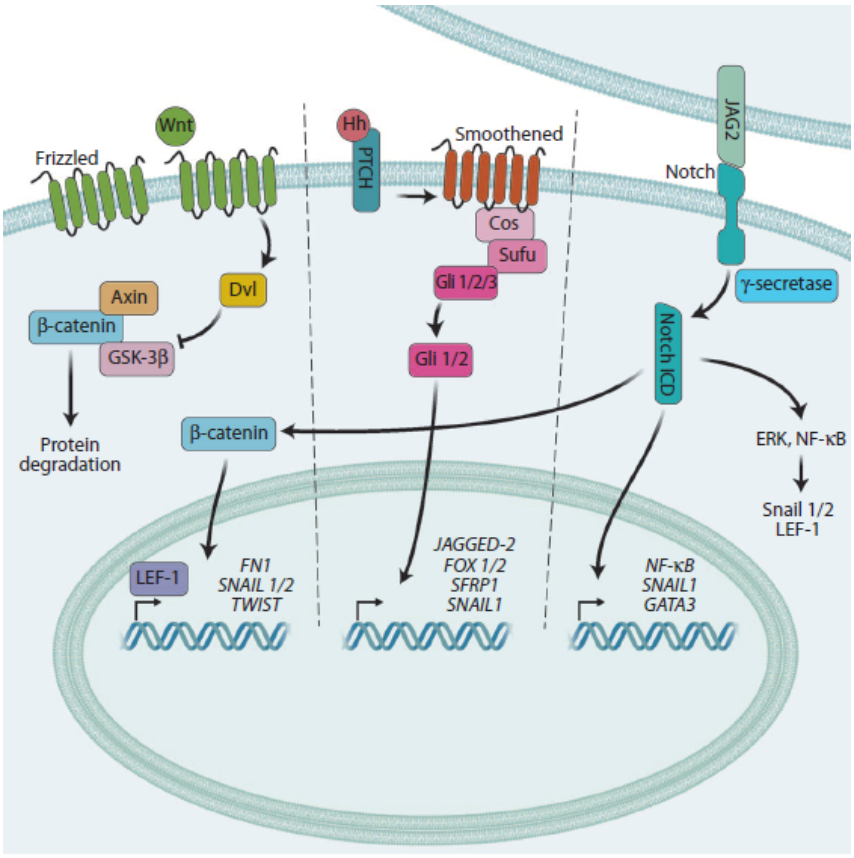

Figure 1.7: Representation of the Wnt, Notch and Hedgehog signaling pathways involved in EMT [25].

1.7.5 Other signaling pathways

For other signaling pathways, like the Matrix and Hypoxia mentioned in section 1.7, check ”Signaling Mechanisms of the epithelial-mesenchymal transition” by David Gonzalez and Damian Medici [25], and the references mentioned there.

1.7.6 Other observations

It has been observed, and reported in some of the pathways presented, that some oncogenes also work to induce, while others inhibit, the EMT process [10, 69]. Studies have also shown the many of the signaling pathways involved in the EMT/MET process are also responsible for CSC generation and regulation [70, 71].

Other EMT inducers recently discovered include: Bleomycin [72], IL-6 [73], TLE1 [74], among many others.

A somewhat unexplored signaling pathway in EMT is calcium. A study in cancer cells shows that the stimulus caused by the induction of EMT lead to variations in calcium levels in the cells, and that silencing said signal had an inhibitory effect in EMT [75].

1.8 Novelties and Motivation

General opinion on this topic is that EMT and its opposite, MET, similarly to many chemical reactions, require a significant jump start (contrary to others that are constantly occurring in the background) in order to actually occur, in either direction. This idea that EMT is a process to be induced by external factors like TGF-β only when there is the need to fulfill a certain function, may be incorrect.

Previous work wielded data that showed the epithelial cell line nMuMG constantly maintains, in equilibrium (once again, similar to a standard reversible reaction) a mix of both epithelial and mesenchymal cells when in culture. This could occur through exposure to one of the afore mentioned many inducible factors. However, in that case the exposure would need to be low, as there are a mix of both cell types, but therein lies the issue, as a continuous exposure would be needed to maintain the cells in a mesenchymal state (by the same logic, the stimuli needed to convert epithelial cells to mesenchymal would need to be removed to re-stimulate the conversion back to the epithelial phenotype), and a continued exposure would surely increase the % of population in the mesenchymal state. As such, the hypothesis that the EMT process is spontaneous and is constantly occurring in the background, maintaining a balance of both types of cells was thought up.

This line of thought led to the creation of two new cell lines, using as a progenitor the nMuMG one. By serial dilution of a nMuMG cell suspension, single cells were separated, observed and annotated as epithelial or mesenchymal through their morphology. From a single epithelial cell, and by clonal expansion, the E clone cell line was grown. In a similar way, but using a mesenchymal cell, the M clone cell line was obtained.

This led to the opportunity of researching the EMT process in a way that is completely different to everything that is done today, not only that , but also to compare the newly (suspected) discovered EMT with the previously studied induced EMT.

1.9 Proposal

The first stage of this proposed work is to confirm the theory that the nMuMG cell line, being categorized as epithelial, contains mesenchymal cells in its midst, when in culture. Following that, the two cell lines E clone and M clone (derived by clonal expansion from an epithelial and a mesenchymal cell, respectively, originating from the paternal nMuMG line) are to be analyzed in the same manner. This is meant as additional confirmation that the EMT process is spontaneous, as if it is not, E and M clone cell lines should accuse only one type of cell, epithelial and mesenchymal, respectively, instead of both.

Assuming evidence is favorable for a spontaneous and continuously occurring EMT, the next step is to observe with time-lapse phase contrast microscopy the nMuMG line. This occurrence should be observable as a number of cells in the sample that switch morphology from epithelial to mesenchymal and *vice-versa*, further supporting the assumption tested.

Deviating from the live-cell experiment, but essential to confirm whether or not it is EMT being observed, further studies on the available cell lines are to be performed. Using E-cadherin and Vimentin as

epithelial and mesenchymal markers, their expression level is to be studied, both at the total population but also at a single cell level. This is to be complemented with the same analysis on nMuMG samples that have been exposed to a traditional EMT inducer, TGF-β, used both as a standard, but also as comparison between what is assumed to be induced EMT versus the hypothesized spontaneous one.

Complementing this protein analysis is the mRNA study. In a similar manner to the previous experiment, the same samples are to be analyzed for their mRNA content, quantifying not only the two phenotypical markers, but also two additional genes, Snai1 and Snai2, both being noted as E-cadherin transcriptional repressors. With the addition of this study, a better understanding should be attained as to whether the three cell lines show evidence of EMT, and some inkling as to the similarities and disparities between what is being studied and what is known to occur when EMT is induced.

As a final objective, the effects of the cell density in culture are also to be glanced upon, both at the protein and the mNRA level. This should tell how each cell line responds to the cell crowding effect.

1.10 Techniques

This section gives a basic understanding of the experimental techniques used in this work, comprised of the several types of microscopies, the Western Blot and the RT-qPCR.

1.10.1 Microscopy

The invention of the microscope allowed us to observe what was previously unseen, to turn a world controlled by unknown quantities that were unobservable at the naked eye and turned it into one where these agents were visible, named and even in some cases prevented. It allowed scientists and researchers to observe and study microscopic creatures and structures.

The common microscope contains a set of lens kept in a certain orientation, a platform for stable viewing of samples and additional components to improve the image quality. More complex microscopes may operate in different manners, but the underlying scientific principle is maintained.

Besides several types of microscopes, this field is made more complex by the addition of several types of imaging techniques that can be used in said microscopes. The following summaries will focus on the three techniques used in this work, phase contrast and immunofluorescence imaging and also live-cell (or time-lapse) microscopy.

Phase contrast

Phase contrast was a very important achievement in the field of microscopy, as before this was discovered researchers were limited to brightfield microscopy, without the capacity to observe living beings. What happens is that living biological samples are transparent, they do not absorb light, which made it hard to observe with any detail under the microscope. However, Frits Zernike found a way to take advantage of both direct and refracted light to increase the quality and definition of unstained, transparent and/or colorless objects. Nowadays, there are two types of phase contrast imaging: positive

and negative. This imaging technique allows the ability to observe living organisms in their natural state, with high detail and good contrast, although it carries some disadvantages like lower resolution or distortions. Nonetheless, it is a basic feature in most modern microscopes [76].

Immunofluorescence

Immunofluorescence microscopy is an imaging technique used with a fluorescence microscope, the most used in the medical and biological fields. This type of equipment use high-powered light waves at a specific wavelength to excite electrons in specimens, so that during the relaxation, the same specimen releases the energy again in the form of light in the visible spectrum [77].

Immunofluorescence is used because very few biological structures naturally emit fluorescence in the visible spectrum. This evolved into the use of fluorescent stains, and eventually the use of antibodies as stains. In order to make this technique more accessible, a two step process was designed (indirect immunofluorescence) [78]. A primary antibody was developed in an animal, specific for the protein of interest. Then, a secondary antibody, this time specific for any protein produced by the animal that was used to collect the primary antibody, was created and attached with a fluorophore. The primary antibody attached to the protein that is being studied, the secondary attached to the primary, and when excited under a certain wavelength, the secondary protein emitted light at a longer wavelength. To execute that, besides a prepared sample, a fluorescence microscope is needed, and it must be able to emit light at the required wavelengths, but also capture the wavelengths of the light emitted by the sample. In a microscope, the capability to capture a certain wavelength is usually referred to as a channel, so the number of channels is usually the limit to how many proteins can be studied at once in a sample (not completely true, seeing as each channel must be sufficiently apart from the others on the light spectrum, in order to avoid overlap, which in the end also limits the total number of proteins that can be observed at once).

Live-cell microscopy

Time-lapse microscopy involves replicating the conditions verified in the cell culture incubators (temperature, CO_2 level and humidity) but in a ”portable” chamber that fits in with the microscopes sample holder. In this way, the cells can be imaged in a microscope without the need to fix them, keeping them alive. The benefit is that, by taking time-lapsed images (one image every x minutes/hours), a sequence can be created where the growth/movement/insert event of choice can be observed later on. It is most useful in motility studies, with a little less emphasis in growth studies (other techniques provide more accurate information in this area), and is being more and more used to study the occurrence of specific events, either induced ones, or even natural ones like cell division. Time-lapse microscopy can be used together with many imaging techniques, as long as a solution is found to obtain that type of images with live cells (for example, for observing a certain protein with fluorescence, cells can be manipulated into producing said protein with a fluorescence tag attached to it).

1.10.2 Western Blot

The Western Blot is an analytical technique developed for the study and identification of protein samples. In biological research sometimes it is of interest to know whether or not a specific protein is being expressed in a certain cell line under a specific condition, and if it is, how does it relate to the expression in another cell line/condition. For that, Western Blot analysis, along with immunostaining and ELISA are the more commonly used techniques.

What happens is that a sample containing a mix of proteins is ran in gel electrophoresis, where they will be separated (either by 3D structure for native proteins or by polypeptide length for denaturated ones). The proteins are then transferred to a membrane (most common are nitrocellulose or PVDF membranes), where they are stained with antibodies specific for the proteins of interest. These steps share similarities with indirect fluorescence staining, as a primary antibody is needed to link to the relevant proteins and a secondary one is needed to attach to the primary and supply the signal to be detected and further on analyzed (although the secondary antibodies in Western Blot are different than for staining, as there are several possible detection mechanisms).

With this, it can be used to identify whether or not a protein is being expressed in a sample, or even if multiple samples are being compared, the relative expression levels of the said protein can be plotted for each sample (through the use of a control protein that is known to have almost no variation in expression throughout the samples).

The technique involves several steps: sample preparation, gel electrophoresis, transfer, blocking, incubation with antibodies, detection.

The initial stage, sample preparation, involves releasing the proteins from the sample to be analyzed. To do that usually a combination of mechanical techniques like blending, homogenizing or sonication are used side-by-side with detergents, salts and buffers to lyse the sample (cells, tissue or even viruses) and obtain their protein content. Protease and phosphatase inhibitors are a must to ensure that the enzymes present in the sample don't degrade the proteins. Also common is to perform this stage at low temperatures to avoid any denaturation or degradation. Further operations can be added if the need is for the proteins of a specific cellular location or organelle.

The second stage, gel electrophoresis, is where the proteins in the sample are separated, and this can be done through a large amount of parameters. The most common method is SDS-PAGE, meaning using polyacrilamide gels to run the electrophoresis, with buffers loaded with sodium dodecyl sulfate. This keeps proteins in a denaturated state, removing their secondary and tertiary structure, allowing them to be separated by their molecular weight. The SDS confers them with a negative charge, so that when a current is run through it, they migrate to the opposite end of the gel where they are loaded (the positive charge). The smaller the protein, the faster the migration, meaning that when the electrophoresis is stopped, the proteins present in a sample are displayed in the gel in a gradient, according to peptide size. To be able to keep track of the protein size, some wells are loaded with ladders, commercial mixes of proteins with known sizes, and usually stained to form visible bands.

The transfer stage involves removing the proteins from the gel and passing them to a membrane, with the purpose of afterwards performing antibody detection. The most common method is electroblotting,

running a current to pull the proteins from the gel to the membrane. At the end, the proteins are displayed on a thin surface layer for detection.

After the transfer the membrane is always blocked for unspecifing binding, seeing as the membrane was chosen for its ability to bind to proteins, and antibodies also fall into that category. This would mean that the antibodies would bind to the membrane, invalidating any observable result. Blocking involves using a dilute solution of either Bovine Serum Albumine or non-fat dry milk in Tris Buffered Saline, along with low amounts of detergent, like Tween 20 for instance. The proteins in the solution will bind to the membrane everywhere there currently isn't any protein, removing any free membrane space that could bind to the antibodies during incubation, thus diminishing background noise.

Then comes the incubation with the antibodies. Similarly to immunofluorescence, this is usually done in a two step process, mainly because a signal amplification step may be added to the protocol, and also because it is cheaper. As such, the membrane is incubated in a TBS-Tween solution containing a primary antibody to attach to the protein of interest. Following this is an incubation in TBS-Tween containing the secondary antibodies (usually one for the ladder, which functions as its own primary, and one for the secondary). These secondary antibodies are species specific and contain the reporter that will allow for detection (it is cheaper to have a protein-specific antibody obtained from a certain species and a species-specific antibody that contains the reporter than having to make a protein-specific antibody with the reporter for every possible protein).

Detection can be done through a variety of methods, as long as the required equipment is present and the necessary antibodies are used. The most common methods are: colorimetric, chemiluminescent, radioactive and fluorescent detection. The chosen method was chemiluminescence, in which the membrane is further incubated in a substrate that in contact with the secondary antibodies and exposed to light will luminesce, which will then be captured by a digital camera. The protein bands are analyzed by densiometry, to correlate positively to protein amount, and a structural protein (protein whose expression does not vary significantly in the conditions studied) needs to be analyzed for each sample, in order to normalize the results for each sample. One sample is also usually chosen to act as control, to normalize other samples against it (relative quantification).

For analysis of multiple proteins, once the detection is done for one protein, the membrane is incubated with a stripping buffer that removes all the antibodies and even the blocking proteins. So, afterwards, blocking step must be redone for the next protein in the study to be detected with the necessary antibodies. Although it is possible to do stripping on nitrocellulose membranes, the PVDF membranes are sturdier and allow for easier stripping and more reuse before background noise makes it impossible.

1.10.3 RT-qPCR

The technique PCR (standing for polymerase chain reaction) is a technological duplicate of what occurs to DNA during cell mitosis. As such, the sample DNA is first denaturated from double to single strand, at which point the primers (specific to the genes of interest) attach to the DNA. Using an engineered DNA polymerase enzyme and free nucleotides as a substrate the replication reaction is initiated.

Using specific primers, there is only amplification of the selected areas (parts or the totality of the genes of interest). Seeing as DNA is a double strand, for each gene two primers are used, one for the 'forward' strand, the other for the 'reverse' strand. PCR occurs in cycles, and after each one, the amount of DNA in the mix doubles, as from each copy of DNA, a new one is created. These cycles are comprised of several stages: DNA denaturation, primer annealing, polymerase reaction (elongation); and the driving force behind this is the temperature at which the sample is subjected.

RT-PCR was developed with the thought of amplifying RNA instead of DNA, as research showed that eukaryotic organisms have added regulation mechanisms for gene expression. The discovery that the gene contained areas that didn't encode parts of the protein (introns) along with the coding areas (exons); that and the fact that during the removal of the introns the exons could be rearranged in different orders (splicing) lead to a need of amplifying RNA (as the genetic code there was the necessary one to translate the final protein). The technique was inspired in the action of retro viruses, which contain their genetic material in the form of RNA, and once inside the host cell, use an enzyme called reverse transcriptase, that turns RNA to DNA through base pair complementarity. And thus, again by emulating a living organism RT-PCR was developed.

The RT-PCR technique was further improved and directed when RT-qPCR was developed. The inner working is basically the same, only here a probe is added to the sample mix. There are many types of probes, however the common denominator is that they are used to report on the amount of genetic material present in a sample. As such, after each cycle, seeing as the amount of the DNA of interest doubles, so should the signal provided by the probe. From here, by defining a signal threshold, the number of cycles needed to reach said threshold is measured. This made it possible to perform either absolute or relative measurements of DNA quantity, as the value of the cycle threshold for a sample negatively correlates with the amount of the DNA of interest in said sample.

For absolute measurements a set of several representative samples with known amounts of DNA should also be analyzed parallel to the samples to be studied. With the values obtained for the representative samples, the cycle threshold can be plotted against the total DNA amount. Having this plot and the measurements for each gene of interest in the samples, the absolute expression of each gene can be quantified.

In relative quantification a reference/structural gene (one whose expression is constant under the conditions studied) is analyzed for all samples, parallel to the other genes of interest. This allows to obtain the relative expression of each gene to the reference one, in each sample. In addition, a control condition must be used, as this analysis requires a quantification relative to the reference gene, it also requires it to be relative to the control condition (to allow for comparison between conditions).

Chapter 2

Materials & Methods

2.1 Cell culture

Normal murine mammary gland cells (nMuMG) (#CRL-1636) were purchased from American Type Culture Collection (ATCC, Manassas, VA). M and E clone cell lines were obtained by single cell cloning from the parental line nMuMG, as explained in section 2.2.

All the cell lines were cultured in 10-cm culture dishes (Corning Inc., Corning, NY), using as substrate 10mL of Dulbecco's Modification of Eagle's Medium (DMEM, #10-013-CV, Corning Inc.) supplemented with 10% of fetal bovine serum (FBS, Hyclone, GE Healthcare Life Sciences, Logan, UT) and 100U penicillin and 100μg streptomycin (Sigma-Aldrich, St. Louis, MO). The culture environment was maintained at 37° C and 5% CO_2.

Upon reaching 70-80% confluence in the cell culture, media was removed and replaced with 2mL of calcium and magnesium-free Phosphate-Buffered Saline (PBS, #MRGF-6235, Growcells, Irvine, CA) for washing. Following, cells were detached by replacing the PBS with 1mL of trypsin (#T4049, Sigma-Aldrich) and incubated for 10 minutes at 37° C. Once cells detached from the culture dish, they were resuspended in 10mL of media.

Part of this cell suspension was used in maintaining the cell culture (usually 1mL of the suspension with 9mL of fresh media), while the remaining was plated in separate, to be used in the experiments performed.

Cell numbers were counted with a Countess cell counter (#C10227, Invitrogen, ThermoFisher Scientific, Waltham, MA).

2.2 Single cell cloning

Once nMuMG cells were detached with trypsin, they were ressuspended in GIBCO® Hanks' Balanced Salt Solution (HBSS, ThermoFisher Scientific). The cell concentration for the suspension was estimated, and then adjusted through serial dilution up to 1 cell/10μL. Using a 96-well plate (BD Biosciences, San Jose, CA), a 10μL droplet was added to a well, and immediately observed under an

Eclipse TE-200 microscope (Nikon, Melville, NY). Wells with a single cell were topped to 200μL with media.

Afterwards cells were once again observed under the microscope and their phenotype annotated as epithelial or mesenchymal depending on their morphology. From a well containing an epithelial cell, and through clonal expansion the E clone cell line was obtained. The same for the M clone cell line, but this time by clonal expansion of a mesenchymal cell.

2.3 Low density colony analysis

Cells were seeded in 6-well plates (#353046, Corning Inc.) at around 0,1 to 0,3% confluence, and then left to expand for 3 days.

Following expansion, cells were fixed with 4% paraformaldehyde solution (#15710, Electron Microscopy Sciences, Hatfield, PA) for 12 minutes at room temperature. PBS rinse was conducted three times after the fixing step.

Images were collected with a ORCA-3CCD Digital CCD camera (Hamamatsu, Hamamatsu City, Shizuoka Prefecture, Japan) installed on an Eclipse TE-2000 inverted microscope (Nikon) with a Plan Fluor 10x lens (Nikon), a motorized stage, and motorized excitation and emission filters (Prior Scientific) controlled by Nikon NIS Elements. Two hundred twenty five (15-by-15 grid) fields of views were consistently generated using the software Nikon NIS-Elements. The size of the image acquired from the camera was 1344 x 1024 pixels, and the pixel size was 0,65 μm for a 10x objective. Our ΔX and ΔY for image capture was 780 μm and 600 μm respectively to allow 10% overlay between adjacent fields. The total size of the scanning region (225 images) was ~105,3 mm^2. One channel (phase contrast) was collected.

All of the data was obtained by manually tracing colonies in ImageJ software. At least 11 colonies were analyzed per different technical/biological repeat (i.e. different cells), with a minimum total of ~67 colonies for each tested condition.

The absolute values were then converted to relative frequencies, and plotted in GraphPad Prism, displaying the average % of each fraction on all three samples, along with the SEM, obtained by dividing SD by number of measurements for each sample.

2.4 Live cell imaging

Cells were seeded in to 6-well plates (Corning Inc.) at around 1% confluence and then left to expand. The next day the plate was set into the Live Cell™(Pathology Devices, Inc., Westminster, MD) incubation system.

Images were collected with a ORCA-3CCD Digital CCD camera (Hamamatsu, Hamamatsu City, Shizuoka Prefecture, Japan) installed on an Eclipse TE-2000 inverted microscope (Nikon) with a Plan Fluor 10x lens (Nikon), a motorized stage, and motorized excitation and emission filters (Prior Scientific)

controlled by Nikon NIS Elements. Two hundred twenty five (15-by-15 grid) fields of views were consistently generated using the software Nikon NIS-Elements. The size of the image acquired from the camera was 1344 x 1024 pixels, and the pixel size was 0,65 μm for a 10x objective. Our ΔX and ΔY for image capture was 780 μm and 600 μm respectively to allow 10% overlay between adjacent fields. The total size of the scanning region (225 images) was ~105,3 mm^2. One channel (phase contrast) was collected.

The cells were left under observation for 36 to 72h, during which, images were obtained every 7,5 minutes.

For analysis, a number of cells from the initial time-point were selected and traced. These same cells and their progeny were tracked and traced throughout the whole time of capture. The data presented shows certain time-points, to make it visible the changes occurring in some of the cells analyzed.

2.5 TGF-β treatment

For TGF-β treatment, human TGF-β1 (#T7039, Sigma-Aldrich) was purchased.

4 hours after plating the cells, TGF-β1 was added to the media at a concentration of 1 to 5 ng/mL. Samples subjected to this treatment were incubated at 37° C and 5% CO_2 for 24 or 48 hours (exposure time).

2.6 Immunofluorescence analysis

Cells were grown in Millicell EZ SLIDE 8-well glass, sterile (#PEZGS0816, EMD Millipore, Billerica, MA) and upon reaching the desired density were fixed with 4% paraformaldehyde solution (Electron Microscopy Sciences) for 12 minutes. After fixing, cells were permeabilized with 0.1% Triton X-100 (#X-100, Sigma-Aldrich) for 10 minutes and blocked for nonspecific binding with PBS supplemented with 10% of Bovine Serum Albumine (BSA, #A7096, Sigma-Aldrich) for 30 minutes. All of the steps mentioned previously were performed at room temperature.

For staining, cells were incubated with specific concentration of dye/antibodies diluted from stock solution for 1h (at 37° C) or overnight (at 4° C).

- Nuclear DNA was stained with Hoechst 33342 (#14533, Sigma-Aldrich) at 1:50 dilution for 1h.
- Actin was stained with Alexa Fluor® 568 Phalloidin (#A12380, ThermoFisher Scientific) at 1:50 dilution for 1h.
- E-cadherin was stained with E-Cadherin (24E10) Rabbit mAb primary antibody (#3195, Cell Signaling Technologies, Danvers, MA) at 1:50 dilution overnight, followed by a Goat anti-Rabbit IgG (H+L) Secondary Antibody, Alexa Fluor® 647 conjugate (#A21244, ThermoFisher Scientific) at 1:200 dilution for 1h.

- Vimentin was stained with a Anti-Vimentin antibody [VI-10] primary (#ab20346, Abcam, Cambridge, MA) at 1:1000 dilution overnight, followed by a Goat anti-Mouse IgG (H+L) Secondary Antibody, Alexa Fluor® 488 conjugate (#A11001, ThermoFisher Scientific) at 1:200 dilution for 1h.

The order of staining went as such: primary E-cadherin and Vimentin antibodies were used simultaneously in the first step, the secondary antibodies, the DNA and Actin dyes were all used at the same time following the previous step. PBS rinse was conducted three times between each step.

Fluorescent images were collected with a Zyla 5.5 sCMOS (Andor Technology Ltd., Belfast, UK) camera installed on an Eclipse N*i*-U upright microscope (Nikon) with a CFI Plan Apo VC 20x lens (Nikon), a motorized stage, and excitation and emission filters (Prior Scientific) controlled by Nikon NIS Elements. Sixteen (4-by-4 grid) fields of views were consistently generated using the software Nikon NIS-Elements. The size of the image acquired from the camera was 2560 x 2160 pixels, and the pixel size was 0.32 μm for a 20x objective. The ΔX and ΔY for image capture was 737 μm and 622 μm respectively to allow 10% overlay between adjacent fields. The total size of the scanning region (16 images) was ~7.3 mm^2. Four different wavelengths (342nm: blue, 488nm: green, 568nm: red and 647nm: far red) were collected for every field of view.

The data analyzed was obtained through the use of a custom-built MatLab algorithm for the purpose of tracing cells and gathering phenotypical data from fluorescence images. Typically ~2000 cells were analyzed per samples' technical repeat (i.e. duplicate samples for each cell line) for a total of ~4000 cells for each tested condition.

The data obtained, E-cadherin signal intensity, Vimentin signal intensity and cell area were analyzed in GraphPad Prism. All the intensity values were used to create a scatterplot, showing E-cadherin in the X-axis and Vimentin in the Y-axis. A plot was built for all three datasets, containing the average value for each sample and the SEM, obtained by dividing the SD by the square root of the number of cells analyzed. Using Kruskal–Wallis test all conditions were compared to the untreated nMuMG cell line using significance criteria p-value$\leq$0,05. *:$p\leq$0,05; **:$p\leq$0,01; ***:$p\leq$0,001; ****:$p\leq$0,0001.

The same set of data was obtained for the experiment testing the density effects. Plots were built, three (one for each dataset) for each cell line, showcasing the average value and SEM for high, medium and low densities. Using Kruskal–Wallis test each density was compared to the remaining two using significance criteria p-value$\leq$0,05.

2.7 Western Blot

Cells were plated in 10-cm tissue culture dishes and incubated until 80% confluence.

Sample preparation required extracting the totality of the proteins in the cell culture, and for that cells were washed with cold PBS and lysed. This step involved scraping the culture dish after the addition of 300μL Radioimmunoprecipitation assay buffer (RIPA, #R0278, Sigma-Aldrich) containing 1mM Phenylmethylsulfonyl fluoride (PMSF, #10837091001, Roche, Basel, Switzerland), Protease inhibitors (PI, Roche), and Dichlorodiphenyltrichloroethane (DDT, #R0861, ThermoFisher Scientific). Following the extraction the resulting suspension was kept on ice for 30 minutes, being vortexed every 10 minutes.

Alliquoting was performed to avoid damaging all of the sample with freeze-thaw cycles, and the vials were stored at -80° C.

Protein concentration was measured with a Bradford assay, using Quick Start Bovine Serum Albumin 2 mg/ml Std. (#5000206, Bio-Rad, Hercules, CA) to obtain solutions with known protein concentrations. A mix composed of 1mL Bradford reagent and 100μL of RIPA buffer and sample/protein standard (19:1 ratio for samples to be tested, ratios differed for the protein standards). Protein concentration was obtained by measuring solution absorbance at 595nm, in a Spectrophotometer.

Protein samples were boiled at 100° C for 5 minutes, and then loaded into a Mini-Protean TGX precast 12% polyacrylamide gel (#456-1044, Bio-Rad). For each sample, 30 μL were loaded in the gel, this being a mix of the sample, RIPA buffer and Laemlli 6X (Bio-Rad). The composition of each suspension was defined so that Laemlli concentration was 1X and so that an equal amount of protein, 25μg, was loaded for each sample (the volume of sample in the suspension was based on the quantification obtained by the Bradford Assay).

The gel was placed in the support and fit into the container, which was afterwards filled with 10X Tris/Glycine/SDS Electrophoresis Buffer (TGS, #1610772, Bio-Rad) diluted to 1X. SDS-PAGE was run at 120 V (constant voltage) for 75 minutes at 4° C to separate protein samples before executing the transfer from the gels onto PVDF membranes (#162-0219, Bio-Rad).

The transfer step was done placing the gel and the PVDF membrane, soaked in Methanol (Sigma-Aldrich), in between two sheets of Blot paper (#1703966, Bio-Rad) soaked in TransBlot Turbo Transfer Buffer in a TransBlot Turbo Transfer System (#170-4155, Bio-Rad), and running the High Molecular Weight program.

Membranes were blocked with 5% of Blotting-Grade Blocker (#170-6404, Bio-Rad) in Tris-Buffered Saline (#46-012-CM, Corning Inc.) with 1:1000 Tween 20 (P9416-100ML, Sigma-Aldrich) for 1 hour at constant agitation and then incubated with primary antibody at 1:1000 dilution (E-Cadherin (24E10) Rabbit mAb or Anti-Vimentin antibody [VI-10]) overnight at 4° C.

Following a washing off of the primary antibody, a 2 hour incubation at room temperature with secondary antibodies: Anti-rabbit IgG, HRP-linked Antibody (#7074, Cell Signaling Technologies) or Goat Anti-Mouse IgG H&L (HRP) (#ab6789, Abcam), and ladder secondary (#1610380, Bio-Rad) was performed. After, blots were developed with an Clarity™ Western ECL Blotting Substrate (#1705061, Bio-Rad).

Staining for β-Actin required only one step, incubating the membrane with ladder secondary and beta Actin Antibody HRP (#HRP-60008, ProteinTech Group, Chicago, Il) for 2h at room temperature under constant agitation.

In between each protein analysis, membrane was incubated at room temperature for 2 hours with Gentle Review Stripping Buffer (#N552-500ML, Amresco, Solon, OH), followed by a new blocking step with TBS-Tween 5% Blotting-Grade Blocker for 1 hour.

Membranes were imaged with a Bio-Rad ChemiDoc XRS imaging system, and all bands were quantified by relative intensity using Bio-Rad Image Lab software.

The data obtained, E-cadherin band intensity and Vimentin band intensity were analyzed in Graph-

Pad Prism. A plot was built for both datasets, containing the value for each sample. As a result of having only one measurement, no error was calculated and no statistical analysis was performed.

2.8 RT-qPCR analysis

Cells were grown to 80% confluence in 6-cm dishes (Corning Inc.), after which they were washed with cold PBS. This was followed by addition of 750μL of TRIzol (#15596-018, Ambion, ThermoFisher Scientific). After scraping for about 1 minute, cells were collected in SealRite 1,5 mL microcentrifuge tubes (#1615-5500, USA Scientific, Ocala, FL), where 750μL of 200 Proof Ethanol was added (#111000200, Pharmco-Aaper, Brookfield, CT).

Total RNA was extracted from the samples through the Directzol RNA Prep protocol (#11-330, Genesee Scientific, San Diego, CA). The procedure was as follows:

1. Load the TRIzol/Ethanol mix in an RNAse free column;
2. Add 400μL RNA Wash Buffer;
3. Add 80μL solution (5μL of DNAse in 75μL of DNA Digestion Buffer) and incubate for 15 minutes at room temperature;
4. Add 400μL RNA Pre-Wash Buffer;
5. Repeat step 4;
6. Add 700μL RNA Wash Buffer;
7. Switch the base column for a 1,5 mL microcentrifuge tube;
8. Add 50μL DNAse/RNAse-Free Water;

In between each step described the column was centrifuged at 14.000 rpm for 1 minute (after step 6 the centrifugation time was 2 minutes). Until the last step, when the RNA was eluted, the elutions were deposited in a waste container. The 1,5 mL microcentrifuge tubes with the RNA samples were stored at -80° C.

Total DNA/RNA concentration was measured in a Nanodrop (ThermoFisher Scientific).

For cDNA conversion, 1μg of each RNA samples (calculated using the RNA amount measurements) was loaded into .2mL 8-Well PCR Strip Tubes (#27-125U, Genesee Scientific), supplemented with 1μL of Reverse Transcriptase, and 4 μL of 5X Reverse Transcriptase buffer, both reagents part of iScript cDNA Synthesis Kit (#170-8891, Bio-Rad) . Volume was raised to 20μL by adding the necessary amount of Ultra Pure Water (autoclaved double distilled water).

The reverse transcription reaction was performed in a Techne Genius thermocycler (Techne, Burlington, NJ).

For RT-PCR, 96-well plates (#AB-0990, ThermoFisher Scientific) were used, each well containing 1μL of the required primers, 2μL of the cDNA samples, 7μL of Ultra Pure Water and 10μL of Maxima

SYBR Green/ROX qPCR Master Mix 2X (#K0222, Life Technologies, ThermoFisher Scientific). Triplicates were done for each sample condition. The plates were sealed with Microseal® 'B' Adhesive Seals, Optical (#MSB1001, Bio-Rad)

Primer sequences were as follows. For mouse E-cadherin, forward: 5' TCGGAAGACTCCCGATTCAAA 3'; reverse: 5' CGGACGAGGAAACTGGTCTC 3' (amplifies a fragment of 95 bp). For mouse Vimentin, forward: 5' TCCACACGCACCTACAGTCT 3'; reverse: 5' CCGAGGACCGGGTCACATA 3' (amplifies a fragment of 95 bp). For mouse Snail (Snai1), forward: 5' GTCCAGCTGTAACCATGCCT 3'; reverse: 5' TGTCACCAGGACAAATGGGG 3' (amplifies a fragment of 106 bp). For mouse Slug (Snai2), forward: 5' CAGCGAACTGGACACACACA 3'; reverse: 5' ATAGGGCTGTATGCTCCCGAG 3' (amplifies a fragment of 111 bp). For mouse glyceraldehyde-3-phosphate dehydrogenase (GADPH), forward: 5' AGGTCGGTGTGAACGGATTTG 3'; reverse: 5' GGGGTCGTTGATGGCAACA 3' (amplifies a fragment of 100 bp).

In order to obtain the efficiency of the selected primers, RT-qPCR was run with a single sample at different concentrations (initial concentration, 1:2, 1:4 and 1:8 dilutions). From there the cycle threshold was plotted against relative DNA concentration and the curve was calculated.

To confirm primer quality the result of RT-qPCR was subjected to electrophoresis. A laboratory prepared Tris/Acetic acid/EDTA (TAE) buffer solution with 1% agarose (#E-3119-500, GeneMate, Bioexpress, Kaysville, UT) and Ethidium Bromide (#E1510-10ML, Sigma-Aldrich) at 1:10000 dilution was prepared and boiled under flame. While hot, the solution was poured to a gel frame in a HORIZON® 58 Gel Casting System (#21065-040), and left to cool and solidify for 30 minutes at room temperature.

For each primer, a sample was prepared, containing 2 μL of 6X DNA Gel Loading Dye (#R0611, ThermoFisher Scientific), 200 ng of DNA (volume calculated after measuring the cDNA concentration) and enough Ultra Pure Water to make a total volume of 12 μL.

Samples and 1μL of the Thermo Scientific GeneRuler 1 kb DNA Ladder (#SM0313, ThermoFisher Scientific) were loaded in the wells of the gel in a container filled with TAE buffer, and electrophoresis was run at 90V for 30 minutes in a Horizontal Gel Electrophoresis System.

The gel was imaged with a Bio-Rad ChemiDoc XRS imaging system, under UV light.

The data obtained, cycle threshold values for E-cadherin, Vimentin, Snai1 and Snai2 genes were converted to relative gene expression using the Pfaffl method and the values of cycle threshold for reference gene GAPDH. The gene expression values for all four genes were analyzed in GraphPad Prism. A plot was built for each dataset, containing the average value for each sample and the SEM, obtained by dividing the SD by the square root of the number of measurements. Using 2-way ANOVA all conditions were compared to the untreated nMuMG cell line using significance criteria p-value$\leq$0,05. *:$p\leq$0,05; **:$p\leq$0,01; ***:$p\leq$0,001; ****:$p\leq$0,0001. The data obtained from low cell density cultures for the experiment testing the density effects was subjected to the same treatment.

Chapter 3

Results and Conclusions

3.1 nMuMG: an epithelial cell line

The normal Murine Mamary Gland cell line (nMuMG) is an epithelial cell line, originating from the mamary gland tissue of an adult NAMRU mouse [79]. It is only one of many epithelial cell lines derived from this type of tissue. The cells exhibit features of secretory epithelium, along with intermediate junctions. This cell line is used frequently in studies of cell adhesion, thus making this one of the most used cell lines in studying the EMT process.

From brightfield microscopy observations of this cell line in culture, however, distinctly shaped cells can be easily observed. Some are present in colonies, showing a more rounded appearance and having straighter edges, while others are found as single cells, with more elongated morphologies and protuberances along the edges.

These differences can be observed in Figure 3.1, images that were obtained by immunocytochemistry, staining this cell line for nucleus and actin.

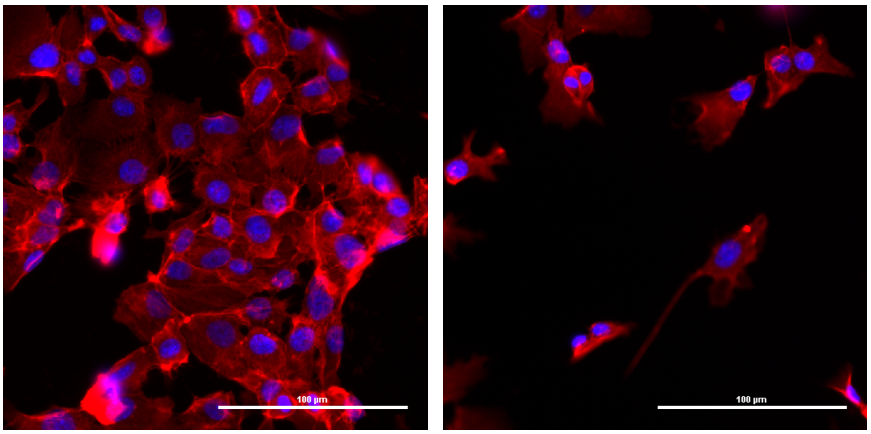

Figure 3.1: Representative images of the nMuMG cell line, obtained by immunocytochemistry (staining for actin and the nucleus). It can be observed that cells are present in colonies and show a more rounded morphology (left), while others appear detached and also present a more elongated shape along with protrusions (right).

These two distinct morphologies are consistent with what is known of epithelial cells (rounder cells

existing in colonies) and mesenchymal ones (growing as elongated single cells).

As these observations come from the original unmodified cell line, it is evidence in favor of a previously made hypothesis: EMT is a process that can occur spontaneously at a single cell level, at the least in the nMuMG cell line.

From these results, two new cell lines had already been established. By serial dilution, a nMuMG culture had been separated into single cells. By clonal expansion of a cell exhibiting epithelial phenotype, the cell line E clone was obtained. Simultaneously, from a mesenchymal cell, the M clone cell line was obtained.

From these three cell lines, performing low density colony analysis, a rough estimate was obtained for the amount of epithelial and mesenchymal cells in each cell line. This was done using phase contrast microscopy, obtaining 225 fields of view (a 15-by-15 grid), as depicted in Figure 3.2).

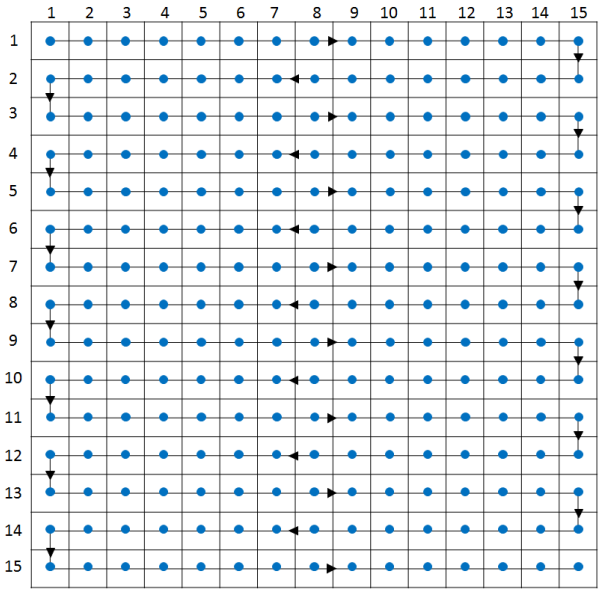

Figure 3.2: A representation of the grid scheme applied in the low density grow-outs to obtain a larger field of view. Each square represents a field of view, centered around a blue circle. The arrowed path that links all the circles is the order the images were obtained, useful for stitching afterwards. It is not displayed but in between each adjacent square there is an overlap of 10%, so that when stitching the final image there is less likelihood of blank spaces.

Afterwards, these 225 fields of view were stitched in a single large image, for each of the samples, which can be observed in Figure 3.3.

For analysis, the cell colonies in each image were traced, and annotated as either epithelial or mesenchymal according to the predominant morphology in said colony. Figure 3.4 provides an example to better understand what is considered to be an epithelial and a mesenchymal colony.

The absolute frequency of each type of colony was recorded, and the relative frequencies of epithelial and mesenchymal cells were calculated for the conditions studied. The final results are displayed in Figure 3.5.

The results here observed support the theory that the EMT process is not something that only occurs when it is needed (induced), but that it is indeed a background process (spontaneous). That is proven not only by the fact that nMuMG cell line samples contain evidence of both epithelial and mesenchymal

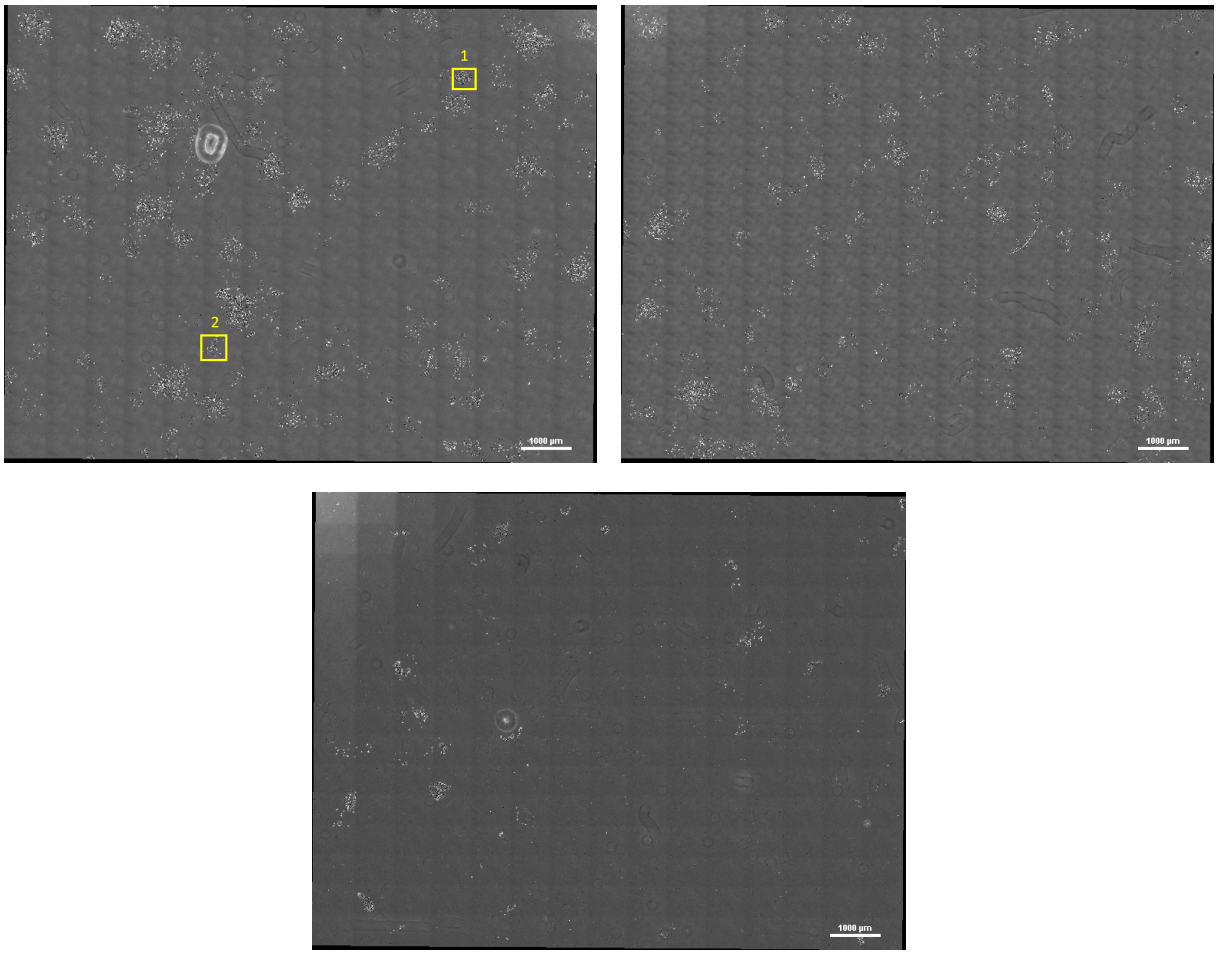

Figure 3.3: Low density phase contrast images for the nMuMG (top left), M clone (top right) and E clone (bottom) cell lines, large images stitched from the 225 fields of view. The yellow rectangles numbered 1 and 2 encompass an epithelial and a mesenchymal colony, respectively, that can be seen at an increased size in Figure 3.4

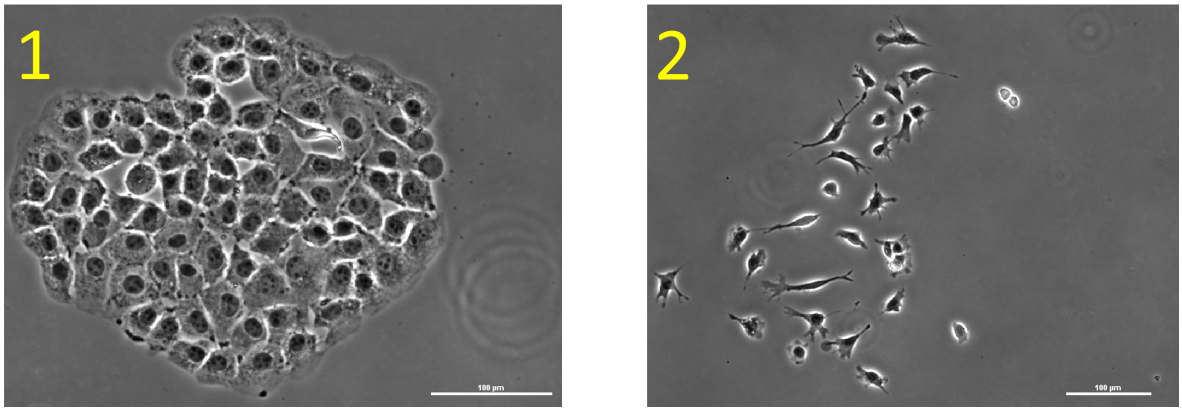

Figure 3.4: Representative images of an epithelial (1) and mesenchymal (2) colonies

cells, but the clone cell lines, obtained by clonal expansion of an epithelial and a mesenchymal cell, respectively, also show evidence of both types of cells.

Ideally this analysis would be done cell by cell (for better accuracy) and also at higher cell densities, so as to compare if the composition of a sample in epithelial and mesenchymal cells is affected by such conditions. However, with phase contrast it becomes a very laborious process, as this type of

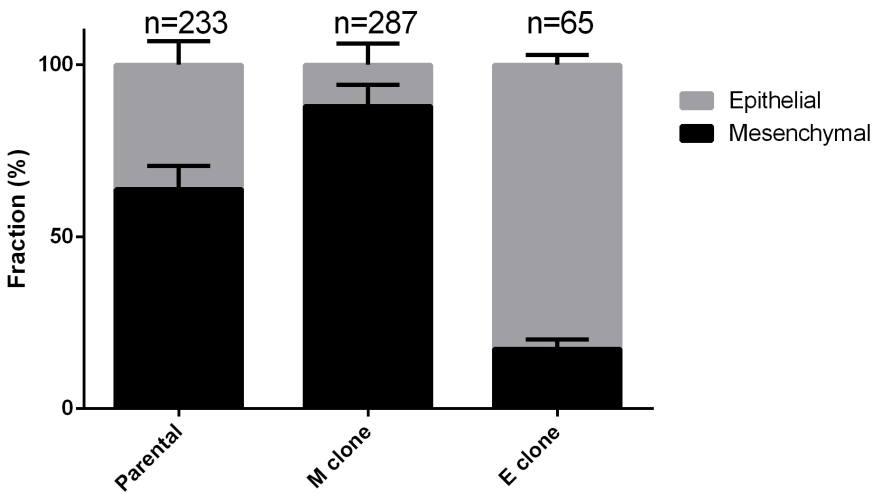

Figure 3.5: Result of colony analysis for the low density grow-outs performed for all three cell lines.

microscopy image is not suited for automated analysis (and even if it were, the higher the cell density, the harder the analysis is, automated or by hand).

3.2 Tracking the transition

The EMT process has been suspected of occurring spontaneously at a single-cell level, working similarly to what is defined in thermodynamics as a reversible process (as seen in Figure 3.6), without the need for external induction (addition of molecules, like TGF-β, Snai1, Snai2, ZEB, microRNA-200, *etc.*, to create a stimulus). The hypothesis is that, once again similar to thermodynamics, the purpose of this process occurring as described is to maintain a balance in the proportion of epithelial and mesenchymal cells.

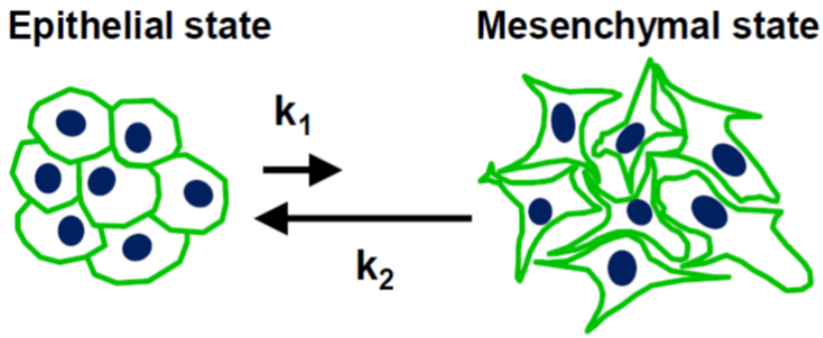

Figure 3.6: Representative design of the hypothesis created for the EMT process. It is thought that similarly to a thermodynamical reversible reaction that this process is composed of two states, between which cells spontaneously change; and each transition has a kinetic reaction associated with it. The idea that the M to E transition occurs at a higher rate then the E to M one stems from observations that the E clone cell line grows at a lower rate than the M clone one; and the fact that nMuMG cell line contains almost a 50/50 mix of both types of cells

Time-lapse microscopy is the act of, over a period of time, taking images of a sample (while maintaining it in culture, deriving the term live-cell microscopy) to observe the occurrence of cellular phenomenons. Caveats associated with this type of microscopy include: the process to be observed/studied must occur on an adequate timescale to be able to be recorded; it must also be observable (if whatever is being studied leads to no appreciable change that can be reported, some auxiliary/alternative method must be designed to that effect); and unless there is a clear way to distinguish each individual cell from the others, the cell movement may lead to tracking problems (sometimes two cells ”cross” each others' paths and it can be difficult to track which is which afterwards), which may be further compounded upon when the culture reaches higher density.

Assuming TGF-β exposure accurately represents the naturally occurring transition from epithelial to mesenchymal (which might not be a completely correct assumption, seeing as it is not yet a fully understood process [80]), it is known that the exposure must be for at least 24 to 48 hours, so it is a good starting point on a rough timescale for the EMT transitions.

The experiment was designed to run for around 48 to 72 hours, to minimize the probability that the transitions could be interrupted during observation. It also has a secondary purpose of checking if these transitions occur during interphase or mitosis.

A 15-by-15 image grid pattern was acquired, with 7,5 minutes of lapse between each time-point. Using the software ImageJ, a sequence of each field of views' time-lapsed images was obtained and

observed. No stitching was done in this set of data as it would dramatically increase the computational power required and also make it more difficult to track the cells. On the other hand, without stitching, some of the cells observed, the ones that moved towards the edges of a field of view are discarded from the analysis.

To better relate what was observed, the data will be presented as a selection of relevant time-points for one field of view, where a small number of cells (suspected of having undergone EMT), selected and labeled from the first time-point will be tracked, traced and annotated throughout the entire time it was recorded. This can be observed in Figure 3.7.

It is not easily observable here, as the images displayed in Figure 3.7 were carefully selected, but it was somewhat difficult to uncover cells going through this change in the midst of all the cells in all the fields of view captured. But, in the images presented, there were cells undergoing significant morphological changes, here equated with phenotypical changes between epithelial and mesenchymal states. As such, this supports the theory of EMT being a spontaneously occurring phenomenon, instead of a program that is ready to be induced when needed.

Additionally, it is worth noting that cell division was not involved in the EMT process, meaning it does not appear to occur through asymmetrical cell division. It was observed that epithelial-looking cells when dividing gave rise to two epithelial-looking cells, and the same for mesenchymal ones. This would mean that EMT occurs during interphase.

However, the experiment performed presents some issues: the phase contrast data gives nothing but morphological data, which is not exactly an accurate description of epithelial or mesenchymal morphology; and the cell density at which it was performed brings out a difficulty with this type of approach, two cells that cross paths may make it difficult to keep track of both of them, further compounded on by the fact that in many groups of cells, sometimes division would occur for all cells at the same time, making it even harder, at the end, to know which cell came from which progenitor. What this means is that while what is presented here is considered evidence supporting the theory of EMT being a spontaneous process, there is the possibility that said inference is in error.

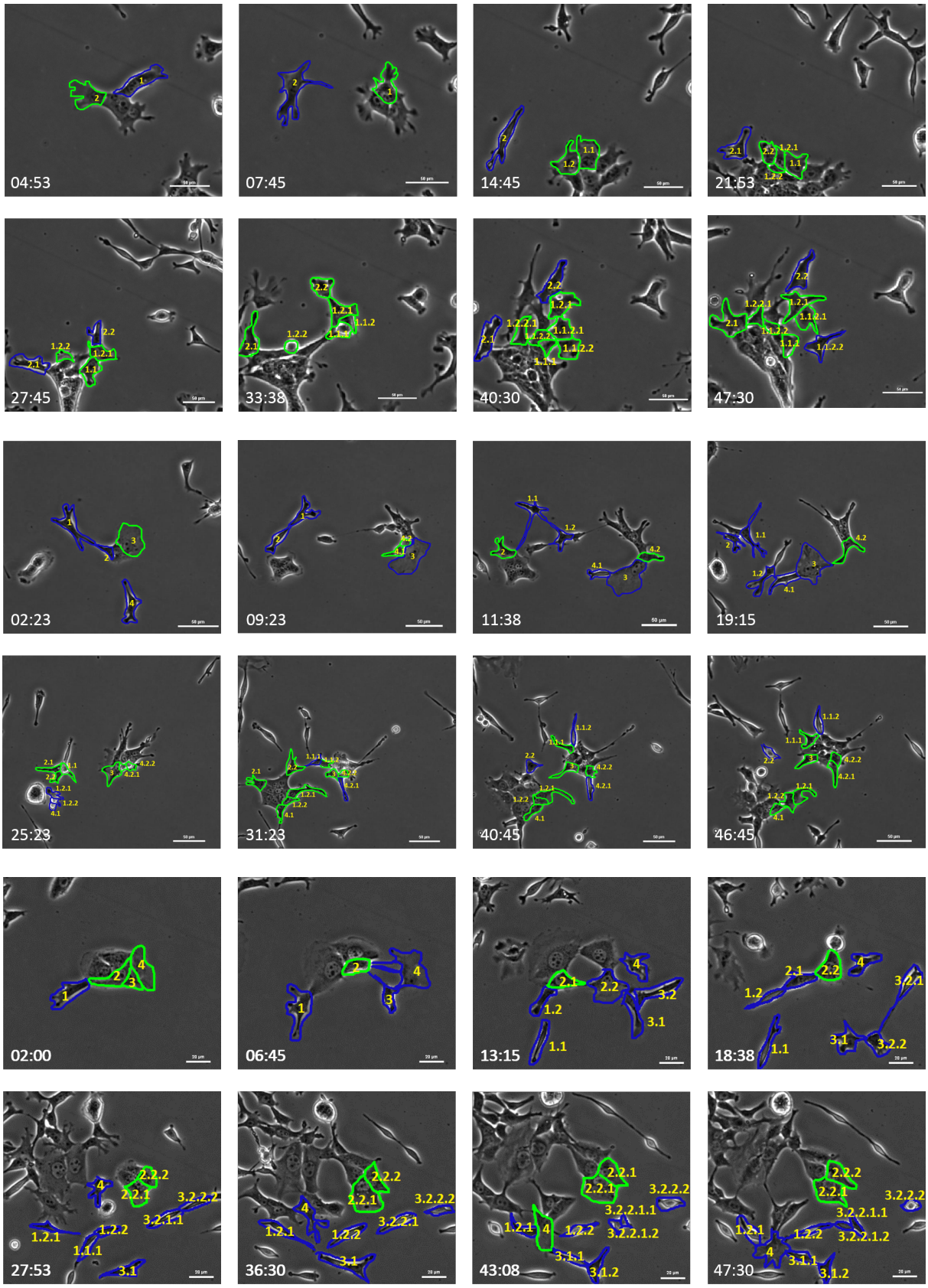

Figure 3.7: Image sequences of nMuMG sample throughout 24h. In the left lower corner of each image is the time at which it was captured. The three sequences correspond to three representative fields of view where it was suspected that EMT was observed. After capture and analysis, a small number of representative cells were chosen from the initial time-point, then labeled and traced. For further time-points, those same cells and progeny were tracked as best as possible and traced. The labeling was done through the assignment of a number to each cell. Progeny from a cell are labeled with the original number plus the addition of .1 or .2 to the new cells created. In green are cells with epithelial morphology while blue represent the mesenchymal ones.

3.3 Epithelial and Mesenchymal: molecular markers

As shown previously, in the cell lines being studied one can observe not only epithelial-like but also mesenchymal-like morphologies. However, morphological data alone is not enough to be able to confirm that cells belong to one or another state.

To do so, molecular markers were used. As mentioned in section 1.7, through years of study on both the epithelial and mesenchymal phenotypes, several molecules have been found to be up or downregulated in these two cell types [3]. In this case, as an epithelial marker, the protein E-cadherin was chosen, while for the mesenchymal phenotype it was the protein Vimentin.

The protein E-cadherin, the "founder" of the cadherin superfamily, is a cell-cell adhesion molecule, dependent on the presence of Ca^{2+} for its function. It has been a long studied protein, with implications in the behavior of epithelial cells, in tissue formation, and even in cancer suppression [81]. This is one of the most, if not the most used proteins in EMT studies as a marker of the epithelial phenotype.

Vimentin is an intermediate filament protein that comprises the majority of a mesenchymal cells' cytoskeleton. It has been implicated of being an organizer protein, ensuring the multitude of proteins involved in functions such as attachment, migration and cell signaling all work as supposed to. It has been gaining ground as one of the most common mesenchymal cell marker [82].

In order to obtain a positive control, the nMuMG cell line was exposed to TGF-β, in order to induce EMT and compare the results from this sample with the data obtained from the other cell lines.

3.3.1 Western Blot

As a confirmation step, a Western Blot was performed, in order to study the three cell lines at the total population level. The cells were seeded and left to grow to high density, upon which the proteins were extracted. After extraction, the total protein amount was measured with the Bradford Assay.

For that, using BSA at known concentration, several solutions were created and measured absorbance at 595nm. From this, absorbance was plotted against total protein concentration (which in the standards used, was the only protein present), as can be observed in Figure 3.8.

The curve can be described by Equation 3.1, which was then used to measure protein concentration in the samples, the results are displayed in Table 3.1.

$$A_{595nm} = 4,213 \times Total\ Protein\ Concentration\ (mg/mL) \tag{3.1}$$

Table 3.1: Total Protein Concentration results obtained the samples tested with the Bradford Assay method

Sample ID	A_{595nm}	Sample Concentration (mg/mL)	Actual Total Protein Concentration (mg/mL)
nMuMG	0,535	0,127	2,540
nMuMG + 1ng/mL TGF-β 24h	0,413	0,098	1,961
nMuMG + 5ng/mL TGF-β 24h	0,333	0,079	1,581
M clone	0,715	0,170	3,394
E clone	0,758	0,180	3,598

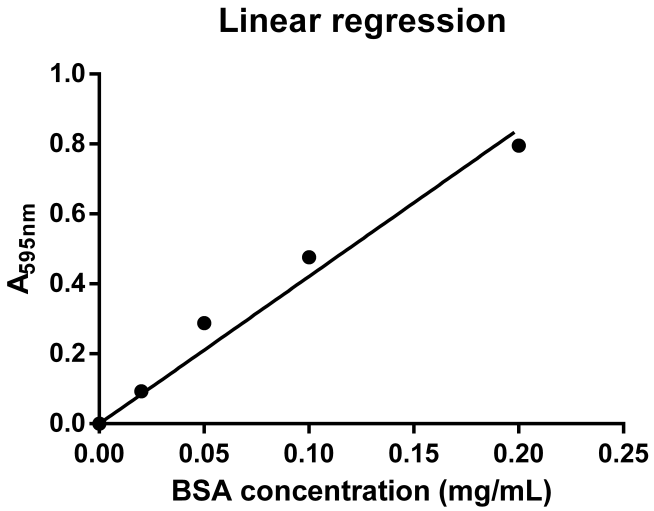

Figure 3.8: Plot of absorbance at 595nm versus BSA concentration.

Once amount was known, samples were prepared so that only 25μg of total protein was loaded into a pre-cast 12% poliacrilamide gel, where electrophoresis was ran at 120V for around 75 minutes. Once finished, the gel was inserted into a TransBlot Turbo so that the proteins could be transferred from the gel to a PVDF membrane.

The membrane was blocked with TBS Tween 5% milk for 1h at room temperature, after which it was stained for E-cadherin, Vimentin and Actin (in between each protein staining the membrane was stripped for 2h at room temperature and blocked again with TBS Tween 5% milk).

The images obtained through the staining of each protein are displayed in Figure 3.9.

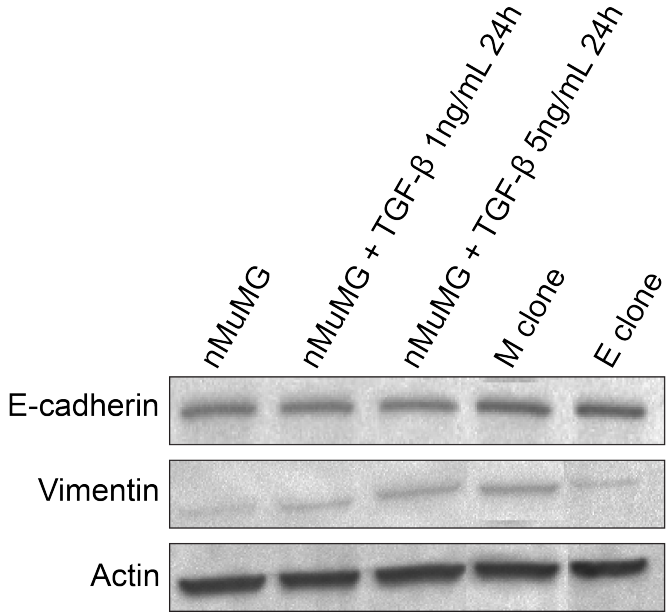

Figure 3.9: Compilation of the E-cadherin, Vimentin and Actin (from top to bottom) bands observed in the Western Blot experiment for the samples indicated above the bands.

The images obtained were analyzed with the Bio-Rad ImageLab software complementary with the

Bio-Rad ChemiDoc XRS imaging system so that band quantification could be done.

For the quantification results, the data obtained was all relative to a control sample, defined as the nMuMG untreated cell line. Furthermore, the Actin protein levels were used as a reference for each sample, meaning that they were used to normalize the protein level of both E-cadherin and Vimentin. The quantification was performed by analyzing the density of each band present in the Western Blot.

The initial results can be seen in the Table 3.2, while the final set of data (normalized to Actin levels) is displayed in Figure 3.10.

Table 3.2: Western Blot Preliminary Results: band intensity analysis, relative to untreated nMuMG cell line (not yet normalized against Actin content)

	nMuMG	nMuMG + 1ng/mL TGF-β 24h	nMuMG + 5ng/mL TGF-β 24h	M clone	E clone
E-cadherin	1	0,87	0,73	1,08	1,06
Vimentin	1	1,61	3,46	3,81	0,87
Actin	1	0,95	1,12	1,2	1,09

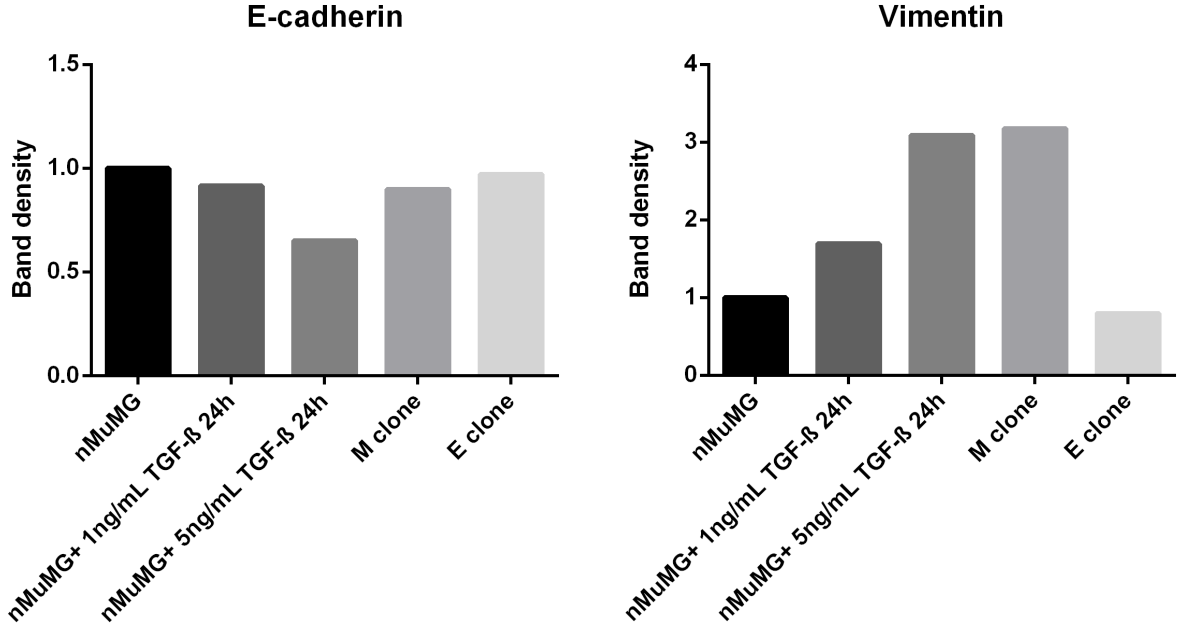

Figure 3.10: E-cadherin (left) and Vimentin (right) protein expression levels in various samples. Values relative to untreated nMuMG cell line.

The results obtained from the Western Blot seem for the most part to agree with what has already been published and with what was previously observed. The effects of the addition of TGF-β to the nMuMG culture are as expected: decreased E-cadherin levels and increased Vimentin ones. The variation seems to be the fact that the E-cadherin decrease appears to be minimal (with 1ng/mL it is ~0,9-fold, while it only goes to ~0,65-fold with 5ng/mL). The M clone cell line also registers a slight E-cadherin decrease (around the levels of the 1ng/mL TGF-β sample) and a sharp Vimentin increase (~3,1-fold, slightly above the 5ng/mL TGF-β sample value), which can be taken as evidence that it is a mostly mesenchymal cell line. The E clone cell line however, despite registering a slight Vimentin decrease (~0,8-fold) shows almost no change in the E-cadherin level (it registered at 0,97-fold), which shouldn't

happen seeing as E clone cell line seems to be composed mostly of epithelial cells while nMuMG appears to have only around 40% of this type of cells.

A possible explanation for this is the fact that, as mentioned in section 3.1, all three cell lines show difference in terms of composition in epithelial and mesenchymal cells, but that may not be the only difference. An epithelial or mesenchymal cell in one cell line is not guaranteed to have the same average E-cadherin or Vimentin protein expression level as the same type of cell in another cell line (they may all have the same origin, an almost equal genome, but there are possible mutations or epigenetics to take into account), which can in the end lead to variation between the results obtained, and the published data. Also of note is the fact that M and E clone cell lines were created not bought, so there is no other cell line for a proper comparison to be made.

3.3.2 Immunocytochemistry high-throughtput analysis

As mentioned previously, the Western Blot experiment only provides results at a population level, which is why an immunocytochemistry experiment was done to observe the differences between the cell lines at the single cell level.

This experiment consisted of plating the cells in slides, allowing expansion and after fixing them, perform the staining for E-cadherin and Vimentin with the necessary antibodies. Samples were also stained for nucleus and actin, with the main objective of analyzing the images obtained with a custom designed algorithm.

Once stained the samples were imaged with a fluorescence microscope, a 4-by-4 fields of view grid that was then stitched together. However, said images make it hard to observe individual cells when displayed, that being the reason for Figure 3.11 containing just representative images of the big picture for each condition.

As mentioned, the resulting images were analyzed with a custom designed high-throughput cell phenotyping Matlab algorithm, in order to obtain the data on the expression level of E-cadherin and Vimentin for every visible cell in all the samples.

Once the software analyzed all the viable cells in the images for the expression levels of E-cadherin and Vimentin, scatter plots were created, to try and observe if there are any distinguishable trends in the samples. These plots are displayed as Figure 3.12.

These values were averaged out for each sample, and afterwards compared, this can be seen in Figure 3.13.

An average of the cell area per sample is also in display as it is known that the action of TGF-β when inducing EMT also leads to increase in cell size and protein content [83], as such there are concerns that the differences observed in the experiment are due to the different cell sizes.

When looking at the results presented in Figures 3.12 and 3.13 one can find both similarities and discrepancies with the Western Blot data. Regardless, the nMuMG cell line appears to have a generally lower intensity, both at the E-cadherin and Vimentin levels, as it is shown to be the lowest value in both plots. As expected, the nMuMG treated with TGF-β and M clone samples show an increase of

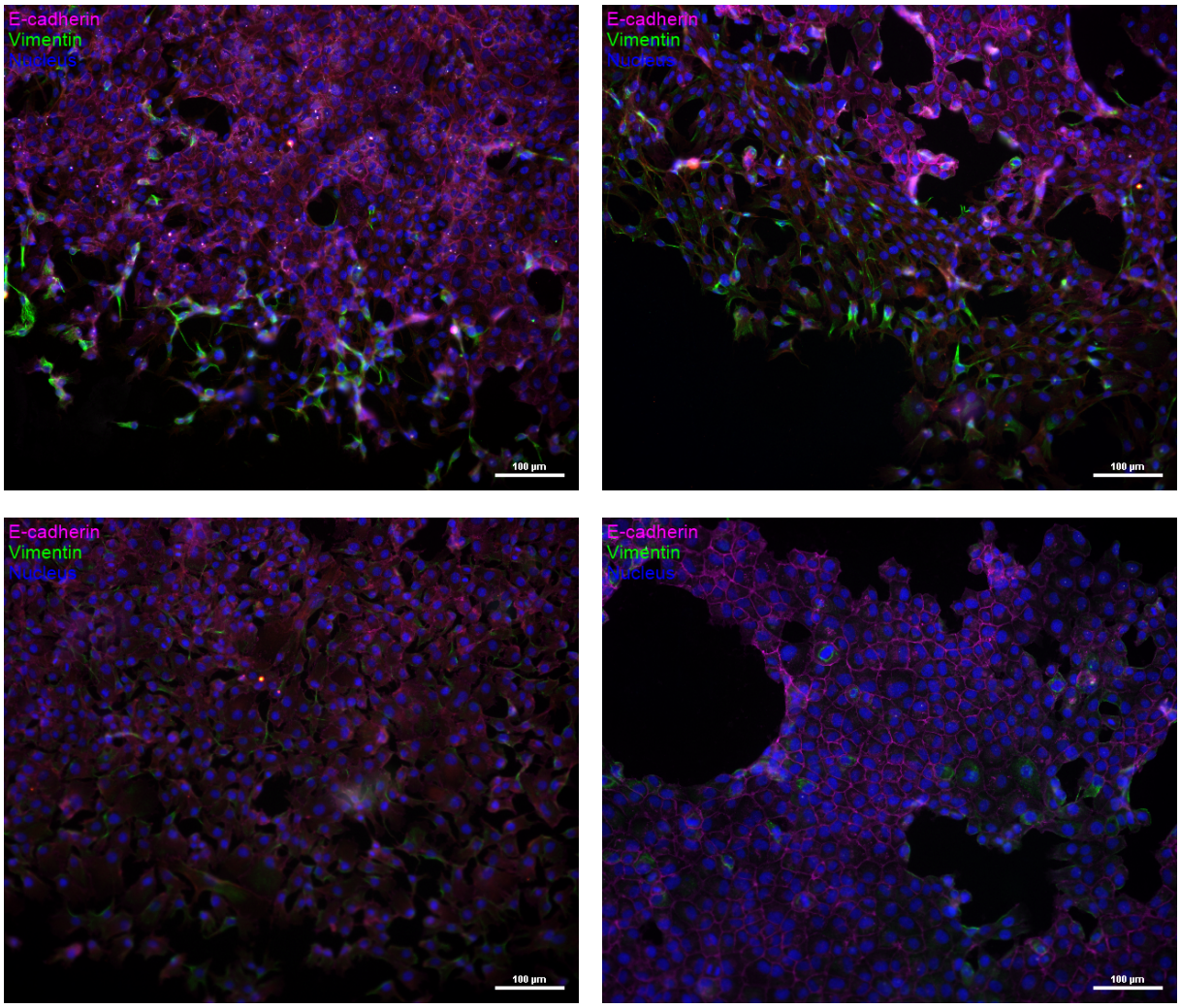

Figure 3.11: Representative images obtained by immunocytochemistry of the untreated nMuMG cell line (top left), 5 ng/mL 24h TGF-β exposed nMuMG cell line (top right), M clone cell line (bottom left) and E clone cell line (bottom right). Samples show staining for E-cadherin (magenta), Vimentin (green) and Nucleus (blue).

Vimentin expression as verified previously, while E clone sample demonstrated a marked increase in E-cadherin levels (according to the literature that is to be expected, but it was not observed previously). Against both what was written and the previous results nMuMG exposed to TGF-β and M clone samples presented increased E-cadherin compared to untreated parental cell line, while E clone sample also signaled increased Vimentin.

Analyzing the scatterplots provide the same information, only not as quantifiable. When comparing nMuMG samples, the exposure to TGF-β seems to create a marked increase in Vimentin as mentioned in the literature, while also inducing a smaller but still significant increase in E-cadherin. Comparing clone samples the shift is visible, as the M clone values tend towards higher levels of Vimentin, and the E clone remark increased E-cadherin (as seen in the average values, M clone has higher Vimentin expression, while E clone shows more E-cadherin intensity). Comparing the nMuMG line with either clone, it is easy to see that M clone seem shifted towards both Vimentin and E-cadherin increase, while E clone shows a similar trend, where the E-cadherin increase seems more obvious.

These similarities and discrepancies may have a variety of origins, from experimental error (the

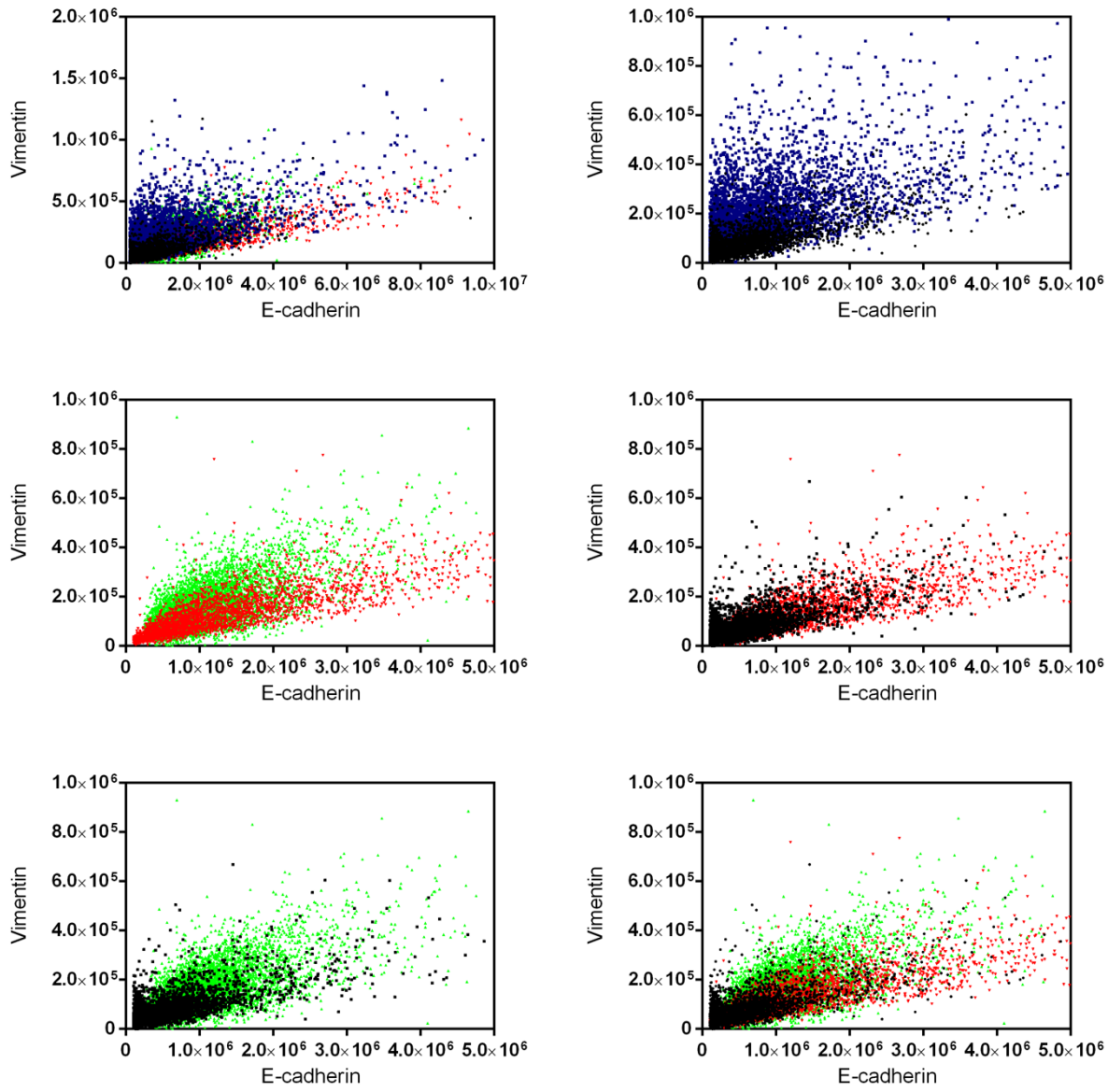

Figure 3.12: Scatterplots of the E-cadherin (X axis) and Vimentin (Y axis) data obtained by analysis of the immunocytochemistry images. Comparisons between all conditions (top left); untreated and TGF-β treated nMuMG (top right); M and E clone (middle left); nMuMG and E clone (middle right); nMuMG and M clone (bottom left) and parental, E and M clone (bottom right). nMuMG condition represented in black, TGF-β treated nMuMG condition in blue, M clone in green and E clone in red.

immunocytochemistry process is more finicky, requiring more care with the samples, therefore having higher probability of small errors impacting the results), to analysis error (the algorithm used identifies cells and measures intensity levels, but the identification process is not infallible, and although data was filtered, some false positives may have come through). As mentioned previously, the average area of the cells may have also effected the results, as it is expected that with larger cells, more protein is expressed, there is no way to confirm that the correlation between expression levels and cell size is the same in all the samples. Finally, the Western Blot data was obtained for a set of samples while the immunocytochemistry experiments were being tested throughout the project, so the samples are not exactly equals (probably not from the same thawed out batch, not even same age).

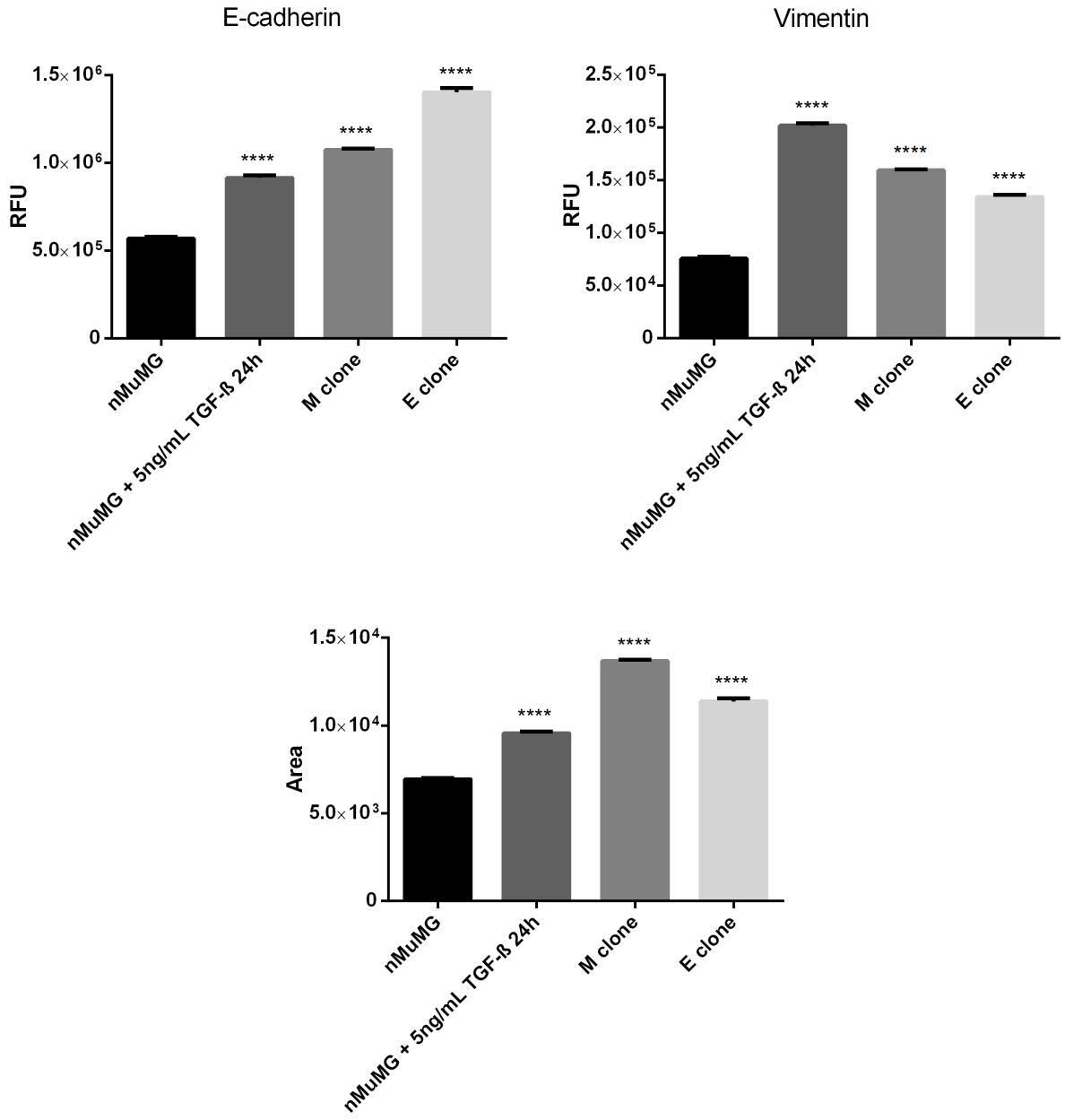

Figure 3.13: Average fluorescence intensity measures of E-cadherin (top left) and Vimentin (top right) expression and also cellular area (bottom) for nMuMG, TGFβ treated nMuMG, M and E clone samples.

3.4 Gene expression: transcription studies

3.4.1 RT-qPCR

As shown previously, differences had been found in the expression levels of E-cadherin and Vimentin (proteins chosen as markers of epithelial and mesenchymal phenotypes) in the three different cell lines analyzed. To better understand the expression of these protein markers, a study at the mRNA level was performed.

The idea was to check if the trends observed at the protein level are also verified when looking at mRNA (meaning the expression regulation of these genes occurs at the transcriptional level), because as it is beginning to be known, expression differences at the protein level are not always directly correlated with the mRNA expression [84].

To evaluate gene expression at the mRNA level there are several possible techniques: Northern Blot; RPA; DNA Microarray Analysis; RT-qPCR; SAGE; ADGE; among others [85]. The SAGE, ADGE and DNA Microarray techniques are more commonly used to compare transcriptomes, meaning that the objective is to study the whole of the mRNA expressed by a sample; as such not being as adequate for a study of a low number of genes in several samples. The RPA method is tailored for a more in-depth analysis of a single genes' mRNA and Northern Blot has a lower sensitivity limit, being less useful the less the gene is expressed, which led to the choice of the RT-qPCR method, a simple and cheap method. However, this process carries an inherent problem associated with the effects of the reaction efficiency on the final results.

Once the analysis method was defined, it was time to select which genes should be studied. E-cadherin and Vimentin were selected as these proteins had been selected as markers for epithelial and mesenchymal cells, respectively, in the Western Blot and immunocytochemistry experiments. As RT-qPCR studies rely on primers, which are much easier to obtain then antibodies, it was decided to study some more genes in this system. From there, the genes Snai1 and Snai2 (more commonly referred to as Snail and Slug) were selected, as the proteins they express are very important and common transcription factors involved the repression of E-cadherin expression when EMT is induced by exposure to TGF-β.

It was also necessary to define a reference gene, whose expression didn't differ too much in all the conditions to be studied, a baseline for the experiment. For this, GAPDH was chosen due to its designation as a housekeeping gene and it's prolific use as a real-time RT-qPCR reference gene in EMT studies [86, 87, 88, 89].

3.4.2 Primer Validation

A concern when performing RT-qPCR studies is that the technique hinges on the primers used, each pair is supposed to be specific for just the gene of interest and it should guarantee a good reaction efficiency.

As such, there is a need to confirm the quality of the chosen primers. This series of experiments are performed to prove that for each pair of primers only one region (amplicon) is amplified, and that the

annealing and amplification efficiency is high enough not to affect the results (one assumes that for each cycle, the DNA amount doubles, but the lower a primers' efficiency is, the less true that assumption is).

The genes studied and the reasons for their choice have been mentioned in subsection 3.4.1, while the details of the chosen pairs of primers can be found in section 2.8.

To study whether a pair of primers is specific for just one amplicon, several techniques can be used, most common amongst them the Melt Curve Analysis and DNA Gel Electrophoresis. In this case, both methodologies were performed, for added confirmation but also because Melt Curve Analysis is not always an accurate method.

Melt Curve Analysis

The Melt Curve Analysis technique requires no additional equipment, as it is performed in the thermocycler used for the RT-qPCR. After the cycle threshold measurements have been done, each sample contains a high amount of the target cDNA, which provides a high value of fluorescence due to the probe.

Melt Curve Analysis provides a plot of the fluorescence intensity against temperature. To do so, it starts at a lower temperature (around 40-50° C), and gradually increases that up to a maximum limit. While it is doing so, the fluorescence intensity is measured for every sample. A representative scheme is displayed in Figure 3.14.

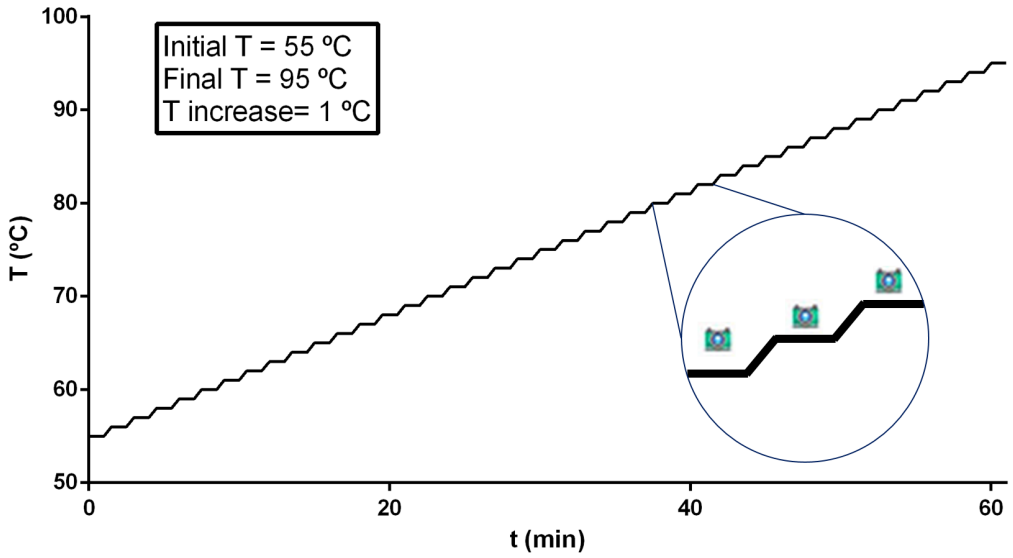

Figure 3.14: Scheme representative of the Melt Curve Analysis Procedure. As mentioned, the temperature is gradually increased from a low to a high value (so that the DNA can be captured as double stranded in the beginning and single stranded towards the end), and fluorescence is measured after every increment, as indicated by the camera icons.

Based on the DNA structure, a specific amplicon has a melting point (temperature at which 50% of the total DNA is double stranded and the other 50% is single stranded). This melting point is inserted in a small temperature range in which the total DNA switches from double to single strand, providing a sharp decrease in fluorescence intensity (the probe attaches only to double stranded DNA), which can

easily be seen in a plot of RFU vs T, and can also be identified as a single peak in a plot of -d(RFU)/dT vs T. Seeing as slight differences in the composition of a DNA strand can lead to melting point variations, the presence of a single peak in the derivative plot is a strong signal of the presence of just one amplicon in a sample.

The plots of RFU vs T and d(RFU)/dT vs T for all five pairs of primers are displayed in Figure 3.15.

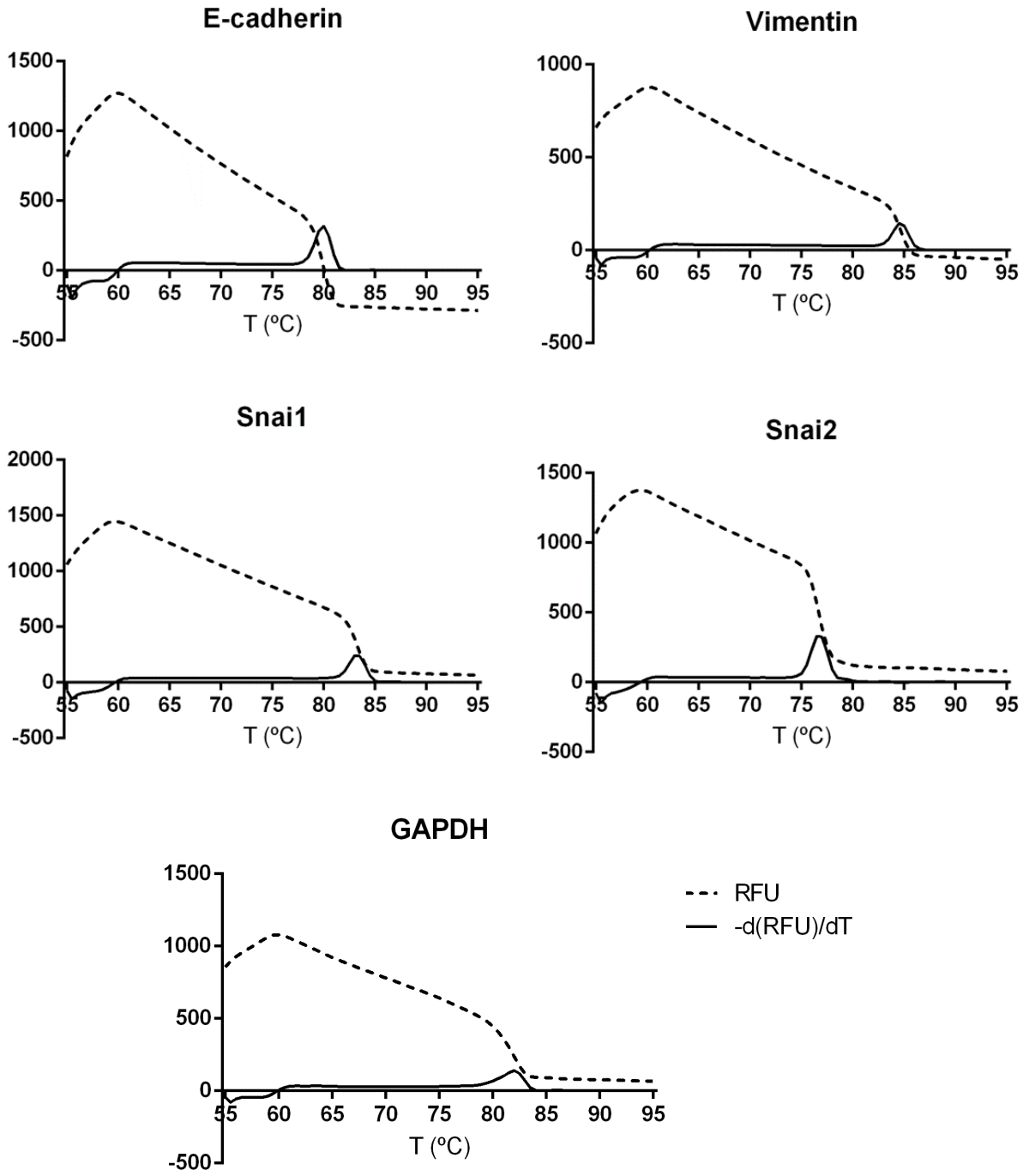

Figure 3.15: Plot of RFU (dotted line) and -d(RFU)/dT (full line) against T (in degrees Celsius) for the primer selected to the E-cadherin (top left), Vimentin (top right), Snai1 (middle left), Snai2 (middle right) and GAPDH (bottom) genes.

From what can be observed from these plots, the chosen primers seem to lead to the amplification

of a single region each, as expected, meaning they are so far, valid for the use in the experiments.

DNA Gel Electrophoresis

The Melt Curve Analysis method has been proven however, to have some limitations, the most glaring of which are: there is only guarantee of the presence of one amplicon in the samples, there is no extra information to prove that it is indeed the expected one and not the result of binding to a different gene; and that in some cases (amplicons with concentrated GC-rich regions) more than one peak may show up on the results (AT regions are easier to denaturate, lower melting point than GC regions) but it may still originate from just one amplicon.

As such, for added confirmation, DNA Gel Electrophoresis was conducted. In this procedure, the samples (after RT-qPCR) are loaded into a 1% agarose gel containing EtBr (ethidium bromide), and electrophoresis is run. DNA naturally has a negative charge, so it migrates from the negative electrode to the positive one when an electrical current is applied, and along the gel it will be separated by its molecular weight (size).

This allows to see if in any sample, there is more than one region of DNA being amplified (more than one DNA band would show up on the gel then), but also to check the molecular weight of a band and compare it with the amplicon that the primer should have amplified.

For that, after running the gel it is photographed under UV light, and the resulting imaged is displayed as Figure 3.16.

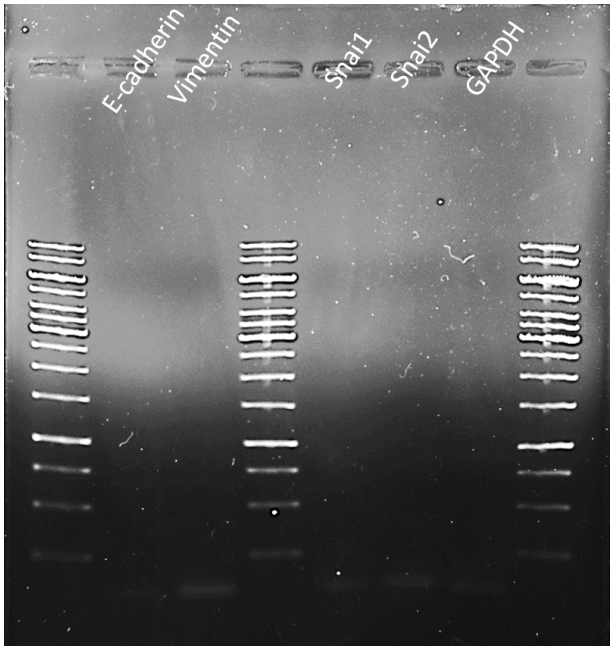

Figure 3.16: UV imaging of the agarose gel on which the DNA Electrophoresis was run. Hard to distinguish in some cases, but for the second, third, fifth, sixth and seventh lanes there seems to be only one DNA band. Above each of the mentioned lanes is the indication of which primer sample was on said lane

Once again, it can be observed that, for any of the 5 primer pairs that were chosen, there is only one amplicon, evidenced by the presence of only one DNA band in each well. Regrettably, the ladder avail-

able to keep track of a bands' molecular weight was not adequate enough for an accurate measurement, as all the primers selected were for amplicons with a size of around 100bp, and the lowest band on the ladder was 250bp, which would lead to an extrapolation with a large error margin.

Further testing to check the specificity of the primers against the theoretical data used to select them would require sequencing the amplicons in each sample, to compare with the sequence mentioned in the database.

3.4.3 Primer Efficiency

Calculating the efficiencies of the primers to be used in an assay is an important part in establishing a robust and precise qPCR experiment [90]. Ideally, a primer pair should only be used if the efficiency recorded is around 95-105%, but newer analysis methods have already started taking into account that it doesn't always happen and try to compensate. However, the most important part is that when analyzing more than 1 gene, the primers used should have as little a difference as possible between their efficiencies, this because at higher values, a variation of 1 unit in one genes' analysis may not actually correspond to the same variation in another [90].

As such, experiments were performed to obtain the efficiency of every pair of primers used. The experimental setup involved using one sample, at various dilutions (as can be seen in Figure 3.17), so that a measure of the cycle threshold was obtained and plotted against the logarithm of the relative DNA amount.

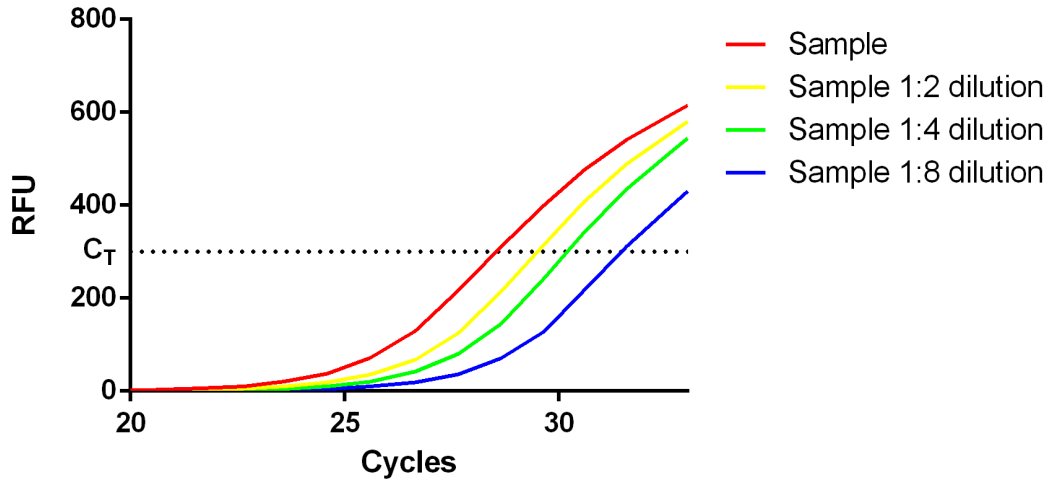

Figure 3.17: Example of the procedure performed to obtain the efficiency of a pair of primers. Knowing the cycle threshold values for various concentrations of the same sample, the curve can then be plotted and efficiency calculated as seen further on.

The sample chosen for this was the cDNA obtained from the untreated nMuMG cell line, and it was tested for each primer in the following conditions: undiluted, 1:2, 1:4 and 1:8 dilutions. For a 100% efficiency primer, the dilutions tested should have a cycle threshold one unit above the value of the immediately higher concentration sample. Triplicates were performed for measurement accuracy.

From there, and according to Equation 3.2, the slope of the plots obtained were used to calculate primer efficiency.

$$E\ (\%) = (10^{-\frac{1}{slope}} - 1) \times 100 \tag{3.2}$$

This is displayed in Figure 3.18.

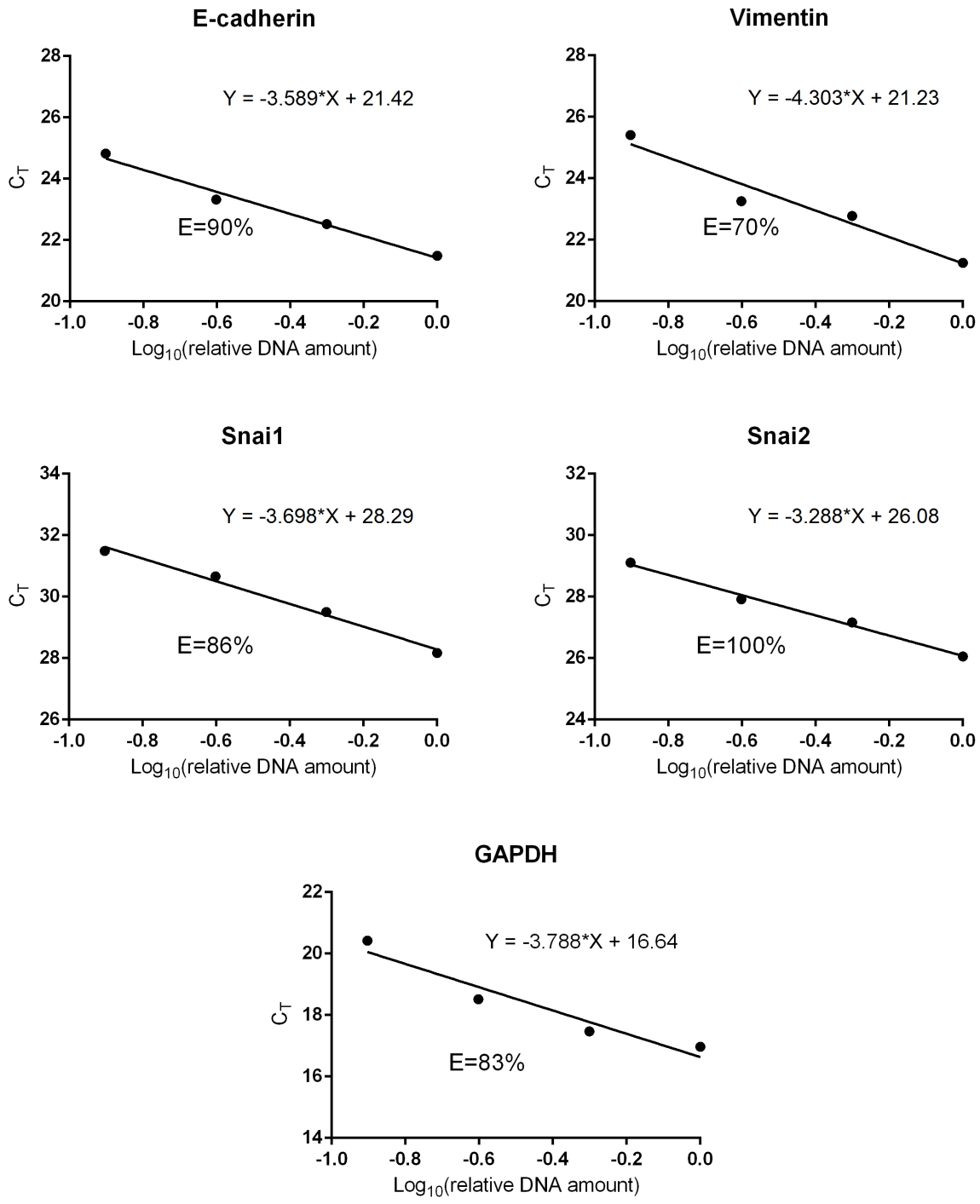

Figure 3.18: Plot of Cycle Threshold (C_T) versus decimal logarithm of relative DNA quantity for the E-cadherin (top left), Vimentin (top right), Snai1 (middle left), Snai2 (middle right) and GAPDH (bottom) gene primers. Curve was obtained by linear regression and efficiency calculated using Equation 3.2.

In terms of primer efficiencies, it is clear that the whole RT-qPCR experiment could have been better optimized. While the efficiency for the Snai2 primer pair is 100% and the E-cadherin pair registered a velue of 90%, the three other pairs (Vimentin, Snai1 and GAPDH) show efficiencies below 90%, and in the Vimentin primers case, a value of 70%. This is indication that the method used to analyze the data obtained from these experiments needs to take into account that not all the primers have 100% efficiency, but most important, as the efficiency of the reference gene is not similar to that of the others, the final results may present a larger error margin than is accounted.

3.4.4 Gene transcription results

The RT-qPCR experiments were done in a 96-well plate, analyzing each primer in every condition, with triplicates of each sample, to ensure precision. A duplicate plate was performed for every set of samples to ensure consistency in the results. To validate the quality of the results, two or three biological repeats were performed for every type of experiment performed.

The schematic for the 96-well plate is displayed as Figure 3.19.

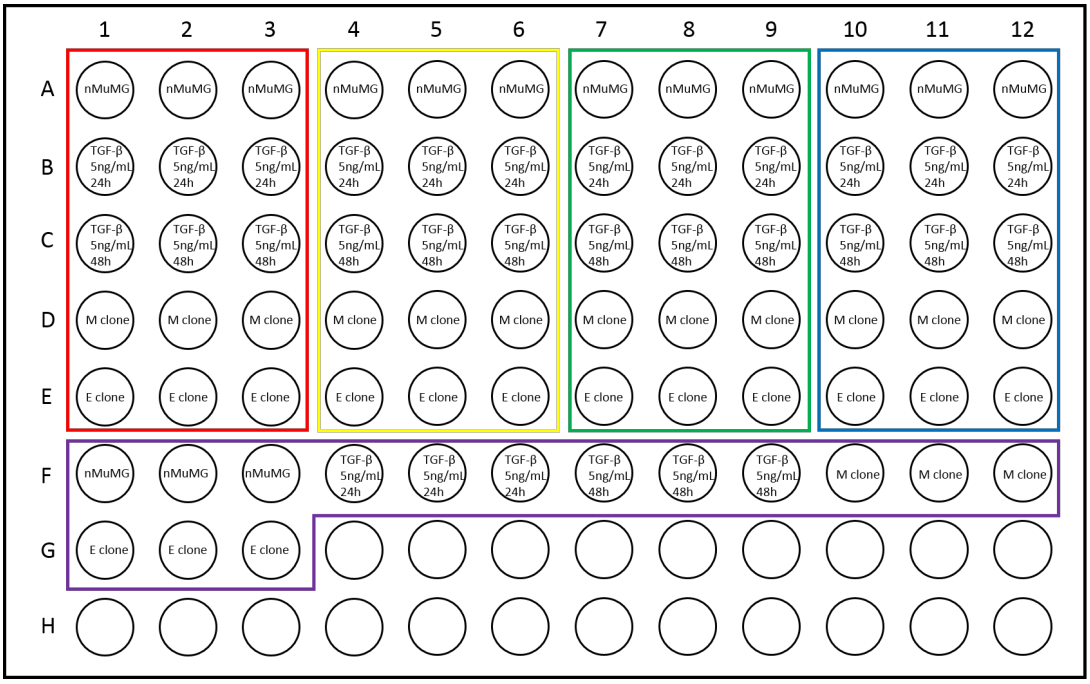

Figure 3.19: Schematic of the design of a 96-well plate, which was used to run the qPCR. The squares encompass the wells that were used to test for the gene expression of E-cadherin (red), Vimentin (yellow), Snai1 (green), Snai2 (blue) and GAPDH (purple).

From the three technical repeats that were performed for each sample and primer, the two closest cycle threshold values were selected for analysis.

To analyze qPCR data there are three methods: the Livak method, the ΔC_T method using a reference gene and the Pfaffl method, which was the one chosen for this analysis. The Livak method was discarded as it required primer efficiencies close to 100% and with a maximum difference between them of 5%; and the ΔC_T method was not used as it is a simple variation of the first one. The difference between the Livak and the Pfaffl methods is simply that the last one takes into account a primers actual

efficiency, therefore being usable for efficiencies below 100% or with more than 5% differences between primer pairs. A representative sample will be used as example of the workings of the method.

The data obtained, once screened for the more accurate values, is presented as shown in example Table 3.3.

Table 3.3: Cycle threshold values obtained from RT-qPCR for the samples tested and primers selected (Example)

	Gene								
	E-cadherin		**Vimentin**		**Snai1**		**Snai2**		**GADPH**
nMuMG	22,497	22,449	21,162	21,200	28,037	28,001	26,663	26,726	18,137
nMuMG + TGF-β 5 ng/mL 24 h	23,002	23,025	20,601	20,625	25,221	25,206	25,706	25,760	18,219
nMuMG + TGF-β 5 ng/mL 48 h	23,523	23,628	20,152	20,158	25,745	25,757	25,395	25,450	18,225
M clone	22,256	22,119	20,722	20,722	26,683	26,606	26,760	26,806	18,310
E clone	21,934	21,891	21,680	21,688	30,635	30,740	29,047	29,042	18,052

The first step involves normalizing the measures of every genes expression to the values of the reference gene, using Equation 3.3. This leads to example Table 3.4.

$$\Delta C_T(sample, gene) = C_T(sample, gene) - C_T(sample, reference\ gene) \tag{3.3}$$

Table 3.4: ΔC_T values obtained by applying Equation 3.3 to Table 3.3 (Example)

	Gene							
	E-cadherin		**Vimentin**		**Snai1**		**Snai2**	
nMuMG	4,360	4,311	3,025	3,063	9,900	9,864	8,526	8,589
nMuMG + TGF-β 5 ng/mL 24 h	4,783	4,807	2,382	2,406	7,003	6,988	7,487	7,541
nMuMG + TGF-β 5 ng/mL 48 h	5,298	5,403	1,927	1,933	7,520	7,532	7,170	7,225
M clone	3,946	3,809	2,412	2,412	8,373	8,296	8,450	8,496
E clone	3,883	3,840	3,628	3,636	12,583	12,688	10,996	10,990

Afterwards it is necessary to once again normalize the expression values, but this time against a control condition (as there is no data that could be used to obtain absolute gene expression values, the results need to all be relative to one of the samples). For this, Equation 3.4 is used and unmodified nMuMG sample is chosen as control. Note that for every PCR experiment there are three technical repeats for each sample primer, and from these three the two closer values are chosen. As such, in this step, the control sample value used to normalize each sample is actually the average of both duplicates.

$$\Delta\Delta C_T(sample, gene) = \Delta C_T(sample, gene) - \Delta C_{T_{average}}(control\ sample, gene) \tag{3.4}$$

From this step, the data obtained would be $\Delta\Delta C_t$, as exemplified in Table 3.5.

The final step turns the $\Delta\Delta C_t$ values into relative gene expression, using Equations 3.5a and 3.5b. This method, as mentioned before does not assume primer efficiency to be 100%, so it also requires the

Table 3.5: $\Delta\Delta C_T$ values obtained by applying Equation 3.4 to Table 3.4 (Example)

	Gene							
	E-cadherin		**Vimentin**		**Snai1**		**Snai2**	
nMuMG	0,024	-0,024	-0,019	0,019	0,018	-0,018	-0,031	0,031
nMuMG + TGF-β 5 ng/mL 24 h	0,448	0,471	-0,662	-0,638	-2,879	-2,894	-1,070	-1,016
nMuMG + TGF-β 5 ng/mL 48 h	0,963	1,068	-1,117	-1,111	-2,362	-2,350	-1,387	-1,332
M clone	-0,414	-0,502	-0,613	-0,651	-1,526	-1,568	-0,076	-0,093
E clone	-0,477	-0,472	0,603	0,573	2,683	2,824	2,469	2,402

slopes calculated in subsection 3.4.3. The result of this step is displayed in example Table 3.6.

$$a = 10^{-\frac{1}{primer\ slope}} \tag{3.5a}$$

$$RE\ (Relative\ Expression) = a^{\Delta\Delta C_T(sample,gene)} \tag{3.5b}$$

Table 3.6: Relative Gene Expression Values obtained by applying Equation 3.5b to Table 3.5 (Example)

	Gene							
	E-cadherin		**Vimentin**		**Snai1**		**Snai2**	
nMuMG	0,985	1,016	1,010	0,990	0,989	1,011	1,022	0,978
nMuMG + TGF-β 5 ng/mL 24 h	0,750	0,739	1,425	1,407	6,005	6,061	2,116	2,037
nMuMG + TGF-β 5 ng/mL 48 h	0,539	0,504	1,818	1,812	4,350	4,318	2,642	2,542
M clone	1,304	1,380	1,389	1,417	2,587	2,654	1,055	1,067
E clone	1,358	1,354	0,724	0,736	0,188	0,172	0,177	0,186

The accumulation of the technical duplicate gene expression values obtained, the duplicates results tested for each sample and the biological repeats performed, the gene expression results were all organized by sample and primer, and the final results are found in Figure 3.20.

Looking at the TGF-β treated samples, everything appears to be as would be expected, the E-cadherin levels decreased (~0,75-fold, not found significant), Vimentin expression increased to ~1,5-fold and both Snai1 and Snai2 also show increased expression (~6 and ~3-fold, expected as these two genes are known to encode for E-cadherin transcriptional inhibitors). Observing the M clone results show some peculiarities: while Vimentin expression increases as expected (~1,75-fold), despite the fact that there is increased levels of Snai1 (~4-fold) and Snai2 (~1,5-fold, not considered significant) mRNA, increased E-cadherin gene expression is also observed (~1,5-fold, though not considered significant). This is not in agreement with what a mesenchymal cell line should exhibit, not only the fact that it has higher levels of E-cadherin than the parental cell line (comparable to the levels of the E clone cell line), but also that this occurs at the same time that there is an increase in expression levels of two E-cadherin transcriptional repressors. This could be taken to indicate that either the Snai1 and Snai2 mRNA produced is not being converted to its active protein form (therefore being unable to act on the increased E-cadherin levels), or that in this cell line in study these two repressors are no longer part of the E-cadherin expression regulation process (and that whatever other transcriptional inhibitors are

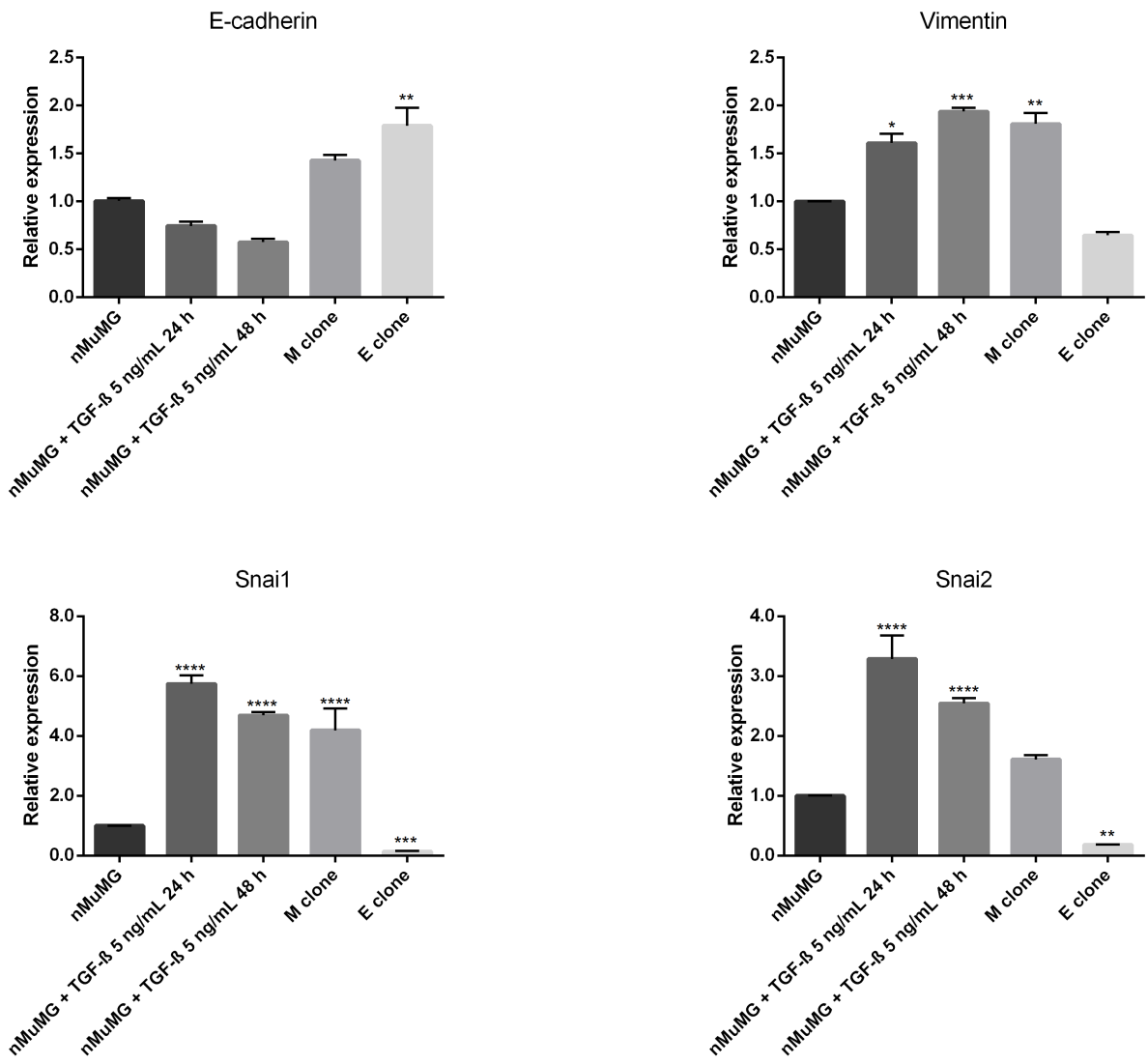

Figure 3.20: Comparison of the expression of the E-cadherin (top left), Vimentin (top right), Snai1 (bottom left), Snai2 (bottom right) genes in the samples indicated above relative to the untreated parental cell line (nMuMG).

similarly not exerting their function). Finally, the E clone cell line shows the results that were expected, an increase in E-cadherin to ~1,8-fold and decreased expression of Vimentin (~0,65, not significant), Snai1 and Snai2 (both ~0,1 to 0,3-fold, these last two accounting for the high levels of E-cadherin mRNA registered).

3.5 Cell density and it's effects

In cell culture, once a plate is reaching confluence, the cells are detached and then passed to a new dish, so as to maintain the culture, as some cells exhibit contact inhibition, and slow expansion at high densities, but also because of the fast nutrient consumption which would then lead to apoptosis. It has also been observed that many cell lines from different species begin behaving differently once the culture reaches high levels of confluency, from changes in gene and protein expression to differences in organization [91, 92, 93].

Some of the experiments performed used samples gathered from high density cell cultures, mostly out of an attempt to ensure that there is enough sample for the experiments to be performed. This, coupled with the knowledge presented in the previous paragraph led to concerns that the density of the culture could be affecting the results.

For this, the immunocytochemistry, and RT-qPCR experiments were repeated, but this time evaluating the effects of cell density.

3.5.1 Immunocytochemistry

For the immunocytochemistry, three 8-well chambered slides (one for each cell line) were seeded with three different cell concentrations (referred to as low, medium and high density) according to the design displayed in Figure 3.21.

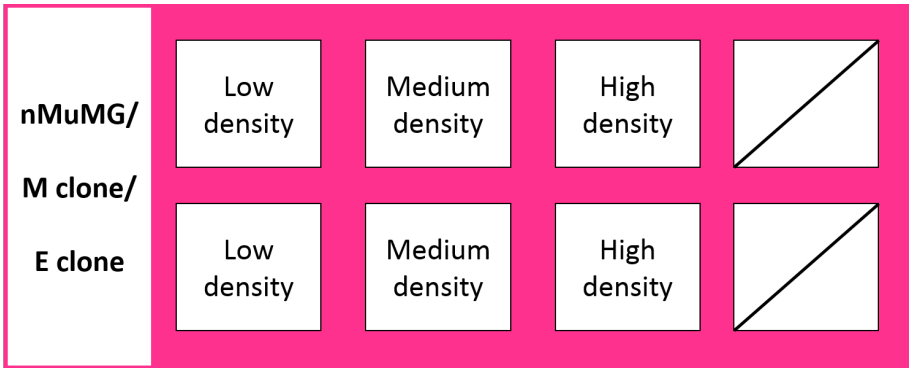

Figure 3.21: Schematic representing the 8-well slides in which these immunocytochemistry experiments were performed. Each slide contained duplicates for each density condition, for each cell line (3 cell lines studied means 3 separate slides)

After 2 days of expansion these samples were fixed and prepared for staining. Similar to previous observations, they were stained for nucleus, actin, E-cadherin and Vimentin.

As previously, a 4-by-4 image grid was acquired, and later on stitched. Representative images are displayed in Figure 3.22.

The results from the quantification analysis are similar but in a manner different, as it is a comparison of the effects of cell density in each one of the cell lines. As samples were analyzed in different slides a comparison between the three cell lines is risky, there is a lot of variation originating from the experimental conditions (since the samples weren't all in one slide, human error induces a higher grade

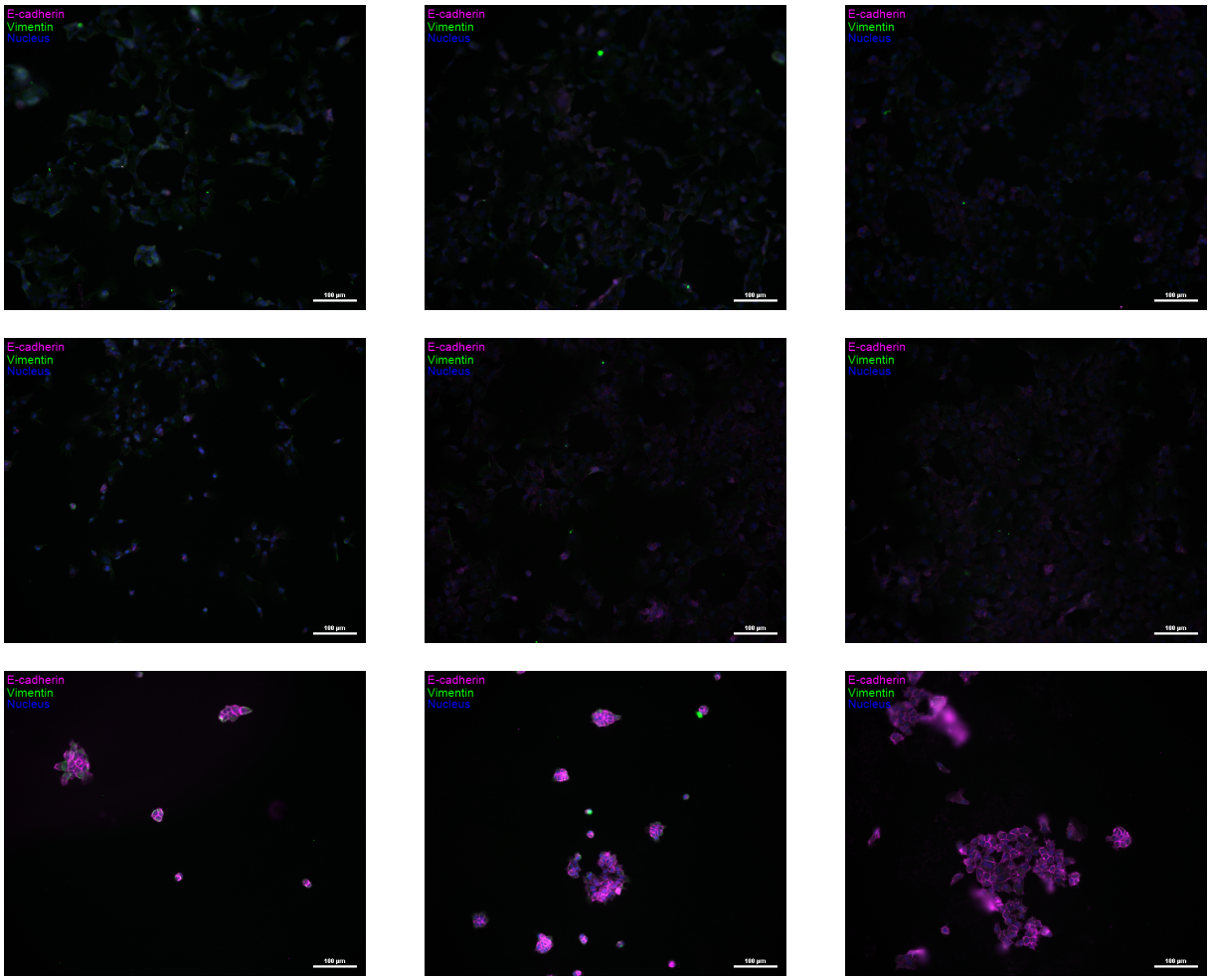

Figure 3.22: Representative images of all three cell lines, nMuMG (top), M clone (middle) and E clone (bottom); at different cell densities, low density (left), medium density (center) and high density (right). Images obtained through immunocytochemistry, by staining nucleus, E-cadherin and Vimentin

of variation). In order to be able to perform an accurate comparison between conditions at different densities, for example, one of the conditions would have to be repeated throughout all three slides so the others could be analyzed relative to that common one, similarly to what is done in mRNA analysis with a reference gene. This leads to Figure 3.23.

As mentioned previously, any comparison between conditions in this experiment is subject to large error, for it can be easily observed comparing the last line of images in Figure 3.22, the E clone cell line with the other two cell lines. Despite using the same intensity range in all images, the E clone ones show higher intensity, by a very wide margin, in all three channels; easily identifiable by the fact that it is hard to distinguish cells in either nMuMG or M clone samples, as opposed to E clone ones. However, it is possible to confirm that the density at which the cells are stained, most likely affects the results obtained. The average cell area was maintained at a constant level in the nMuMG samples, and throughout most of the E clone ones, while the M clone samples show different area averages for all three conditions (when comparing E clone high density with medium density, area is also significantly different). In terms of protein expression, both nMuMG E-cadherin and Vimentin showed a decrease in expression with the increase in density, while M clone presented a maximum of E-cadherin and a

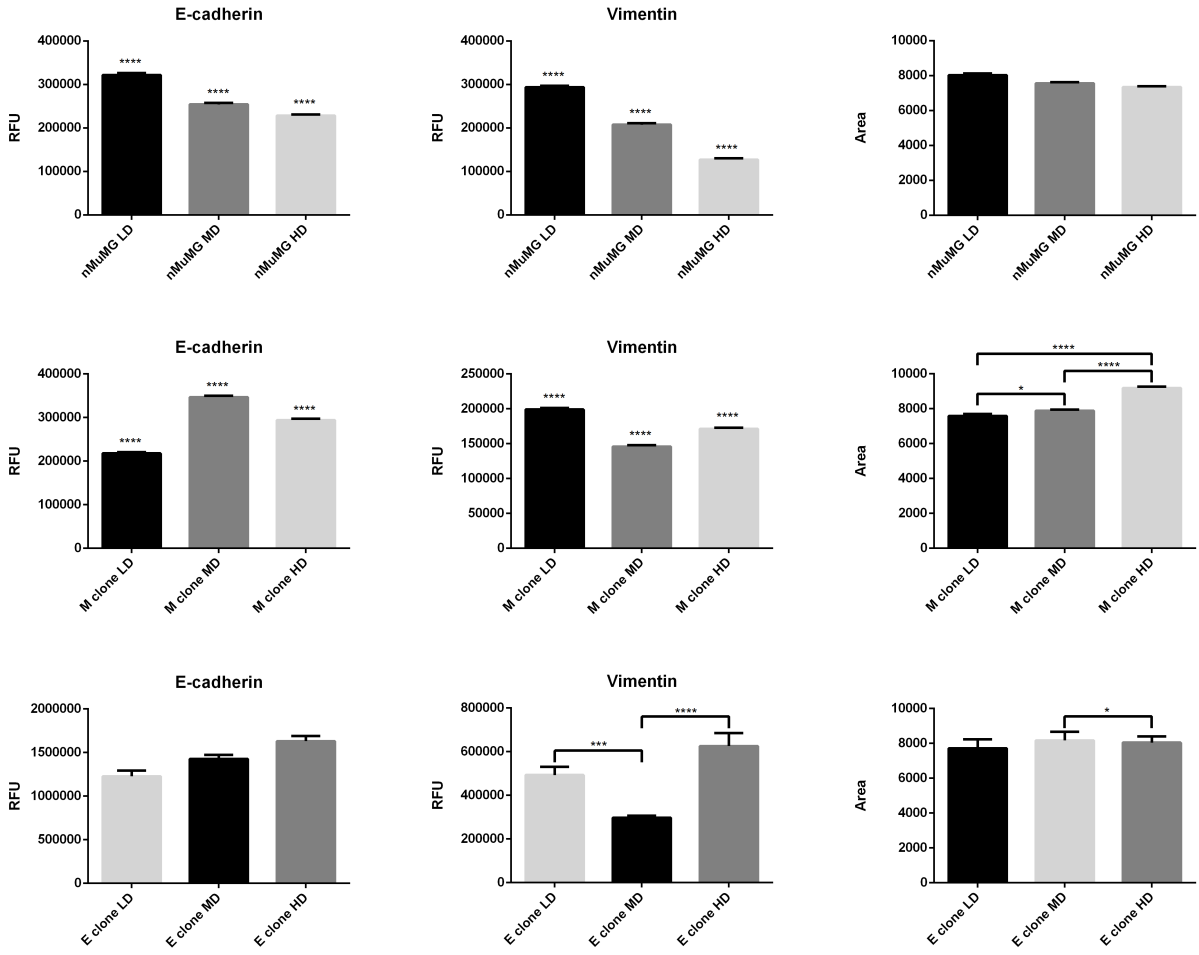

Figure 3.23: Averages of the Areas, E-cadherin and Vimentin Intensity in all three cell lines, nMuMG (top), M clone (middle) and E clone (bottom).

minimum of Vimentin expression at medium density. Looking at the E clone samples, the E-cadherin expression appears to steadily increase with density (although the values reported are not statistically significant), while Vimentin levels showed the same trend as in M clone samples, but with much more marked difference. This appears to tell us, once again, that despite both of the M and E clone cell lines having been originated from nMuMG cells, that their response to cell crowding due to density is somewhat different.

3.5.2 RT-qPCR

For the RT-qPCR, each cell line was seeded (in triplicate) at a low density, and allowed to grow for 1 day to around 30% confluence. From there, RNA was extracted.

The RNA samples were purified, converted to cDNA and then analyzed by qPCR, which led to the results found in Figure 3.24.

The results observed in Figure 3.24 agree almost fully with the mRNA analysis performed in subsection 3.4.4. M clone cell line shows increased E-cadherin and Vimentin expression levels along with a similar increase in Snai1 and Snai2; while the E clone cell line shows increase in E-cadherin levels

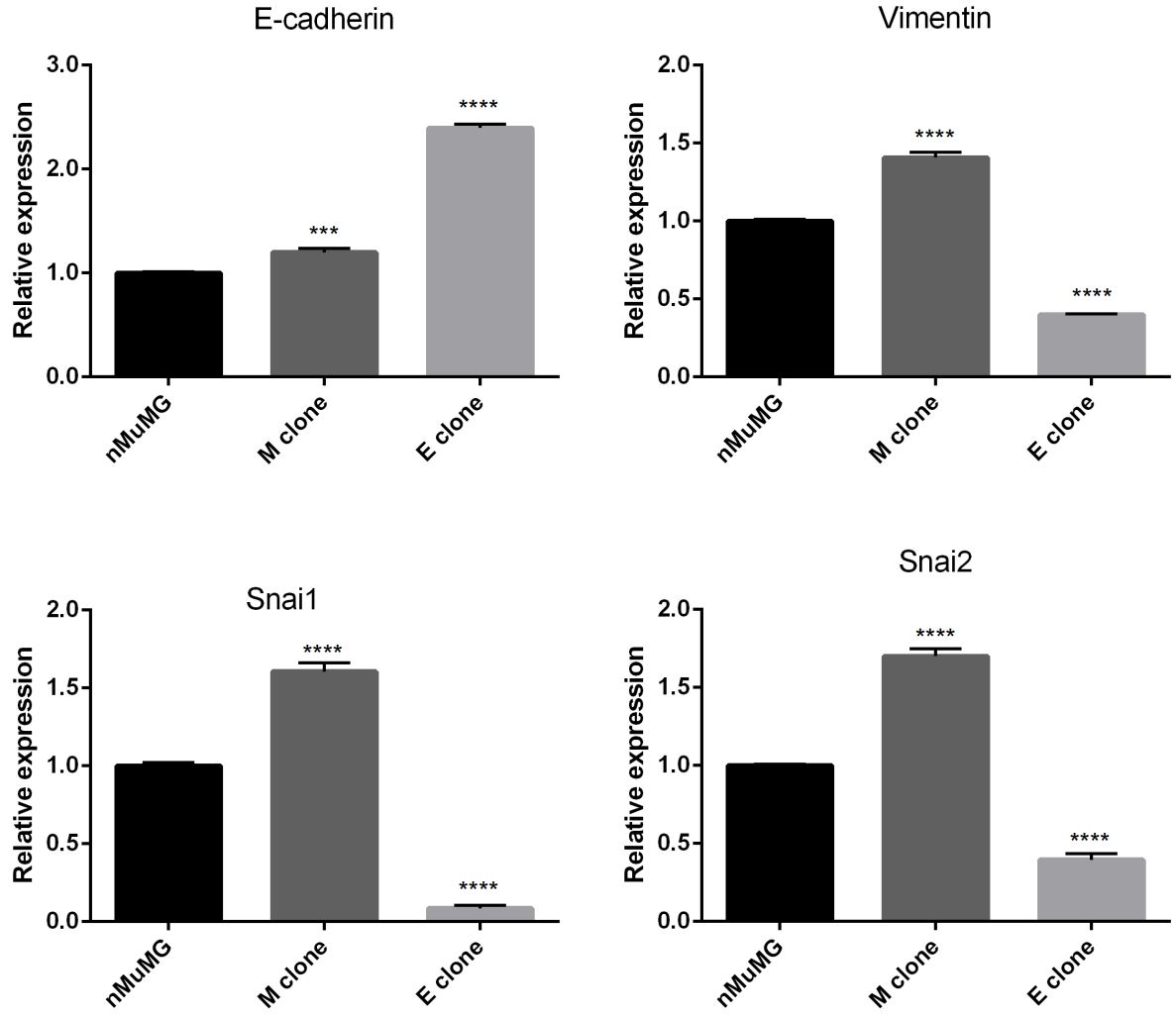

Figure 3.24: Comparison of the expression of the E-cadherin (top left), Vimentin (top right), Snai1 (bottom left), Snai2 (bottom right) genes at low cell density in the samples indicated above relative to the untreated parental cell line (nMuMG).

and decrease in the other 3 genes. The more notable difference here is that the M clone E-cadherin level is much closer to the parental cell line than the E clone (the opposite was observed in the previous analysis).

This time, the situation is the opposite to the one verified in the protein analysis, as the trends observed for each gene at both low and high densities can be compared but a comparison can't be made between the two cell densities in each sample condition, as that would need a parallel analysis of higher density samples. Case in point, despite showing the same trend, the M clone cell line values seem to be much lower, in almost every gene, than before, but without a parallel comparison with all the samples at different densities, there is no way of knowing what type of density effect in which cell line caused the apparent difference observed.

Chapter 4

Discussion and Future Work

4.1 Result Discussion

Along this work, it became clear, as can be seen in section 3.1, that despite being classified as an epithelial cell line, in culture, nMuMG presents a significant amount of what appear to be mesenchymal-like cells. Further on, in Figure 3.11 from section 3.3 it can be seen that these cells showing a mesenchymal morphology appear to show a difference in the expression of E-cadherin and Vimentin when compared to typical epithelial-like cells (the scatterplots in Figure 3.12 show in the nMuMG sample a great range of E-cadherin and Vimentin expression, from which can be assumed that high E-cadherin and low Vimentin represent the stereotypical epithelial cells, while the opposite expression levels refer to mesenchymal ones). This is then consistent with the presence of mesenchymal cells in an epithelial cell line.

From the live-cell images obtained, it was hard but in some cases possible to observe single cells switching from being present in epithelial colonies to surviving by themselves with evidence of mesenchymal morphology and *vice-versa*. However the data obtained from these experiments is, while not exactly insufficient, somewhat inconclusive, making it unlikely to be able to confidently affirm the existence of spontaneous EMT as a reversible process. Not only was it very hard to actually observe any change in the cells analyzed, but also the data obtained was purely morphological, which doesn't allow a reliable annotation of cells as either epithelial or mesenchymal, as such, a more accurate method would have to be used.

The two new engineered cell lines (E clone and M clone), and the colony analysis they were subject to lends some more credibility to the idea of EMT occurring in either direction without an induced external stimulus. The data showed that despite the fact that these two cell lines had been obtained by clonal expansion of a single epithelial and mesenchymal cell from the parental nMuMG cell line, respectively, both of the samples analyzed showed the presence of both types of cells. If EMT occurred only through outside intervention, such occurrences would not have been verified. Once again, this analysis was done from a purely morphological standpoint, making this another evidence of the possibility of spontaneous EMT, but not actionable proof (the samples would have to be further analyzed to confirm whether there was presence of both types of cells).

One of the staples of the hypothesis on the spontaneity and reversibility of EMT was that this process occurred in balance, meaning that some of the epithelial cells would be changing to mesenchymal and *vice-versa*, and that in the grand scheme of things, these transitions would all orbit a steady state, meaning a certain fixed percentage of epithelial and mesenchymal cells (noted to be around 40% epithelial cells and 60% mesenchymal ones in the nMuMG cell line). However, looking at the new cell lines that were studied, these percentages look significantly different from the values in the parental line. The question here then becomes: why? If cells spontaneously switch between both states, the issue must not be at the genetic level.

It was thought that a possibility for this may be epigenetics (it is known that this is the reason behind dynamic alterations in transcriptional potential). As such, there would be different ”groupings” of cells (some alterations may be hereditable) in a culture plate that all have, through external factors, different variations in their transcriptome, that may lead to changes in the epithelial mesenchymal balance, and so the final balance in a sample is an average of all those smaller groups. As such, when separating the cells for clonal expansion, it would only be natural that each new cell line would give rise to a sample with a different composition in epithelial and mesenchymal cells.

Still referring to the new cell lines, they were then verified for use as the epithelial and mesenchymal standards from which the parental cell line should be a mixture. The use of the nMuMG cell line treated with TGF-β was also useful to compare the long known and studied mesenchymal cells with the new standards. From looking at the protein levels in the Western Blot, subsection 3.3.1, it was noted that the M clone cell line assumed a behavior very similar to the nMuMG cells treated with TGF-β (higher Vimentin and lower E-cadherin when compared to parental sample), a good start in establishing at least a viable sample of the mesenchymal cells that compose part of the nMuMG cell line. However the E clone cell line, despite evidencing a decrease in Vimentin levels, displayed E-cadherin levels very similar to that of the parental cell line, which was not expected considering E clone cell line is composed mostly of epithelial-looking cells (which if actually epithelial should have increased E-cadherin expression), as opposed to the 40% epithelial cells in parental line. This would mean that either some of the epithelial-looking cells annotated previously in this cell line were not actually epithelial, or that it is a similar case than observed in the previous paragraph, where by unknown external factors, the cells regulate the E-cadherin gene expression to a lower level.

As such, the immunocytochemistry experiments were performed, to attempt to have a closer look at the cells that make up each cell line individually. Here the data obtained became somewhat conflicting. When looking at the images (as an example, Figure 3.11) it is possible to see the two different types of cells, the more elongated and green mesenchymal cells, and the colony growing rounder delineated in magenta epithelial ones. It is even possible, comparing the conditions studied, to observe some population shifts between the samples (as mentioned in section 3.1, the different cell lines seem to be composed of epithelial and mesenchymal cells in different percentages). When looking at the scatter plots, some of the expectable differences can be observed, M clone and TGF-β exposed nMuMG cells showed a shift towards increased Vimentin levels while E clone cells seem to have increased E-cadherin expression. However, observable in the comparison of average intensity levels is the conflicting data,

indicating that E clones also show increased Vimentin levels, and that the E-cadherin expression in M clones and TGF-β exposed nMuMG cells is higher than in the untreated nMuMG cell line. As already mentioned, it may have to do with experimental, analytical errors or even caused by possible differences in the samples tested in each experiment.

Looking at the cell lines at the protein level revealed some interesting data, which lead to the formulation of new questions. Then the decision was made to study the differences at a ”lower” gene expression level and once again compare the available samples. Here, there were some expected results, but it was not without its contradictions. The TGF-β treated nMuMG cells exhibited a very similar behavior at the E-cadherin and Vimentin mRNA level as they did at the protein level; and while that happened the mRNA results for the extra genes tested (Snai1 and Snai2, identified as transcriptional repressors of E-cadherin) are also in agreement with logic, as an increase in mRNA for these repressors was noted, fitting together with the complementary decrease in E-cadherin mRNA (and protein) levels. This time, it was the E clone cell line that showed the expected results, Vimentin levels were decreased, E-cadherin was increased, with also decreased Snai1 and Snai2 mRNA levels. When compared side by side with the Western Blot results, this then sheds some light on the possibilities. While the E-cadherin mRNA levels show an increase compared to parental, at the protein level there is no appreciable difference, which leads to the belief that another E-cadherin repressor is in action at this point, and this one either acts at a translational or post-translational level, somewhat decreasing the suspect list (unless it is another repressor not yet known, or simply not yet known to act in this manner). The M clone results show conflicting data by themselves. Looking at Vimentin levels, an increase is noted, much similar to the protein level increase. However, the questions arise when looking at the E-cadherin levels. Despite the E-cadherin protein levels being lower in this cell line than the parental one, an increase in E-cadherin mRNA was consistently verified when comparing the same cell lines. This comes together with the data from the Snai1 and Snai2 genes, that show an increase in mRNA levels, which would not agree with the before mentioned E-cadherin increase. Possibilities for this are that Snai1 and Snai2 are not fully translated and activated (so there is no action at the transcriptional level), and that there is the presence of another E-cadherin repressor at work (that would have to act post-transcriptionally to explain the disparities between mRNA and protein levels).

Interest came up in how the density of the samples used in the experiments performed affected the results, whether or not this parameter led to increases or decreases in expression of proteins and mRNA. For that just the three cell lines were studied.

At the protein levels it was observed that each cell line behaved somewhat differently. NMuMG steadily decreased both E-cadherin and Vimentin expression with cell density increase. M clones showed a maximum and a minimum, respectively, for the E-cadherin and Vimentin expression, at the medium cell density levels. While that, the E clone samples maintained E-cadherin levels about constant and showed a more abrupt minimum level of Vimentin expression at medium density. This shows that some of the previous results may have another possible source for the discrepancies mentioned, especially because, although no tests were performed, throughout the maintenance of these three cell lines, it was observed that E clone cell line showed the slowest growth, followed by nMuMG and finally

M clone. Ideally, a valid comparison between the three cell lines at the different densities would have been useful in order to understand if this was indeed a source causing the unexpected results observed earlier.

Looking at the mRNA, the trends observed seem to be the same at low density as verified at higher density. There isn't data that allows us to compare each cell line at both densities, so if there are comparable variations in mRNA expression from density effects, they are seemingly uniform in all of the three cell lines.

4.2 Final Conclusions

The data obtained along this project is a step towards proving the existence of a spontaneous and naturally occurring EMT process, trying to replace the concept of the induced EMT model that is the current mode of thought. Concrete proof was not supplied, as the experiments tried were very basic, and still showed some results that prove hard to conciliate (along with some that appear to be incorrect), but it is a first step, that builds on to the possibility of, with some more study, proving this theory correct.

The two clonal cell lines (already created at the beginning of the project) were tested along with the parental nMuMG line, and evidence seemed to point towards the two new cell lines being a viable standard for what could be considered the averages epithelial and mesenchymal cell (E and M clone, respectively).

The initial colony analysis appears to support the spontaneous EMT theory (how would mesenchymal cells show up in a cell line derived by clonal expansion from an epithelial cell otherwise?). The study of E-cadherin and Vimentin at the protein level brought conflicting results, as besides the discrepancies observed when comparing the Western Blot and immunocytochemistry experiments, the data analyzed showed various unexpected results (low E-cadherin in E clone sample on Western Blot, low E-cadherin and Vimentin levels in nMuMG sample from immunocytochemistry). While most certainly there is some error that accounts for the discrepancy, the issue remains whether it was an experimental error, an analytical one, or the other possibilities (that during the clonal expansion, some external factor affected the clone cell lines, leading to aberrant results, or even the unlikely fact that spontaneous EMT does not follow the conventionally established rules for EMT). The mRNA analysis once again also brings attention to those questions, as although the response of E-cadherin repressors Snai1 and Snai2 in M clone is similar to TGF-β treated samples, the E-cadherin levels in this cell line still register higher than the nMuMG cell line. Several possibilities arise here: seeing as the action of these repressors is at transcriptional level, the Snai1 and Snai2 mRNA may not be translated, or they may not even be involved in the spontaneous EMT process, but instead other molecules regulate E-cadherin levels post-transcriptionaly.

In the end, the mRNA analysis, despite not being completely consistent (the E-cadherin levels for both clone cell lines had a different trend in each analysis) with the Western Blot data, was still closer to this data then the immunocytochemistry. Thought was given to this, and the more likely possibilities are that of human error, or that there was an issue with the data gathering in the immunocytochemistry

images, as the algorithm used also depends on human factors (choice of several parameters) to operate, and the cell tracing may not always be accurate.

Also noted were some effects of cell crowding (due to the high cell densities in some of the experiments performed), and here it gave support to the idea that it is not exactly the phenotype of the cell that gives rise to the clonal cell line that affects its behavior but instead the epigenetic factors, as all three cell lines exhibited different behaviors at the protein level. At the mRNA level the trends observed were seemingly the same, but further testing would be needed to understand if the gene transcription is actually not affected by cell crowding or if it is a combination of the different variations in all three cell lines that makes the mRNA trend at low density appear so similar to the one at high density.

4.3 Future Work

Possibilities for future work spanning from this project are plenty, especially since EMT is regarded as an important process in cancer metastasis.

Continuing directly from the conclusion of this project, an interesting idea to consider would be to expand the low density colony analysis. The proposed expansion would be both in breadth of samples and of tests. Additional samples that could be tested in this way would be the exact same three cell lines, but this time exposed to agents known to alter the balance in the ratio of epithelial and mesenchymal cells, for example the EMT inducer TGF-β or the microRNA 200 family, known to prevent EMT. Not only would it be useful to observe the final result in terms of epithelial and mesenchymal cells in the nMuMG line, but it would also allow to check the susceptibility of the new cell lines to factors with a known effect in the parental line.

Another interesting path to take from the conclusion of this project would be to study the effects of different environmental factors: mechanical, biochemical, *etc.* on the distribution of all three cell lines in percentages of epithelial and mesenchymal cells. This would provide knowledge on the possible external factors that favor the EMT or the MET process, to understand how the cell environment contributes to EMT, and possibly metastasis.

As mentioned in section 4.1, the live-cell imaging data obtained and analyzed didn't supply much in the way of conclusive data on the spontaneous occurrence of EMT throughout time. The line of thought is that morphological data is somewhat unreliable in terms of accurately annotating cells as epithelial or mesenchymal, which brings up the need to design a new experiment that simultaneously allows for long term observation of live cells, but also that provides more accurate information than morphology to be able to identify cells as in one of the two states in the transition.

An idea for that would be the creation of a reporter construct.

A plasmid containing a native promoter (the promoter from a gene that is naturally expressed in the cells) would be attached to a fluorescent protein, like GFP or RFP; and the plasmid would then be inserted in a cell line. Through whatever insertion and expression protocol (viral or non-viral; transient or stable), the target cells will then start producing the reporter fluorescent protein, with the expression levels being regulated by the factors that affect the expression of the reporter promoter.

To be able to differentiate between the two states in the transition process, a useful gene promoter would be one that showed different expression levels for the two phenotypes of interest.

It is known that the promoter region of a gene, despite being a non-coding area, is fundamental in deciding how much is a gene transcribed (how many mRNA copies of said gene will be produced). In a simplified manner, it can be inferred that high mRNA levels for a certain gene in one phenotype/condition and low levels for the other indicate a strong promoter candidate.

This would allow us to observe the cells, track them and further on more easily identify their phenotype as a function of time.

From the mRNA analysis results, the best possible candidate for the construct would have to be the Snai1 gene. Assuming that the cell lines E clone and M clone are faithful representations of a nMuMG cell in the epithelial and mesenchymal stages, respectively, the Snai1 gene is the one that at an mRNA level (directly correlated to expression of the gene's promoter) shows the clearest separation between the two (~3 fold expression in the mesenchymal state as opposed to ~0,15 fold in the epithelial state). This would then guarantee that the reporter protein, which would be attached to this gene's promoter, would have very different expression levels in each of the two states of the transition, therefore allowing for a better distinction in tracking and annotating cells in time-lapse.

Something that may be interesting also for further study came from comparing the M clone cell line results with the positive control (nMuMG cells treated with TGF-β). It was shown that despite the similarities found at the protein level between the two samples, that there were differences at the mRNA levels, meaning that even if the final result is similar (and analyzing more parameters may show that it might not exactly be that similar), the means to get there don't appear to be the same. This would mean that a further study into the mechanism of the spontaneous EMT would be necessary, as the assumptions made from studying the EMT induced by external factors may not accurately describe what occurs naturally in a living organism.

Another point of interest to better complete the work done would be to retry the clonal expansion, this time obtaining more than one sample for what is to be called the E and M clone cell lines. It was observed that the three cell lines analyzed showed differences, starting from their composition in percentage of epithelial and mesenchymal cells, up to differences in mRNA and protein levels. However, both the E and M clone cell lines were originated from the parental nMuMG line. This led to the thought that it is possible that despite the cells in all these cell lines being genetically similar, that external factors affect cells differently (epigenetics). The idea is to obtain several more examples of E and M clone cell lines and compare them between their similars, to try and check if there are also significant changes, if there are, this means that each cell line presents differences because the parental cells, despite being very similar genetically, each had their own expression/transition levels/rates due to variations in the effects of external factors.

An interesting proposition, especially considering the possible cancer implications regarding this transition process, would be to try and eventually turn these 2D studies into 3D. The final goal of biological studies is to understand how things work in the human body, so it is natural that the models used in these studies should try to best simulate the environment of interest. Additionally, many studies are proving

that plenty of 2D results are not consistent with the same tests done in 3D. There is always of course the necessity of trying to understand what is being studied in a simpler platform before trying to increase the complexity (not to mention, 3D analysis is much harder and more limited than what is possible in 2D), but once the topic becomes better understood this would be one of the logical steps, especially if it is proven that this process is vital for cancer metastasis (which is the main cause of cancer-related deaths). In this are there are already some attempts being performed, like [9], where an *in vitro* model has been design to recapitulate some features of EMT, and has shown some similarities with clinical breast cancer cell samples.

Another logical step, that could probably be introduced before the switch to 3D studies is the addition of new cell lines, the studies performed here were all in nMuMG, a mouse cell line, however, to better understand this process in humans, naturally that one or more human epithelial cell lines need to be observed. Using more than one cell line might even lead to discover that the EMT process might occur differently depending on the type of cell (non-cancerous, cancerous non-metastatic, cancerous metastatic, *etc.*).

Bibliography

[1] Paolo Mazzarello. A unifying concept: the history of cell theory. *Nature Cell Biology*, 1(1):E13–E15, 1999.

[2] Reviews Microscope Master: Research and Comparisons. Cell theory and microscopes: An introduction to microscopy. `http://www.microscopemaster.com/cell-theory.html`, June 2011. (Visited on 10/12/2015).

[3] Raghu Kalluri and Robert A Weinberg. The basics of epithelial-mesenchymal transition. *The Journal of clinical investigation*, 119(6):1420, 2009.

[4] U.S. Department of Health Bethesda, MD: National Institutes of Health and 2015 Human Services. Stem cell basics: Introduction. in *Stem Cell Information*. Available at `http://stemcells.nih.gov/info/basics/pages/basics1.aspx`, March 2015. [cited Monday, October 12, 2015].

[5] Jean Paul Thiery. Epithelial–mesenchymal transitions in tumour progression. *Nature Reviews Cancer*, 2(6):442–454, 2002.

[6] Davidson College Biology Department. Epithelial cells introduction. `http://www.bio.davidson.edu/people/kabernd/BerndCV/Lab/EpithelialInfoWeb/`, 2010. (Visited on 02/27/2015).

[7] Buzz Baum, Jeffrey Settleman, and Margaret P Quinlan. Transitions between epithelial and mesenchymal states in development and disease. In *Seminars in cell & developmental biology*, volume 19, pages 294–308. Elsevier, 2008.

[8] Jeff H Tsai, Joana Liu Donaher, Danielle A Murphy, Sandra Chau, and Jing Yang. Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer cell*, 22(6):725–736, 2012.

[9] Elad Katz, Sylvie Dubois-Marshall, Andrew H Sims, Philippe Gautier, Helen Caldwell, Richard R Meehan, and David J Harrison. An in vitro model that recapitulates the epithelial to mesenchymal transition (emt) in human breast cancer. *PLoS One*, 6(2):e17083, 2011.

[10] Lionel Larue and Alfonso Bellacosa. Epithelial–mesenchymal transition in development and cancer: role of phosphatidylinositol 3' kinase/akt pathways. *Oncogene*, 24(50):7443–7454, 2005.

[11] Michael W Klymkowsky and Pierre Savagner. Epithelial-mesenchymal transition: a cancer researcher's conceptual friend and foe. *The American journal of pathology*, 174(5):1588–1593, 2009.

[12] Raghu Kalluri and Eric G Neilson. Epithelial-mesenchymal transition and its implications for fibrosis. *Journal of Clinical Investigation*, 112(12):1776, 2003.

[13] Anna Ferrer-Vaquer, Manuel Viotti, and Anna-Katerina Hadjantonakis. Transitions between epithelial and mesenchymal states and the morphogenesis of the early mouse embryo. *Cell adhesion & migration*, 4(3):447–457, 2010.

[14] Chris Jopling, Stephanie Boue, and Juan Carlos Izpisua Belmonte. Dedifferentiation, transdifferentiation and reprogramming: three routes to regeneration. *Nature reviews Molecular cell biology*, 12(2):79–89, 2011.

[15] Kazutoshi Takahashi and Shinya Yamanaka. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *cell*, 126(4):663–676, 2006.

[16] Frank Rattray Lillie. *The development of the chick: an introduction to embryology*. Holt, 1908.

[17] Raul Fleischmajer. Epithelial-mesenchymal interactions. *Science*, 157(3795):1472–1491, 1967.

[18] Robert L Trelstad, Elizabeth D Hay, and Jean-Paul Revel. Cell contact during early morphogenesis in the chick embryo. *Developmental biology*, 16(1):78–106, 1967.

[19] Gary Greenburg and Elizabeth D Hay. Epithelia suspended in collagen gels can lose polarity and express characteristics of migrating mesenchymal cells. *The Journal of cell biology*, 95(1):333–339, 1982.

[20] MICHAEL Stoker and MARION Perryman. An epithelial scatter factor released by embryo fibroblasts. *Journal of cell science*, 77(1):209–223, 1985.

[21] Michael Stoker, Ermanno Gherardi, Marion Perryman, and Julia Gray. Scatter factor is a fibroblast-derived modulator of epithelial cell mobility. *Nature*, 327(6119):239–242, 1987.

[22] Ermanno Gherardi, Julia Gray, Michael Stoker, Marion Perryman, and Robert Furlong. Purification of scatter factor, a fibroblast-derived basic protein that modulates epithelial interactions and movement. *Proceedings of the National Academy of Sciences*, 86(15):5844–5848, 1989.

[23] Samy Lamouille, Jian Xu, and Rik Derynck. Molecular mechanisms of epithelial–mesenchymal transition. *Nature reviews Molecular cell biology*, 15(3):178–196, 2014.

[24] M Angela Nieto. Epithelial plasticity: a common theme in embryonic and cancer cells. *Science*, 342(6159):1234850, 2013.

[25] David M Gonzalez and Damian Medici. Signaling mechanisms of the epithelial-mesenchymal transition. *Science signaling*, 7(344):re8–re8, 2014.

[26] Jean Paul Thiery, Hervé Acloque, Ruby YJ Huang, and M Angela Nieto. Epithelial-mesenchymal transitions in development and disease. *Cell*, 139(5):871–890, 2009.

[27] Dianbo Yao, Chaoliu Dai, and Songlin Peng. Mechanism of the mesenchymal–epithelial transition and its relationship with metastatic tumor formation. *Molecular Cancer Research*, 9(12):1608–1620, 2011.

[28] NPA Devika Gunasinghe, Alan Wells, Erik W Thompson, and Honor J Hugo. Mesenchymal–epithelial transition (met) as a mechanism for metastatic colonisation in breast cancer. *Cancer and Metastasis Reviews*, 31(3-4):469–478, 2012.

[29] Christine L Chaffer, Janelle P Brennan, John L Slavin, Tony Blick, Erik W Thompson, and Elizabeth D Williams. Mesenchymal-to-epithelial transition facilitates bladder cancer metastasis: role of fibroblast growth factor receptor-2. *Cancer research*, 66(23):11271–11278, 2006.

[30] Wenjing Yu, Zhen Liu, Shu An, Jinyi Zhao, Lan Xiao, Yongchao Gou, Yunfeng Lin, and Jun Wang. The endothelial-mesenchymal transition (endmt) and tissue regeneration. *Current stem cell research & therapy*, 9(3):196–204, 2014.

[31] Jeff H Tsai and Jing Yang. Epithelial–mesenchymal plasticity in carcinoma metastasis. *Genes & development*, 27(20):2192–2206, 2013.

[32] David Shook and Ray Keller. Mechanisms, mechanics and function of epithelial–mesenchymal transitions in early development. *Mechanisms of development*, 120(11):1351–1383, 2003.

[33] Jonathan M Lee, Shoukat Dedhar, Raghu Kalluri, and Erik W Thompson. The epithelial–mesenchymal transition: new insights in signaling, development, and disease. *The Journal of cell biology*, 172(7):973–981, 2006.

[34] Konrad Steinestel, Stefan Eder, Andres Jan Schrader, and Julie Steinestel. Clinical significance of epithelial-mesenchymal transition. *Clinical and translational medicine*, 3:17, 2014.

[35] Min Yu, Aditya Bardia, Ben S Wittner, Shannon L Stott, Malgorzata E Smas, David T Ting, Steven J Isakoff, Jordan C Ciciliano, Marissa N Wells, Ajay M Shah, et al. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *science*, 339(6119):580–584, 2013.

[36] Shinji Hirano, Naomi Kimoto, Yutaka Shimoyama, Setsuo Hirohashi, and Masatoshi Takeichi. Identification of a neural α-catenin as a key regulator of cadherin function and multicellular organization. *Cell*, 70(2):293–301, 1992.

[37] Christopher S Williams, Baolin Zhang, J Joshua Smith, Ashwath Jayagopal, Caitlyn W Barrett, Christopher Pino, Patricia Russ, Sai H Presley, DunFa Peng, Daniel O Rosenblatt, et al. Bves regulates emt in human corneal and colon cancer cells and is silenced via promoter methylation in human colorectal carcinoma. *The Journal of clinical investigation*, 121(10):4056, 2011.

[38] Jose A Galván, Aurora Astudillo, Aitana Vallina, Paula J Fonseca, Lourdes Gómez-Izquierdo, Rocío García-Carbonero, and Maria Victoria González. Epithelial-mesenchymal transition markers in the

differential diagnosis of gastroenteropancreatic neuroendocrine tumors. *American journal of clinical pathology*, 140(1):61–72, 2013.

[39] Mercedes Lioni, Patricia Brafford, Claudia Andl, Anil Rustgi, Wafik El-Deiry, Meenhard Herlyn, and Keiran SM Smalley. Dysregulation of claudin-7 leads to loss of e-cadherin expression and the increased invasion of esophageal squamous cell carcinoma cells. *The American journal of pathology*, 170(2):709–721, 2007.

[40] Scott L Kominsky, Pedram Argani, Dorian Korz, Ella Evron, Venu Raman, Elizabeth Garrett, Alan Rein, Guido Sauter, Olli-P Kallioniemi, and Saraswati Sukumar. Loss of the tight junction protein claudin-7 correlates with histological grade in both ductal carcinoma in situ and invasive ductal carcinoma of the breast. *Oncogene*, 22(13):2021–2033, 2003.

[41] Kai J Lorenz, Klaus Kraft, Franziska Graf, Christian Pröpper, and Konrad Steinestel. Role of reflux-induced epithelial–mesenchymal transition in periprosthetic leakage after prosthetic voice rehabilitation. *Head & neck*, 37(4):530–536, 2015.

[42] Yongli Shi, Hongyan Wu, Mingyi Zhang, Lei Ding, Fanqing Meng, and Xiangshan Fan. Expression of the epithelial-mesenchymal transition-related proteins and their clinical significance in lung adenocarcinoma. *Diagn Pathol*, 8:89, 2013.

[43] Amir F Kagalwalla, Noorain Akhtar, Samantha A Woodruff, Bryan A Rea, Joanne C Masterson, Vincent Mukkada, Kalyan R Parashette, Jian Du, Sophie Fillon, Cheryl A Protheroe, et al. Eosinophilic esophagitis: epithelial mesenchymal transition contributes to esophageal remodeling and reverses with treatment. *Journal of Allergy and Clinical Immunology*, 129(5):1387–1396, 2012.

[44] Frans Van Roy. Beyond e-cadherin: roles of other cadherin superfamily members in cancer. *Nature Reviews Cancer*, 14(2):121–134, 2014.

[45] M Pichler, AL Ress, E Winter, V Stiegelbauer, M Karbiener, D Schwarzenbacher, M Scheideler, C Ivan, SW Jahn, T Kiesslich, et al. Mir-200a regulates epithelial to mesenchymal transition-related gene expression and determines prognosis in colorectal cancer patients. *British journal of cancer*, 110(6):1614–1621, 2014.

[46] Haoxuan Zheng, Wenjing Li, Yadong Wang, Tingting Xie, Yidong Cai, Zhiqing Wang, and Bo Jiang. mir-23a inhibits e-cadherin expression and is regulated by ap-1 and nfat4 complex during fas-induced emt in gastrointestinal cancer. *Carcinogenesis*, 35(1):173–183, 2014.

[47] Preeti Shah, Yael Gau, and Gauri Sabnis. Histone deacetylase inhibitor entinostat reverses epithelial to mesenchymal transition of breast cancer cells by reversing the repression of e-cadherin. *Breast cancer research and treatment*, 143(1):99–111, 2014.

[48] Lai Jin, Jiandong Chen, Li Li, Chuanhua Li, Cheng Chen, and Shengnan Li. Crh suppressed tgfβ1-induced epithelial–mesenchymal transition via induction of e-cadherin in breast cancer cells. *Cellular signalling*, 26(4):757–765, 2014.

[49] R YJ Huang, MK Wong, TZ Tan, KT Kuay, AHC Ng, VY Chung, YS Chu, N Matsumura, HC Lai, YF Lee, et al. An emt spectrum defines an anoikis-resistant and spheroidogenic intermediate mesenchymal state that is sensitive to e-cadherin restoration by a src-kinase inhibitor, saracatinib (azd0530). *Cell death & disease*, 4(11):e915, 2013.

[50] Yuting Sun, Bu-Er Wang, Kevin G Leong, Peng Yue, Li Li, Suchit Jhunjhunwala, Darrell Chen, Kyounghee Seo, Zora Modrusan, Wei-Qiang Gao, et al. Androgen deprivation causes epithelial–mesenchymal transition in the prostate: implications for androgen-deprivation therapy. *Cancer research*, 72(2):527–536, 2012.

[51] Mei-Yi Lee, Cheng-Yang Chou, Ming-Jer Tang, and Meng-Ru Shen. Epithelial-mesenchymal transition in cervical cancer: correlation with tumor progression, epidermal growth factor receptor overexpression, and snail up-regulation. *Clinical Cancer Research*, 14(15):4743–4750, 2008.

[52] Junichi Ikenouchi, Miho Matsuda, Mikio Furuse, and Shoichiro Tsukita. Regulation of tight junctions during the epithelium-mesenchyme transition: direct repression of the gene expression of claudins/occludin by snail. *Journal of cell science*, 116(10):1959–1967, 2003.

[53] Yihong Zhu, Mikael Nilsson, and Karin Sundfeldt. Phenotypic plasticity of the ovarian surface epithelium: Tgf-β1 induction of epithelial to mesenchymal transition (emt) in vitro. *Endocrinology*, 151(11):5497–5505, 2010.

[54] Romaine I Fernando, Mary Litzinger, Paola Trono, Duane H Hamilton, Jeffrey Schlom, and Claudia Palena. The t-box transcription factor brachyury promotes epithelial-mesenchymal transition in human tumor cells. *The Journal of clinical investigation*, 120(2):533, 2010.

[55] Junlin Li and Binhua P Zhou. Activation of β-catenin and akt pathways by twist are critical for the maintenance of emt associated cancer stem cell-like characters. *BMC cancer*, 11(1):49, 2011.

[56] Hiroshi Tanaka, Evelyn Kono, Chau P Tran, Hideyo Miyazaki, Joyce Yamashiro, Tatsuya Shimomura, Ladan Fazli, Robert Wada, Jiaoti Huang, Robert L Vessella, et al. Monoclonal antibody targeting of n-cadherin inhibits prostate cancer growth, metastasis and castration resistance. *Nature medicine*, 16(12):1414–1420, 2010.

[57] Seema Sethi, Jill Macoska, Wei Chen, and Fazlul H Sarkar. Molecular signature of epithelial-mesenchymal transition (emt) in human prostate cancer bone metastasis. *American journal of translational research*, 3(1):90, 2011.

[58] Julie Steinestel, Marcus V Cronauer, Johannes Müller, Andreas Al Ghazal, Peter Skowronek, Annette Arndt, Klaus Kraft, Mark Schrader, Andres J Schrader, and Konrad Steinestel. Overexpression of p16 ink4a in urothelial carcinoma in situ is a marker for mapk-mediated epithelial-mesenchymal transition but is not related to human papillomavirus infection. *PLoS One*, 8:e65189, 2013.

[59] Heather Dawson, Viktor H Koelzer, Eva Karamitopoulou, Mary Economou, Caroline Hammer, Dominique-Elisabeth Muller, Alessandro Lugli, and Inti Zlobec. The apoptotic and proliferation

rate of tumour budding cells in colorectal cancer outlines a heterogeneous population of cells with various impacts on clinical outcome. *Histopathology*, 64(4):577–584, 2014.

[60] Sivan Elloul, Mari Bukholt Elstrand, Jahn M Nesland, Claes G Tropé, Gunnar Kvalheim, Iris Goldberg, Reuven Reich, and Ben Davidson. Snail, slug, and smad-interacting protein 1 as novel parameters of disease aggressiveness in metastatic ovarian and breast carcinoma. *Cancer*, 103(8):1631–1643, 2005.

[61] J Chen, N Imanaka, and JD Griffin. Hypoxia potentiates notch signaling in breast cancer leading to decreased e-cadherin expression and increased cell migration and invasion. *British journal of cancer*, 102(2):351–360, 2010.

[62] Hui-Wen Lo, Sheng-Chieh Hsu, Weiya Xia, Xinyu Cao, Jin-Yuan Shih, Yongkun Wei, James L Abbruzzese, Gabriel N Hortobagyi, and Mien-Chie Hung. Epidermal growth factor receptor cooperates with signal transducer and activator of transcription 3 to induce epithelial-mesenchymal transition in cancer cells via up-regulation of twist gene expression. *Cancer research*, 67(19):9066–9076, 2007.

[63] Zhou Yang, Xiaohong Zhang, Haiju Gang, Xiaojun Li, Zumao Li, Tao Wang, Juan Han, Ting Luo, Fuqiang Wen, and Xiaoting Wu. Up-regulation of gastric cancer cell invasion by twist is accompanied by n-cadherin and fibronectin expression. *Biochemical and biophysical research communications*, 358(3):925–930, 2007.

[64] SS Sohal, A Soltani, S Weston, R Wood-Baker, and H Walters. Intermediate filament vimentin and potential role in epithelial mesenchymal transition (emt). In RA de Mello, editor, *Vimentin Concepts and Molecular Mechanisms*. Nova Science Publishers, 2013.

[65] Keun Hur, Yuji Toiyama, Masanobu Takahashi, Francesc Balaguer, Takeshi Nagasaka, Junichi Koike, Hiromichi Hemmi, Minoru Koi, C Richard Boland, and Ajay Goel. Microrna-200c modulates epithelial-to-mesenchymal transition (emt) in human colorectal cancer metastasis. *Gut*, 62(9):1315–1326, 2013.

[66] Ester Sánchez-Tilló, Oriol de Barrios, Laura Siles, Pier G Amendola, Douglas S Darling, Miriam Cuatrecasas, Antoni Castells, and Antonio Postigo. Zeb1 promotes invasiveness of colorectal carcinoma cells through the opposing regulation of upa and pai-1. *Clinical Cancer Research*, 19(5):1071–1082, 2013.

[67] JY Lee, MK Park, JH Park, HJ Lee, DH Shin, Y Kang, CH Lee, and G Kong. Loss of the polycomb protein mel-18 enhances the epithelial–mesenchymal transition by zeb1 and zeb2 expression through the downregulation of mir-205 in breast cancer. *Oncogene*, 33(10):1325–1335, 2014.

[68] Jing Yang and Robert A Weinberg. Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. *Developmental cell*, 14(6):818–829, 2008.

[69] Rehana Qureshi, Himanshu Arora, and MA Rizvi. Emt in cervical cancer: Its role in tumour progression and response to therapy. *Cancer letters*, 356(2):321–331, 2015.

[70] Srishti Kotiyal and Susinjan Bhattacharya. Breast cancer stem cells, emt and therapeutic targets. *Biochemical and biophysical research communications*, 453(1):112–116, 2014.

[71] Linlin Zhang, Min Jiao, Kaijie Wu, Lei Li, Guodong Zhu, Xinyang Wang, Dalin He, and Dapeng Wu. Tnf-α induced epithelial mesenchymal transition increases stemness properties in renal cell carcinoma cells. *Int J Clin Exp Med*, 7(12):4951–4958, 2014.

[72] Li-Jun Chen, Hong Ye, Qian Zhang, Feng-Zhi Li, Lin-Jie Song, Jie Yang, Qing Mu, Shan-Shan Rao, Peng-Cheng Cai, Fei Xiang, et al. Bleomycin induced epithelial–mesenchymal transition (emt) in pleural mesothelial cells. *Toxicology and applied pharmacology*, 283(2):75–82, 2015.

[73] Che Dehai, Pan Bo, Tian Qiang, Shang Lihua, Liu Fang, Jin Shi, Cao Jingyan, Yu Yan, Wang Guangbin, and Yuan Zhenjun. Enhanced invasion of lung adenocarcinoma cells after co-culture with thp-1-derived macrophages via the induction of emt by il-6. *Immunology letters*, 160(1):1–10, 2014.

[74] Xin Yao, Shubha Kale Ireland, Tri Pham, Brandi Temple, Renwei Chen, Madhwa HG Raj, and Hector Biliran. Tle1 promotes emt in a549 lung cancer cells through suppression of e-cadherin. *Biochemical and biophysical research communications*, 455(3):277–284, 2014.

[75] Felicity M Davis, Iman Azimi, Richard A Faville, Amelia A Peters, Kees Jalink, JW Putney, Geoffrey J Goodhill, Erik W Thompson, Sarah J Roberts-Thomson, and Gregory R Monteith. Induction of epithelial–mesenchymal transition (emt) in breast cancer cells is calcium signal dependent. *Oncogene*, 33(18):2307–2316, 2014.

[76] Reviews Microscope Master: Research and Comparisons. Phase contrast microscope, application in microscopy, advantages & disadvantages. `http://www.microscopemaster.com/phase-contrast-microscope.html`, November 2011. (Visited on 10/14/2015).

[77] Reviews Microscope Master: Research and Comparisons. Microscopy imaging techniques. `http://www.microscopemaster.com/microscopy-imaging-techniques.html`, November 2011. (Visited on 10/15/2015).

[78] Reviews Microscope Master: Research and Comparisons. Immunofluorescence in microscopy - applications, direct and indirect methods. `http://www.microscopemaster.com/immunofluorescence-microscopy.html`, November 2011. (Visited on 10/15/2015).

[79] Nmumg — sigma-aldrich. `http://www.sigmaaldrich.com/catalog/product/sigma/94081121?lang=en&region=US`. (Visited on 10/20/2015).

[80] Shaheen Ahmed and Ali Nawshad. Complexity in interpretation of embryonic epithelial-mesenchymal transition in response to transforming growth factor-β signaling. *Cells, tissues, organs*, 185(1-3):131, 2007.

[81] Frans Van Roy and Geert Berx. The cell-cell adhesion molecule e-cadherin. *Cellular and molecular life sciences*, 65(23):3756–3788, 2008.

[82] Johanna Ivaska, Hanna-Mari Pallari, Jonna Nevo, and John E Eriksson. Novel functions of vimentin in cell adhesion, migration, and signaling. *Experimental cell research*, 313(10):2050–2062, 2007.

[83] Samy Lamouille and Rik Derynck. Cell size and invasion in tgf-β–induced epithelial to mesenchymal transition is regulated by activation of the mtor pathway. *The Journal of cell biology*, 178(3):437–451, 2007.

[84] Christine Vogel and Edward M Marcotte. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics*, 13(4):227–232, 2012.

[85] Gene quantification & mrna analysis methods & mrna quantification. `http://www.gene-quantification.com/mrna.html`. (Visited on 09/24/2015).

[86] F Ishikawa, H Miyoshi, K Nose, and M Shibanuma. Transcriptional induction of mmp-10 by tgf-β, mediated by activation of mef2a and downregulation of class iia hdacs. *Oncogene*, 29(6):909–919, 2010.

[87] K Horiguchi, K Sakamoto, D Koinuma, K Semba, A Inoue, S Inoue, H Fujii, A Yamaguchi, K Miyazawa, K Miyazono, et al. Tgf-β drives epithelial-mesenchymal transition through δef1-mediated downregulation of esrp. *Oncogene*, 31(26):3190–3201, 2012.

[88] Miki Kondo, E Cubillo, K Tobiume, T Shirakihara, N Fukuda, H Suzuki, K Shimizu, K Takehara, A Cano, M Saitoh, et al. A role for id in the regulation of tgf-β-induced epithelial–mesenchymal transdifferentiation. *Cell Death & Differentiation*, 11(10):1092–1101, 2004.

[89] Ulrich Valcourt, Marcin Kowanetz, Hideki Niimi, Carl-Henrik Heldin, and Aristidis Moustakas. Tgf-β and the smad signaling pathway support transcriptomic reprogramming during epithelial-mesenchymal cell transition. *Molecular biology of the cell*, 16(4):1987–2002, 2005.

[90] Stephen A Bustin, Vladimir Benes, Jeremy A Garson, Jan Hellemans, Jim Huggett, Mikael Kubista, Reinhold Mueller, Tania Nolan, Michael W Pfaffl, Gregory L Shipley, et al. The miqe guidelines: minimum information for publication of quantitative real-time pcr experiments. *Clinical chemistry*, 55(4):611–622, 2009.

[91] Avri Ben-Ze'ev, Gregory S Robinson, NL Bucher, and Stephen R Farmer. Cell-cell and cell-matrix interactions differentially regulate the expression of hepatic and cytoskeletal genes in primary cultures of rat hepatocytes. *Proceedings of the National Academy of Sciences*, 85(7):2161–2165, 1988.

[92] Toru Sakairi, Yoshifusa Abe, Parmijit S Jat, and Jeffrey B Kopp. Cell-cell contact regulates gene expression in cdk4-transformed mouse podocytes. *American Journal of Physiology-Renal Physiology*, 299(4):F802–F809, 2010.

[93] J Bizik, E Kankuri, A Ristimäki, A Taieb, H Vapaatalo, W Lubitz, and A Vaheri. Cell–cell contacts trigger programmed necrosis and induce cyclooxygenase-2 expression. *Cell Death & Differentiation*, 11(2):183–195, 2004.