Research Report

Mechanisms of Hereditary Diffuse Gastric Cancer Initiation

The role of E-cadherin in the Epithelial-Mesenchymal Transition

Henry Beetham

A report submitted in partial fulfilment of the Bachelor of Biomedical Sciences with Honours (Molecular Basis of Health and Disease) University of Otago, Dunedin, New Zealand

November 2012
ABSTRACT

Hereditary diffuse gastric cancer (HDGC) is a dominantly inherited cancer syndrome caused by germline mutations in the CDH1 gene which encodes the calcium-dependent, cell-cell adhesion protein E-cadherin [1,2]. Early stage HDGC is characterized by multiple foci of stage T1a signet ring cell carcinomas (SRCC). These foci are relatively indolent [3], however some eventually undergo an epithelial-mesenchymal transition (EMT) changing them into aggressive mesenchymal cells. At present it is not known definitively how the loss of E-cadherin and CDH1 germline mutations contribute to tumour initiation primarily in HDGC. The aim of this research was to determine whether the abrogation of E-cadherin expression alone is sufficient to induce an EMT, or if other mechanisms are also required. In order to analyse the early events of HDGC in vitro, an isogenic breast cell line model system with CDH1 knockout (-/-) cells was chosen. E-cadherin loss was found to affect cell morphology, proliferation rate, nucleoli number and cell adherence of the mammary epithelial cells. Transcriptome profiling performed on the isogenic cell lines showed E-cadherin loss resulted in an upregulation of genes involved in the tight junction complex such as claudin and occludin. The majority of EMT markers expected to be involved in cells undergoing an EMT were not upregulated. CDH1 loss also affected migration rates and growth in 3D culture. Overall, this study showed that E-cadherin loss alone was insufficient to cause a complete EMT in our model system. However, some genes associated with gastric and breast cancer progression were differentially expressed: S100-calcium binding proteins, matrix metallopeptidases (MMPs), and mucins. Hence, these CDH1 -/- cells show an increased ‘cancer-like’ phenotype but still remain relatively indolent, at least in the observed time frame.
ACKNOWLEDGEMENTS

I would like to thank my supervisor Parry Guilford, for all his literary wisdom, good nature and humour. Cheers to all of the members of the Cancer Genetics Lab for making the lab an enjoyable workplace and great social environment from squash games and half marathons to a weekend in Queenstown! Especially Augustine Chen who provided me with guidance, and helped with day to day questions and problems. Special thanks to: Joanne Guest for being a great bench buddy and for help with my Western blot. Sofie van Huffel for her expertise on soft agar. George Wiggins, Tanis Godwin, Augustine Chen and Anne Scott for proof-reading my thesis. Mik Black, Chris Hakkaart and Tyler McInnes for their help with learning R programming language. Bryony Telford for her input to my thesis. And Luxmanan Selvanesan for randomizing my nucleoli images. Thanks to everyone in the CGL lab for their patience and help making my honours research year a great learning experience.

To my parents thank you for your support in whatever I choose to do. And lastly to my friends for keeping me motivated, determined and stress-free (mostly).
TABLE OF CONTENTS

Abstract........................................................................................................................................... i
Acknowledgements......................................................................................................................... ii
Table of Contents .............................................................................................................................. iii
List of Figures .................................................................................................................................. vi
List of Tables .................................................................................................................................... vii
Abbreviations ................................................................................................................................... viii
1 Introduction ................................................................................................................................. 1
  1.1 Gastric Cancer ......................................................................................................................... 1
    1.1.1 Pathology and Clinical Features ...................................................................................... 2
  1.2 The role of E-cadherin ............................................................................................................. 5
  1.3 The second CDH1 hit ............................................................................................................... 7
    1.3.1 Potential causes of the second CDH1 hit ......................................................................... 7
  1.4 The Epithelial-Mesenchymal Transition (EMT) ...................................................................... 7
  1.5 Does CDH1 downregulation alone cause an EMT? ............................................................... 10
Overall Aim ..................................................................................................................................... 12
Experimental Background .............................................................................................................. 12
2 Materials and Methods ............................................................................................................... 13
  2.1 Materials ............................................................................................................................... 13
    2.1.1 Reagents ....................................................................................................................... 13
    2.1.2 Equipment .................................................................................................................... 15
    2.1.3 Software ....................................................................................................................... 16
    2.1.4 Cell lines ...................................................................................................................... 17
  2.2 Cell culture ............................................................................................................................. 17
    2.2.1 Complete Growth Media ............................................................................................... 17
    2.2.2 Cell counting ............................................................................................................... 19
    2.2.3 Thawing cell lines ........................................................................................................ 19
    2.2.4 Cryogenic preservation ................................................................................................. 20
    2.2.5 Cell Culture Maintenance ............................................................................................ 20
  2.3 Western Blotting .................................................................................................................... 21
    2.3.1 Determination of Protein concentration ...................................................................... 21
    2.3.2 SDS-PAGE .................................................................................................................. 22
    2.3.3 Western analysis .......................................................................................................... 22
  2.4 Cell Density assay .................................................................................................................. 23
3 Results .......................................................................................................................... 31
  3.1 Characterisation of in vitro phenotypes resulting from CDH1-/- .......................................................................................... 31
    3.1.1 Verification of E-cadherin loss .................................................................................. 31
    3.1.2 Cellular Morphology ............................................................................................... 33
    3.1.3 Proliferation rates of MCF-10A WT and MCF-10A CDH1-/- .................................. 34
    3.1.4 Nucleoli numbers .................................................................................................... 37
  3.2 Gene Expression profiling of MCF-10A WT and MCF-10a CDH1-/- .................................. 39
    3.2.1 Quantitation of RNA in samples ............................................................................... 39
    3.2.2 Quality of RNA in samples ...................................................................................... 41
    3.2.3 Transcriptome sequencing (NZGL) ......................................................................... 44
  3.3 Impact of E-cadherin loss in a mammary epithelial cell line ................................................................................................. 45
    3.3.1 Distribution of differentially expressed genes for MCF-10A WT and MCF-10A CDH1-/- .................................................. 45
    3.3.2 Most differentially expressed genes ........................................................................ 47
    3.3.3 GO analysis ............................................................................................................ 50
    3.3.4 E-cadherin loss effects genes involved in Cell adhesion ......................................... 51
  3.4 Genes associated with cancer ...................................................................................... 53
  3.5 Gene criteria for an EMT ............................................................................................ 56
  3.6 Cell migration is affected by E-cadherin loss ............................................................. 61
  3.7 The effect of CDH1-/- on cell anchorage independent growth .................................... 65
  3.8 The differentiation capability of the two cell lines ....................................................... 67
4 Discussion ...................................................................................................................... 70
  4.1 Cell line characterisation ......................................................................................... 70
  4.2 Cellular Impact of E-cadherin loss ............................................................................. 71
    4.2.1 Cell adhesion ......................................................................................................... 72
    4.2.2 CDH1 loss causus reduced migration ................................................................... 74
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Proposed model for early progression in E-cadherin mutation carriers</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>HE section of mucosa containing early HDGC</td>
<td>3</td>
</tr>
<tr>
<td>1.3</td>
<td>Anatomical map of stomach showing size and location of SRCC foci</td>
<td>4</td>
</tr>
<tr>
<td>1.4</td>
<td>Schematic overview of the E-cadherin-catenin complex</td>
<td>6</td>
</tr>
<tr>
<td>1.5</td>
<td>The cycle of EMT and MET</td>
<td>8</td>
</tr>
<tr>
<td>1.6</td>
<td>Schematic of overlay method for 3D culture of MCF-10A cells on Matrigel</td>
<td>30</td>
</tr>
<tr>
<td>2.1</td>
<td>Western blot showing E-cadherin is not present in MCF-10A CDH1/-</td>
<td>32</td>
</tr>
<tr>
<td>2.2</td>
<td>Differential growth morphology between MCF-10A WT and MCF-10A CDH1/-</td>
<td>33</td>
</tr>
<tr>
<td>2.3</td>
<td>Three-dimensional culture of MCF-10A WT and MCF-10A CDH1/- cells</td>
<td>34</td>
</tr>
<tr>
<td>2.4</td>
<td>Cell density assay</td>
<td>35</td>
</tr>
<tr>
<td>2.5</td>
<td>Exponential curve regression</td>
<td>37</td>
</tr>
<tr>
<td>2.6</td>
<td>Randomised, double blinded analysis of nucleoli number</td>
<td>38</td>
</tr>
<tr>
<td>3.1</td>
<td>Nanodrop, Qubit and Bioanalyser assays were used to quantify total RNA</td>
<td>40</td>
</tr>
<tr>
<td>3.2</td>
<td>DNase1 treated samples have insufficient RNA quality</td>
<td>42</td>
</tr>
<tr>
<td>3.3</td>
<td>RNA extraction Experiments I and II have similar RNA quality</td>
<td>43</td>
</tr>
<tr>
<td>3.4</td>
<td>1,117 genes are significantly differentially expressed between MCF-10A CDH1/- and MCF-10A WT</td>
<td>46</td>
</tr>
<tr>
<td>3.5</td>
<td>Representative ‘wide view’ of a HiQ4 dish in the BioStation under phase contrast</td>
<td>62</td>
</tr>
<tr>
<td>3.6</td>
<td>Diagram of the wounding process</td>
<td>62</td>
</tr>
<tr>
<td>3.7</td>
<td>MCF-10A WT cells migrate faster than MCF-10A CDH1/-</td>
<td>64</td>
</tr>
<tr>
<td>3.8</td>
<td>Soft agar assay to determine anchorage-independent growth</td>
<td>66</td>
</tr>
<tr>
<td>3.9</td>
<td>Day 12 ‘wide view’ of MCF-10A WT and MCF-10A CDH1/- cells grown in Matrigel</td>
<td>68</td>
</tr>
<tr>
<td>3.10</td>
<td>The various proteins associated with tight junctions, adherens junctions and desmosomes</td>
<td>73</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

TABLE 2.1. Medium Recipe for MCF-10A WT and MCF-10A CDH1 -/- ................................................................. 18
TABLE 2.2. Plating densities for T25 and T75 Flasks.......................................................................................... 21
TABLE 2.3. Antibodies for immunohistochemistry............................................................................................. 23
TABLE 2.1. Quality control on library preparations ........................................................................................ 44
TABLE 2.2. Demultiplex statistics: ....................................................................................................................... 45
TABLE 2.3. Top 20 most differentially expressed genes for positive and negative FC. ............................. 48
TABLE 3.4. Gene Ontology analysis. .................................................................................................................. 51
TABLE 3.5. Epithelial cell-cell adhesion proteins. ............................................................................................ 52
TABLE 3.6. S100-calcium binding proteins........................................................................................................ 54
TABLE 3.7. Matrix Metallopeptidases ................................................................................................................. 55
TABLE 3.8. Mucins ............................................................................................................................................... 55
TABLE 3.9. Biomarkers for EMT. ....................................................................................................................... 58
TABLE 3.10. Epithelial and Mesenchymal Markers. .......................................................................................... 60
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>°C</td>
<td>Degrees Celsius</td>
</tr>
<tr>
<td>Ab</td>
<td>Antibody</td>
</tr>
<tr>
<td>adj</td>
<td>Adjusted</td>
</tr>
<tr>
<td>APS</td>
<td>Ammonium persulphate</td>
</tr>
<tr>
<td>ATCC</td>
<td>American Type Culture Collection</td>
</tr>
<tr>
<td>BH</td>
<td>Benjamini &amp; Hochberg</td>
</tr>
<tr>
<td>CDH1</td>
<td>E-cadherin</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>CO₂</td>
<td>Carbon dioxide</td>
</tr>
<tr>
<td>DAPI</td>
<td>4',6-diamidino-2-phenylindole</td>
</tr>
<tr>
<td>DMEM-F12</td>
<td>Dulbecco's modified Eagle's medium and F12 medium</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl Sulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNase</td>
<td>Deoxyribonuclease</td>
</tr>
<tr>
<td>ECL</td>
<td>Enhanced chemiluminescence</td>
</tr>
<tr>
<td>E-cadherin</td>
<td>Epithelial cadherin</td>
</tr>
<tr>
<td>ECM</td>
<td>Extracellular matrix</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EGF</td>
<td>Epidermal growth factor</td>
</tr>
<tr>
<td>EMT</td>
<td>Epithelial mesenchymal transition</td>
</tr>
<tr>
<td>FBS</td>
<td>Fetal bovine serum</td>
</tr>
<tr>
<td>FC</td>
<td>Fold change</td>
</tr>
<tr>
<td>GATHER</td>
<td>Gene annotation tool to help explain relationships</td>
</tr>
<tr>
<td>g</td>
<td>gram(s)</td>
</tr>
</tbody>
</table>
Gb = Gigabyte
HCL = Hydrochloric acid
HDGC = Hereditary diffuse gastric cancer
HE = Hematoxylin and eosin
HRP = Horseradish peroxidase
KO = Knockout (−/−)
L = Litre(s)
m = metre(s)
M = Moles
mg = milligram(s)
 mL = millilitre(s)
mM = millimoles per litre
mRNA = messenger RNA
MTT = Dimethyl thiazolyl diphenyl tetrazolium salt
MMPs = Matrix metalloproteinases
MET = Mesenchymal- Epithelial Transitions
n = nano
nm = nanometre
NaOH = Sodium hydroxide
NZGL = New Zealand Genomics Limited
pM = picomolar
PCR = Polymerase chain reaction
PBS = Phosphate buffered saline
PBS-T = PBS 0.2% tween
PVDF = Polyvinylidene difluoride
RNA = Ribonucleic acid
RNase = Ribonuclease
RPNI = Roswell Park Memorial Institute
RT = Room temperature
s = second
SDS-PAGE = Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM = Standard error of the mean
TBS = Tris buffered saline
TEMED = Tetramethylethylenediamine
TIMPs = Tissue inhibitors of metalloproteinases
TFs = Transcription factors
Tris = Tris(hydroxymethyl)aminomethane
Tween20 = Polysorbate 20
USA = United States of America
μ = micro
μg = micrograms
μL = microlitre(s)
μm = micrometre(s)
WT = Wild type
Vs. = versus
Science is facts; just as houses are made of stones, so is science made of facts; but a pile of stones is not a house and a collection of facts is not necessarily science –Henri Poincaré.
1 INTRODUCTION

1.1 GASTRIC CANCER

Gastric cancer (GC) is the fourth most common cancer worldwide and the second leading cause of cancer mortality in the world with 738,000 deaths estimated in 2008 [4]. Incidence rates can vary considerably between countries. For example, in Egypt the age-standardised rates of stomach (gastric) cancer vary from 3.3 and 2.0 cases per 100,000 in men and women, respectively, compared to rates of 65.9 in men and 25.9 in women in Korea [5]. In New Zealand the incidence is 9.1 per 100,000 for males and 4.5 for females [5]. GC can be classified into two main histological types, the intestinal and the diffuse forms. Although the incidence of intestinal GC has been shown to be decreasing, the incidence of the diffuse form is increasing [6,7].

The notion that diffuse GC may have a genetic link was strongly voiced in 1964 after a Māori family was studied for their predisposition to diffuse gastric cancer (DGC) [8], and subsequently the cancer syndrome hereditary diffuse gastric cancer (HDGC) was characterised in this family in 1998 [1,9]. HDGC is an autosomal dominant cancer caused by a germline mutation in the $CDH1$ gene [1]. 30% to 50% of patients with a clinical diagnosis of HDGC [10,11] have germline $CDH1$ mutations with the remainder due to either mutations in as yet unidentified genes or chance groupings. $CDH1$ encodes the calcium-dependent, cell-cell adhesion protein E-cadherin [1,2]. Other cancers have been reported in $CDH1$ mutation carriers including colorectal, prostate [12], and lobular breast cancer (LBC); but only LBC is over-represented in HDGC individuals [13].
The first discovered germline mutation of *CDH1* was a G1008T nucleotide substitution in the donor splice consensus sequence of exon 7 [1]. There have since been approximately 100 germline mutations published in the *CDH1* gene, with the most common types of mutations being deletions or small insertions [14]. The lifetime risk of developing GC in individuals with *CDH1* germline mutations is approximately 80% (95% CI) [12]. Female mutation carriers have an elevated risk of developing lobular breast cancer [12], although the actual increase in risk has yet to be defined accurately (P. Guilford, personal communication).

1.1.1 PATHOLOGY AND CLINICAL FEATURES

Of the two major histopathological variants for GC [15], the intestinal form has a greater association with environmental risk factors and has components of glandular, solid or intestinal architecture. It is associated with progressive pre-morbid histological changes in the gastric epithelium, which undergoes metaplasia to resemble the epithelium of the small and large intestine [16]. The diffuse form however, has a stronger genetic etiology and is composed of poorly differentiated cells which can infiltrate the gastric wall, leading to extensive thickening and rigidity, in a disease process known as *linitus plastica* [17]. Diffuse GC can be further characterised to include infiltrating carcinomas [18], isolated cell carcinomas [19], signet ring cell carcinomas (SRCC) [20], and *in situ* signet ring cells [21]. A disease model for *CDH1* mutational carriers with *in situ* signet ring cells can be found in Figure 1.1. A loss of E-cadherin within one or more cells lining the gland leads to the ‘pagetoid’ spread of *in situ* signet ring cells below the maintained epithelial lining of the gland and eventually, into the lamina propria [22]. However the relevance of this model to diffuse GC progression has not yet been proven.
Early stage HDGC is characterized by multiple foci of stage T1a SRCC, which have broken through the basement membrane into the lamina propria but have not breached the underlying muscularis mucosae (Figure 1.2). The relationship between \textit{in situ} and T1a foci remains to be determined [14].

Figure 1.2. HE section of mucosa containing early HDGC. The white arrow shows a T1a signet-ring-cell and the white box shows mesenchyme-like cells present in some of the large SRCC foci. Adapted from Humar \textit{et al.} [20].
The largest numbers of signet ring cell foci in germline CDH1 mutation carriers are often seen in the body-antral transitional zone [23] which is between the body and antral gastric mucosal zones (Figure 1.3). These foci are relatively indolent, hypo-proliferative and lack c-Src pathway activation [3]. However, some of these indolent foci eventually undergo an epithelial-mesenchymal transition (EMT) turning them into aggressive mesenchymal cells (Figure 1.2).

Figure 1.3. Anatomical map of stomach showing size and location of SRCC foci and mucosal zones in a 28-year-old male CDH1 germline mutation carrier. The foci are to scale, except for foci of less than 1 mm, which are shown as dots for visibility. F, defect caused by removal of fresh tissue from cryostorage. Adapted from Charlton et al, [23].

Strictly speaking, a cancer cluster cannot be called ‘hereditary’ in the absence of an identified causative germline mutation and should be called ‘familial’. However families meeting the clinical criteria of HDGC are known as HDGC families not familial DGC families even if a CDH1 mutation has not been found [24]. At present, HDGC is diagnosed in a family if there have been: a) two or
more documented DGC cases with at least one under the age of 50 years; b) three or more documented cases independent of age, c) one documented case under 40 years of age; or d) one documented case under 50 years of age and one LBC case [24].

Prophylactic total gastrectomy is the single option to abolish an inherited risk of gastric cancer. Patients with early DGC who undergo gastrectomy have a greater than 90% five year survival rate, but DGC has few specific symptoms so most often patients are diagnosed at a late stage by which time five year survival is reduced to 20% [25]. Unfortunately, all post-gastrectomy patients develop complications resulting from surgery [26], although increasingly these complications are being better managed leading to better quality of life even in younger patients.

1.2 THE ROLE OF E-CADHERIN

E-cadherin (CDH1) is the founding member of the cadherin superfamily, including N- (CDH2), P- (CDH3) and R- (CDH4) cadherins [27]. The cadherin family consists of five major subfamilies: classical cadherins of type 1 (Including CDH1), closely related cadherins of type 2, desmosomal cadherins, protocadherins, and cadherin-related molecules [28]. E-cadherin is a calcium-dependent transmembrane glycoprotein, expressed predominantly at the adherens junction in the epithelium (Figure 1.4). The mature E-cadherin contains a single transmembrane domain, a cytoplasmic domain of around 150 amino acids, and an ectodomain of around 550 amino acids comprising of five tandem repeats [27]. The cytoplasmic domain is associated with p120-catenin, β-catenin or plakoglobin, and α-catenin. The interactions with α-catenin facilitate direct linking to the cytoskeleton (Figure 1.4) [29].
There are several cancers associated with the loss of E-cadherin such as colorectal, prostate [12], and LBC [13]. E-cadherin has long been regarded as an invasion suppressor due to its frequent downregulation in sporadic tumours during invasion and metastasis [20].

At present it is unclear why the loss of E-cadherin contributes to tumour initiation in the stomachs of HDGC patients as opposed to other tissues. Several theories have been proposed, these include increased carcinogen exposure resulting in further genetic and epigenetic damage at the gastric epithelium, and H. pylori infection; high cellular turnover of the gastric mucosa and events of tissue remodelling; and the diffusion of carcinomas as a result of proteolytic enzymes [3].

Figure 1.4. Schematic overview of the E-cadherin-catenin complex at the junction between two neighbouring epithelial cells. Adapted from Roy and Berx [30].
1.3 THE SECOND CDH1 HIT

The foci discovered in the stomachs of HDGC patients (Figure 3) often do not have mutations or deletions in the second CDH1 allele, although these observations have been observed in advanced tumours [31]. In both stage T1a SRCC [32] and late stage HDGC tumours [33], the ‘second CDH1 hit’ is caused by DNA promoter hypermethylation in at least 50% of patients. In lobular breast cancer, promoter hypermethylation also seems to be involved in CDH1 downregulation [34]. Interestingly, transcriptional repressors of CDH1 such as Snail, Slug, ZEB1 and ZEB2 all promote histone deactylase activity [32], suggesting that the ‘second CDH1 hit’ could also be caused by histone modification [35,36].

1.3.1 POTENTIAL CAUSES OF THE SECOND CDH1 HIT

It is not currently known whether the ‘second CDH1 hit’ is exclusively a stochastic event or if it is caused by environmental or physiological pressures. Physiological and pathological processes which have been described to promote the ‘second CDH1 hit’ thereby inducing E-cadherin downregulation include: repair processes [37,38], H. pylori infection [39], the inflammatory response [40,41] and hypoxia [42,43].

1.4 THE EPITHELIAL-MESENCHYMAL TRANSITION (EMT)

The term EMT is defined by a series of events in which epithelial cells are converted into mesenchymal cells by losing their epithelial characteristics and acquiring mesenchymal ones. Important events during the progression of EMT include the dysregulation of the tight junctions, adherens junctions and desmosome, resulting in the loss of cellular adhesion, a gain of fibroblast-like morphology and increase in invasiveness, motility and metastatic capabilities. A number of markers have now been identified that are characteristic of either epithelial (Figure 1.5, Box 1) or
mesenchymal cells (Figure 1.5, Box 2). A loss of epithelial markers such as claudins and occludins and gain of mesenchymal markers are indicative of an EMT event.

Figure 1.5. The cycle of EMT and MET. The diagram shows the cycle of events during which epithelial cells are transformed into mesenchymal cells and vice versa. Taken from Thiery et al., [44].

EMT is important developmentally and is essential for the tissue remodelling that occurs during embryogenesis and gastrulation [44]. It has been suggested that there are three definitive types of epithelial to mesenchymal transitions [45]:

- Type 1) When nascent epithelial cells undergo an EMT to migrate and then reform secondary epithelial cells via a mesenchymal-epithelial transition (MET), following the basic body plan during metazoan development and gastrulation.
• Type 2) When epithelial or endothelial cells form fibroblasts in response to persistent inflammation.

• Type 3) When epithelial carcinoma cells transition into metastatic tumour cells.

EMT induction in the pathogenesis of carcinomas is associated with invasion and carcinoma progression as well as resistance to cancer therapeutics [46]. A ‘complete EMT’ can be defined as a metastable, fibroblastoid phenotype with a loss of E-cadherin and an acquisition of vimentin. A complete EMT is most closely correlated with metastasis and local invasion [47]. The evidence supporting a role for EMT in tumour progression comes from phenotypic, genetic and functional studies. One such phenotypic example is that the expression of markers that are characteristic of mesenchymal cells, such as vimentin, SNAI1 and SNAI2, and loss of epithelial markers, for example E-cadherin, correlates with tumour progression and poor prognosis [44]. Furthermore, the invasion of adenocarcinomas is accompanied by the release of single cells through an EMT [48]. A functional case for this argument is that the manipulation of E-cadherin expression in cancer cell lines triggers both invasive behaviour and change of phenotype [48-50].

Importantly, in order for carcinoma cells to acquire metastatic proficiency these cells develop mesenchymal gene-expression patterns and morphological features. This results in altered adhesive properties and the initiation of proteolysis and motility; therefore allowing the tumour cells to metastasize [51]. In many carcinoma cell lines this sequence of events is followed by a complete or partial EMT, resulting in the gain of metastatic properties in vivo [44]. The EMT is also thought to occur during the progression of HDGC. In particular, the invasion of SRCC lesions beyond the gastric mucosa is proposed to be associated with an EMT [20].
1.5 DOES CDH1 DOWNREGULATION ALONE CAUSE AN EMT?

There is much controversy in the literature on whether loss of E-cadherin alone can induce an EMT [29,52]. Downregulation of E-cadherin has been shown to decrease the strength of cellular adhesion within a tissue, which is thought to result in an increase in cellular motility, enabling cancer cells to cross the basement membrane and invade surrounding tissues [29]. However, there is no clear link between the loss of cell adhesion caused by downregulation of E-cadherin and the tumorigenic consequences characterized in vivo. E-cadherin mediated adhesion is necessary however, in Drosophila for the control of both attachment of stem cells to their niches [53] and the correct spindle orientation during asymmetric divisions required to separate progeny [54].

While the disruption of E-cadherin’s cell-cell contacts alone does not enable metastasis in immortalized human breast epithelial cells, the loss of function of the E-cadherin protein has been shown to enable metastasis through induction of an EMT [29,55].

The support for CDH1 loss as a sole determinant of EMT is not unanimous. Several studies have shown E-cadherin downregulation alone does not seem sufficient to induce a complete EMT [20,52,56]. Evidence supporting this view also comes from murine examples where the loss of E-cadherin induced pre-cancerous lesions in the form of SRCC in the gastric mucosa but was not sufficient for its malignant conversion [32,52]. Thus there is only a causal relationship established between E-cadherin deficiency and the initiation of gastric SRCC in mice, but not further tumour progression.

Consequently, there are various theories for the likely mechanisms behind the EMT if E-cadherin downregulation alone is not sufficient. These include another EMT inducer such as kinase c-Src and its downstream targets FAK and Stat-3 [20]; dysfunction of other tumour suppressor genes such as p53 [57]; DNA damage [44] or over expression of oncogenes such as c-myc [58]. Some
proteins that regulate the EMT, such as p120 [59], Twist [60], LOXL2 and Snail [61] may also independently contribute to a malignant phenotype. The variety of possible mechanisms behind the induction of an EMT explains why there is such unpredictability with the onset of HDGC, as it is not known what causes the sudden shift to aggressive cells. This is a major gap in the current knowledge of HDGC. Better characterisation of the EMT and E-cadherin’s role within this context is vital to gain a clearer understanding of the HDGC disease process. This may allow pathologists to ascertain the likelihood of progression of T1a foci from benign to malignant and provide a better prediction of HDGC risk for CDH1 mutation carriers.
OVERALL AIM

The aim of my research was to determine whether abrogation of E-cadherin expression alone is sufficient to induce an EMT.

EXPERIMENTAL BACKGROUND

It is not possible to analyse all of the early events of human HDGC in vivo, hence there is a need for an in vitro model. As there are no non-malignant gastric cell lines and the EMT is also observed in breast cancer [62], a mammary epithelial cell line model system was chosen to fulfil this role. Both gastric and mammary cell types are likely to have similar mechanisms for an EMT as both lobular breast cancer and gastric tumours have CDH1 gene mutations and are overrepresented in HDGC individuals.

MCF-10A is an immortal, non-transformed mammary epithelial cell line derived from the breast tissue of a 36-year-old woman [63]. The MCF-10A cell line is able to form polarized and glandular structures in 3D culture that resemble the glandular structure of the epithelium in vivo [64]. It is therefore an optimal system to understand how the regulation of differentiation influences the structure and function of glandular epithelium both in early tumour formation and the normal state [64].
2 MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 REAGENTS

0.5M Tris-HCl pH 6.8 – Prepared in the lab (see Appendix A)

1.5 M Tris-HCl pH 8.8 – Prepared in the lab (see Appendix A)

2-Mercaptoethanol – MERCK, USA

30% acrylamide – Bio-Rad, USA

5100 Cryo 1°C Freezing Container – Thermo Fisher Scientific, USA

Actrapid Penfil Neutral Insulin – Novo Nordisk Pharmaceuticals Ltd, New Zealand

Absolute ethanol – Scharlau, Spain

Agar (A0949) – Applichem, USA

Ammonium Persulphate (APS) – BDH Laboratory Supplies, England

Anti E-Cadherin Antibody (Rabbit) – Santa-Cruz SC-7870, USA

Anti-Beta-Actin Antibody (Rabbit) – Santa-Cruz SC-130656, USA

BD Matrigel (356230) – BD Biosciences, USA

Bromophenol blue – Scharlau, Australia

Blocking Buffer – Prepared in the lab (see Appendix A)

Cell culture lysis Buffer – Promega, USA

Cholera Toxsin – Sigma-Aldrich, USA

Dimethyl Sulfoxide (DMSO) – Sigma-Aldrich, USA

Dulbecco’s modified Eagle’s medium and F12 medium (DMEM-F12) – Invitrogen, USA

ECLplus reagents – GE healthcare, USA

Epidermal growth factor (EGF) – Peprotech, USA

Eagle’s minimum essential medium (RPMI 1640) – Invitrogen, USA

Fetal Bovine Serum (FBS) – Invitrogen, USA
Glycerol – Scharlau, Australia

HRP-linked Goat Anti Rabbit polyclonal IgG – Santa-Cruz SC-2004, USA

Horse Serum – Invitrogen, USA

Hydrocortisone – Sigma- Aldrich, USA

Hoechst dye (H33258) – Sigma-Aldrich, USA

Ice-cold freezing media – Prepared in the lab (see Appendix A)

Low Fat Milk Powder – Pams, New Zealand

MTT – Sigma-Aldrich, USA

Novex Sharp pre-stained Protein Standards – Invitrogen, USA

PBS-Tween (PBS-T) – Prepared in the lab (see Appendix A)

Protein loading buffer – Prepared in the lab (see Appendix A)

Phosphate Buffered Saline (PBS) – Prepared in the lab (see Appendix A)

Penicillin/Streptomycin (100X) – Invitrogen, USA

Quick-RNA miniprep kit – Zymo Research, USA

Running Buffer – Prepared in the lab (see Appendix A)

Sodium dodecyl sulphate (SDS) – BDH Laboratory Supplies, England

TEMED – Scharlau, Spain

Tween20 – Sigma-Aldrich, USA

Transfer Buffer for Western blot – Prepared in the lab (see Appendix A)

Ultra-pure distilled water – Invitrogen, USA

TruSeq™ RNA Sample Preparation v2 – Illumina, USA

Trypan blue – Sigma-Aldrich, USA
2.1.2 EQUIPMENT

0.5 mL optical-grade PCR tubes, Invitrogen, USA

1.5 mL Microtube – Axygen, USA

1-200 μL gel loading tip round – Sigma-Aldrich, USA

0.6 mL Microtube – Axygen, USA

15mL and 50mL Falcon Tubes – BD Biosciences, USA

25 mL Falcon Cell Culture Flasks – BD Biosciences, USA

75 mL Cellstar Flasks – Greiner Bio-One, Germany

Automated cell counter – Bio-Rad, USA

Agilent RNA 6000 Pico kit – Agilent, USA

Agilent RNA 6000 Nano kit – Agilent, USA

Agilent 2100 Bioanalyzer – Agilent, USA

Chip priming station – Agilent, USA

Centra 3C Centrifuge – International Equipment Company, USA

Cryovials – Greiner bio-one, Germany

CO₂ cell culture incubator – Binder, Germany

Digital dry-block heater – Bio Products, USA

Falcon Cell Scrapers – BD Biosciences, USA

Falcon Strippettes – BD Biosciences, USA

Falcon 6 well plates – BD Biosciences, USA

Haemocytometer – Propper, USA

Hi 9321 Microprocessor pH meter – Biolab Scientific, New Zealand
HiQ4 dishes – Ibidi, Germany

HiSeq 2000 – Illumina, USA

Immun - Blot PVDF membrane – Bio-Rad, USA

IKA vortex mixer – Thermo Fisher Scientific, New Zealand and Australia

LAS-3000 ECL Imaging system – Thermo Fisher Scientific, New Zealand and Australia

Lab-Tek II Chambered Coverglass (155409) – Thermo Fisher Scientific, New Zealand and Australia

LSM 510 confocal laser scanning microscope – Zeiss, Germany

Milli-Q Ultrapure Water Purification System – Millipore, USA

Mini-Protean II system – Bio-Rad, USA

Mini-TransBlot system – Bio-Rad, USA

Microcentrifuge – Eppendorf, Germany

Nikon BioStation IM – Nikon, USA

Nanodrop – Nanodrop technologies, USA

Olympus CK2 microscope – Olympus, New Zealand

Olympus BX57 – Olympus, New Zealand

Qubit 2.0 Fluorometer – Invitrogen, USA

RNA Clean and Concentrator Kit (R1017 and R1018) – Zymo, USA

Tissue Culture Hood – EMAIL, Australia

Thick Criterion Sheet – Bio-Rad, USA

-------------------------------------------------------------------------------------------------------------------------------

2.1.3 SOFTWARE

EdgeR [65] was used in the differential expression analysis of RNA sequencing data.

Fiji Is Just Image J (FIJI) [66] was used to create, view and quantify images from the BioStation.
GraphPad Prism version 5.00 for Windows [67] was used to create and statistically analyse graphs except where otherwise stated.

GATHER [68] was used for gene ontology (GO) profiles.

Limma, Linear Models for Microarray Data [69] was used to analyse the RNA sequencing data.

Papers, version 1.0.3. [70] was used to manage bibliographies, citations and references.

R version 2.14.2 [71] was used to run Limma for RNA sequencing analysis.

2.1.4 CELL LINES

A MCF-10A WT cell line (CRL-10317) and MCF-10A CDH1-/- cell line (CLLS1042) were purchased from Sigma-Aldrich. The MCF-10A CDH1-/- cells were formed using site specific deletion of four base pairs by zinc finger nucleases, knocking out CDH1 on chromosome 16q22.1.

The MCF-7 cell line (HTB-22) was purchased from the American Type Culture Collection (ATCC).

2.2 CELL CULTURE

2.2.1 COMPLETE GROWTH MEDIA

The complete growth medium used for the culturing of MCF-10A cells was a 1:1 mixture of Dulbecco’s modified Eagle’s medium and F12 medium (DMEM-F12), supplemented with 5% horse serum, hydrocortisone (0.5 μg/ml), cholera toxin (100 ng/ml), insulin (10 μg/ml), EGF (20 ng/ml), and penicillin-streptomycin (100 μg/ml each) (Table 2.1). Pre-mixed complete media was filter sterilised using a 0.22 μm polyethersulfone filter before use. Complete growth media was then aliquoted out into 50 mL centrifuge tubes.
### TABLE 2.1. Medium Recipe for MCF-10A WT and MCF-10A CDH1 -/-

<table>
<thead>
<tr>
<th>Component</th>
<th>Complete Growth Medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMEM/F12 (Invitrogen No. 11965-118)</td>
<td>500 mL</td>
</tr>
<tr>
<td>Horse Serum (Invitrogen No. 16050-122)</td>
<td>25 mL (5% final)</td>
</tr>
<tr>
<td>Actrapid Penfil Neutral Insulin&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.4 mL (10 μg/mL)</td>
</tr>
<tr>
<td>Penicillin/Streptomycin (Invitrogen No. 15070-063)</td>
<td>5.0 mL (100 μg/ml)</td>
</tr>
<tr>
<td>Hydrocortisone&lt;sup&gt;b&lt;/sup&gt;</td>
<td>250 μL (0.5 μg/ml)</td>
</tr>
<tr>
<td>EGF&lt;sup&gt;c&lt;/sup&gt;</td>
<td>100 μL (20 ng/ml)</td>
</tr>
<tr>
<td>Cholera Toxin&lt;sup&gt;d&lt;/sup&gt;</td>
<td>50 μL (100 ng/ml)</td>
</tr>
</tbody>
</table>

<sup>a</sup>=Actrapid Penfil Neutral Insulin- Stock solutions (100 IU/mL stock) were prepared in acidified-water (100 μL). This was then filter sterilized and aliquots stored at -20°C.

<sup>b</sup>=Hydrocortisone (1 mg/mL stock): 50 mg hydrocortisone powder was weighed, dissolved in 25 mL 100% analytical grade Ethanol, then 25 mL ultra-pure distilled water (Invitrogen No. 10977-023) was added. Aliquots were then stored at -20°C.

<sup>c</sup>=EGF (100 μg/mL stock): Resuspend at 100 μg/ml in sterile distilled water, aliquots were then stored at -20°C.

<sup>d</sup>=Cholera Toxin- Resuspended in ultra-pure distilled water (Invitrogen No. 10977-023) and aliquots stored at 4°C.

The MCF-7 cell line was grown in Eagle’s minimum essential medium (RPMI 1640) (ATCC No. 30-2003), supplemented with 10% FBS.
2.2.2 CELL COUNTING

Pelleted cells were resuspended in 5 - 10 mL media. The cell suspension was mixed with Trypan blue at a 1:1 ratio and counted manually and using an automated cell counter. Cell suspension was incubated in Trypan blue for around 2 minutes. As Trypan blue (TPB) is negatively charged, it is excluded by the plasma membrane potential in living cells, but is taken up in dead cells, thereby staining the cells blue.

For manual counting, cells were counted using a haemocytometer under 10X objective. Any unusual clumping was noted; if the clumps could not be dispersed, they were counted as single cells. Live and dead cells were counted separately and averaged in order to calculate percentage cell viability. To calculate the total amount of cells, mean number of cells $\times 10^4$ was used in order to get cells per mL. This number was doubled to account for the 1:1 dilution with Trypan blue.

2.2.3 THAWING CELL LINES

Cells removed from liquid nitrogen storage were thawed quickly in a 37°C water bath. Recovered cells are then resuspended in 9.0 mL warm growth medium, centrifuged for five minutes at 100 x g (rcf), and the supernatant was discarded to remove any remaining DMSO. The pelleted cells were resuspended in warm growth medium and transferred to a tissue culture flask. Cells were inspected using an inverted microscope for morphology, then put in an incubator at 37°C and 5% CO$_2$. The media was changed the following day [64].
2.2.4 CRYOGENIC PRESERVATION

After trypsinization, washed pelleted cells were resuspended in ice-cold cell culture freezing media (Appendix A). The cells in freezing media were then aliquoted into cryovials (1.0 ml each) and transferred to a 5100 Cryo 1°C Freezing Container in a -80°C freezer to allow 1°C decrease per minute before being transferred into liquid nitrogen for long term storage within two days.

2.2.5 CELL CULTURE MAINTENANCE

MCF-10A WT, MCF-10A CDH1-/-, and MCF-7 cells were grown at 37°C, with 5% CO₂ in respective complete growth media (Chapter 2.2.1). Cells were passaged approximately every three days [64] when they reached 90% confluency and were still in the exponential log growth phase. At this confluency per T75 flask the MCF-10A WT cells numbered at 6 x 10⁶ cells while the MCF-10A CDH1-/- and MCF-7 cells cell had 1 x 10⁷ cells per T75 flask. It is not recommended to exceed nine passages post thaw for MCF-10A WT and MCF-10A CDH1-/- or 30 passages for the MCF-7 cell line.

Before passaging, used media was aspirated and cells washed with 10 mL warm phosphate-buffered saline (PBS) which was then aspirated and 2.0 mL Trypsin (0.05% Trypsin: 0.53 mM EDTA) added. The cells were then returned into the incubator at 37°C and 5% CO₂ for approximately 25 minutes (MCF-7 for 5 minutes). Afterwards 3.0 mL of complete media (Table 2.1) was added and the cells resuspended. Cells were transferred to a 15 mL centrifuge tube and the plate rinsed with a further 2.0 mL of resuspension medium in order to procure remaining cells. Flasks were then checked by microscopy to ensure that all of the cells had been transferred. Following this, the cells were pelleted at 100 x g (rcf) for 3 minutes. The media was then aspirated and the cells resuspended in 5 or 10 mL warm complete medium. Cells were plated as per plating densities detailed in Table 2.2 and checked daily.
TABLE 2.2. Plating densities for T25 and T75 Flasks.

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Cells/ T25 Flask</th>
<th>Cells/ T75 Flask</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCF-10A WT</td>
<td>1.0 x10^5</td>
<td>3.0 x10^5</td>
</tr>
<tr>
<td>MCF-10A CDH1-/−</td>
<td>1.5 x10^5</td>
<td>4.5 x10^5</td>
</tr>
<tr>
<td>MCF-7</td>
<td>2.0 x10^5</td>
<td>5.0 x10^5</td>
</tr>
</tbody>
</table>

2.3 WESTERN BLOTTING

For Western blot analysis, cells were seeded in a 6 well plate; MCF-10A was seeded at 2.0 x10^5 cells/well and MCF-10A CDH1-/− was seeded at 3.5 x10^5 cells/well. These densities were chosen in order for cells to reach equal confluency after 48 hours compensating for their different proliferation rates.

Cell harvesting began at 48 hours post seeding. Media was aspirated and each well was then washed twice with 1.0 mL of ice-cold PBS. 200 μL of cell culture protein lysis buffer (Promega) was added to each well and cells were scraped using falcon cell scrapers and transferred to a microcentrifuge tube. Samples were then vortexed, centrifuged and the supernatant transferred into a second microcentrifuge tube and stored at -20°C until required.

2.3.1 DETERMINATION OF PROTEIN CONCENTRATION

Frozen cell lysate samples were thawed on ice then vortexed briefly and heated to 95°C for 10 minutes. Protein concentration was determined using the Qubit 2.0 fluorometer. The Qubit protein assay allows for simple and accurate protein quantitation using three standards which were prepared, along with the working solution, according to the manufacturer’s protocol. Only thin-wall, clear 0.5 mL optical-grade PCR tubes were used. 10 μL of each cell lysate sample was added to
190 μL of working solution, the tubes were then vortexed and incubated for 15 minutes at room temperature (RT). Samples were then read, and stock concentrations calculated using the Qubit.

---

### 2.3.2 SDS-PAGE

10% SDS polyacrylamide gels were cast using the Bio-Rad Mini-Protean II system. Stacking and resolving gels were made (Appendix A) based on a modification of the time-honoured method of Laemmli, 1970 [72]. For SDS PAGE, 15 μg of protein from each sample was used. 37.5 μL of diluted sample, 10 μL protein loading buffer (see appendix) and 2.5 μL 2-mercaptoethanol were mixed to make a total volume of 50 μL. Samples were then vortexed and heated to 95°C for 10 minutes. 40 μL of each sample and 7 μL of prestained protein ladder (not heat denatured) was then loaded.

Electrophoresis was performed in running buffer (Appendix A) at 100 V for 1.5 hours (or until the bromophenol blue dye front had reached the base of the gel). After separation by SDS-PAGE, the proteins were transferred onto polyvinylidene difluoride (PVDF) using the Bio-Rad mini-TransBlot system for one hour at 100 V.

---

### 2.3.3 WESTERN ANALYSIS

After transfer, the PVDF membrane was blocked in blocking buffer (Appendix A) for one hour at RT. The membrane was then washed for 10 minutes in 40 mL of PBS 0.2% Tween (PBS-T), this was repeated three times. The membrane was then sealed in a plastic bag with primary antibody (Table 2.3) and rocked either overnight at 4°C or two hours at RT. The membrane was then washed three times in PBS-T. Then the secondary antibody (Table 2.3) was added to the blot and rocked for two further hours at RT. The membrane was again washed three times in PBS-T and the blot was subjected to ECLplus chemiluminescence analysis to visualise and quantitate the protein of
interest. Imaging was performed on the LAS-3000 imaging system. A chemiluminescence increment standard was used on 10 s exposure. The PVDF membrane was then washed and re-probed with the secondary antibody B-actin (Table 2.3) following the protocol described above.

**TABLE 2.3. Antibodies for immunohistochemistry.**

<table>
<thead>
<tr>
<th>Primary Antibody</th>
<th>Dilution in 5% Milk in PBS-T</th>
<th>Cat. Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>anti-E-cad (rabbit)</td>
<td>1 : 200</td>
<td>SC-7870</td>
</tr>
<tr>
<td>Secondary Antibody</td>
<td></td>
<td></td>
</tr>
<tr>
<td>anti-rabbit HRP</td>
<td>1 : 5,000</td>
<td>SC-2004</td>
</tr>
</tbody>
</table>

### 2.4 CELL DENSITY ASSAY

MCF-10A WT (2.0 x 10^5 cells/well) and MCF-10A CDH1−/− (3.5 x 10^5 cells/well) were seeded in triplicate on three 6-well plates (3 mL per well) in order to be counted at 24, 48 and 72 hours. Adherent cells were trypsinized before counting at 0, 24, 48 and 72 hours.

Doubling times were calculated using counted cell numbers from time 0, 24, 48 and 72 hours. An online calculator was used [73] and compared using the formula,

\[
D = \frac{t \times \log 2}{\log N_t - \log N_0}
\]

- D= Doubling time,
- t= time period (hours),
- Nt= number of cells at time t,
- No= initial number of cells.
2.5 NUCLEOLI COUNTING ANALYSIS

3.0 x 10⁴ MCF-10A WT cells and 4.5 x 10⁴ MCF-10A CDH1-/- cells were added to each well of a Hi-Q² dish made up to a total volume of 500 μL with warm complete media. Cells were then grown for 24 hours until they reached 90% confluency. At 24 hours a final concentration of 1.0 μg/mL Hoechst dye was added and the cells were incubated at 37°C for 15 minutes, to enable the fluorescent staining of the nucleus in living cells. The culture was then washed with growth media twice to remove excess Hoechst dye and further incubated for 45 minutes. Live cell imaging was then performed using the Nikon BioStation.

Thirteen images of adjacent fields within the dish were taken for each cell type (Six of MCF-10A WT and seven of MCF-10A CDH1-/-). Images were taken using phase contrast and DAPI channels at 20X magnification from the BioStation. The images were then randomised using the RAND function in excel followed by a double-blinded analysis which involved counting the nucleoli number per nucleus in MCF-10A WT and MCF-10A CDH1-/- cells. Nucleoli numbers were counted for each image taken under the multichannel view of phase contrast and DAPI channels to determine the distribution of nucleoli within each cell.

2.6 GENOME-WIDE RNA SEQUENCE EXPRESSION ANALYSIS

2.0 x10⁵ MCF-10A WT and 3.5 x10⁵ MCF-10A CDH1-/- cells were plated in duplicate on a 6-well plate (3 mL per well) until 70% confluency was reached (mid-log phase in growth curve). A media change was performed at 24 hours and the RNA was extracted 48 hours post seeding.

The RNA was harvested using Zymo quick-RNA MiniPrep Kit. Cells were removed from the 6-well plate using a cell scraper. The RNA extraction was carried out following the manufacturer’s
protocol with minor modifications: RNase-free water was pre-warmed before use and the stand
time for the column was increased from one to five minutes.

---

### 2.6.1 RNA CLEAN AND CONCENTRATOR

The RNA samples were DNase treated after RNA extraction to remove any genomic DNA. Samples were diluted such that 10 μg of RNA per sample was DNase treated to give an appropriate yield. DNase treatment was performed at RT for 15 minutes, then 1 μL of 25 mM EDTA was added and incubated for 10 minutes at 65°C. In a modification to the manufacturer’s protocol for elution, the column was allowed to stand at RT for five minutes rather than one minute.

---

### 2.6.2 RNA QUANTIFICATION

#### 2.6.2.1 NANOdrop

The Nanodrop is advantageous as only 1 μL of sample is required. It is able to quickly inform the user the concentration of RNA in your sample, or if there are any major contaminants present.

The Nanodrop was used following the manufacturer’s protocol, 2 μL of sample was loaded onto the lower measurement pedestal. A spectral measurement was then initiated using the accompanying operating software.

#### 2.6.2.2 Qubit RNA Assay Kit

The Qubit RNA assay allows for accurate and easy RNA quantitation. It is highly selective for RNA over double-stranded DNA. The Qubit RNA assay uses only two standards, they were prepared along with the working solution according to the manufacturer’s protocol. Only thin-wall, clear 0.5 mL optical-grade PCR tubes were used. 2 μL of each sample was added to 198 μL of working solution, the tubes were then vortexed and incubated for 2 minutes at RT. Samples were then read, and stock concentrations calculated using the Qubit 2.0 Fluorometer.
2.6.2.3 AGILENT RNA 6000 BIOANALYSER

The Agilent RNA 6000 Bioanalyser gives a much more reliable measure of the quality of RNA. The RNA quantification was performed according to the manufacturer’s protocol. RNA samples were diluted to 200 ng/μL and heated to 95°C for 5 minutes before loading 1 μL onto the Agilent RNA 6000 Pico chip using the chip priming station. Chips were then vortexed for one minute at 1000 x g (rcf) using the IKA vortex mixer and read using the Agilent 2100 Bioanalyser.

2.6.3 RNA TRANSCRIPTOME SEQUENCING

Samples for RNA transcriptome sequencing (See appendix) were sent to New Zealand Genomics Limited (NZGL). The library was prepared by NZGL using Illumina TruSeq RNA preparation version 2. One lane of the HiSeq 2000 was used and the libraries were loaded at a yield of 8 pM. The libraries were then sequenced using 208 cycles of chemistry and imaging.

2.6.4 BIOINFORMATICS ANALYSIS

R version 2.14.2 [71] was used to run Limma [69] for RNA sequencing analysis. Mapped sequences were annotated using biomaRt using the Ensemble dataset "hsapiens_gene_ensembl" [74]. Un-annotated genes were removed, and remaining count data were normalised using edgeR [65]. The Benjamini & Hochberg (BH) [75] adjusted P value for FDR control of 0.05 or 5.00E-02 was considered significant. Bioinformatic analysis including mapping and normalisation of counts were performed by NZGL.
2.7 CELL WOUNDING ASSAY

The cell wounding method was based on the classical cell wounding assay by Dupre-Crochet et al, [76]. 3.0 x 10^4 MCF-10A WT cells and 4.5 x 10^4 MCF-10A CDH1−/− cells were seeded into each well of a Hi-Q4 dish to a final volume of 500 μL with warm complete media. Cells were then incubated for 24 hours to allow adherence before a scratch was made using an autoclaved 200 μL gel loading tip on the base of the plate. The media was then aspirated and cells gently washed with 500 μL warm PBS before 500 μL of warm fresh complete media was added to each well. A final concentration of 1.0 μg/mL Hoechst dye was added and the cells were incubated at 37°C for 15 minutes, to enable the fluorescent staining of the nucleus in living cells to facilitate cell counting. The culture was then washed with growth media twice to remove excess Hoechst dye and incubated for a further 45 minutes. Cells were incubated in the Nikon BioStation for 48 hours at 37°C with 5% CO₂. Images were taken every 30 minutes for 48 hours using the BioStation IM software.

There are many approaches to quantify the migration of cells across a wound; some papers used custom-designed computer programs to determine the area remaining after wounding [77], while others used the distance between wound edges at each time point [78]. Another method would be to count the number of cells that crossed the wound ‘front’ over 48 hours. The method chosen in this study was a difference in acellular area. This was due to the fact both wound edges were not able to be visualised on the same image using the BioStation. Additionally, there was no access to software to enable trouble-free counting of the number of cells that cross the wound ‘front’.
Image J was used to analyse the images produced by the BioStation. Cell counts were analysed using the ‘analyse particles’ function and the ‘measure’ function was used in the quantification of acellular area.

### 2.8 SOFT AGAR ASSAY

The idea of studying the effectiveness of an agent on the survival and proliferation of cells in soft agar was first developed by Hamburger and Salmon in 1977 [79]. Cancerous cells are able to form colonies when grown in agar, while non-cancerous cell are not. In the current study, the loss of E-cadherin in mammary epithelial cells was assessed in soft agar to elucidate the ability of cells to grow under anchorage independent conditions.

A base layer of 0.5% agar was prepared in a 6-well dish by diluting sterile agar in the appropriate cell culture medium. 1.5 mL of 0.5% agar solution was dispensed per well and allowed to set on a level surface over two hours in the TC hood. A soft agar layer of 0.35 % agar in medium was made as follows: Cells were trypsinized, counted and diluted to appropriate concentrations. A 1.5 mL mixture of cells, medium and 0.35% agar was carefully overlayed on the base agar layer so as to not disrupt the base layer. For MCF-10A WT and MCF-10A CDH1−/− 2 x 10^4 cells per well was chosen and for MCF-7, 2 x 10^3 cells per well were used. The soft agar was then gently swirled in the 6-well plate in order to evenly distribute the cells. The 0.35% agar was then allowed to cool and set on a level surface in a sterile environment, before moving into the incubator at 5% CO₂ and 37°C. Complete growth media was added to the surface the following day. This was removed once a week and was replenished twice a week.
2.8.1 STAINING SOFT AGAR

Growth medium was removed and each well was overlayed with 1.0 mL of MTT solution in PBS (2 mg/mL). The wells were then allowed to incubate at RT, with gentle shaking, for 4 hours. Then MTT solution was removed and 2 mL water was added every 15 minutes with agitation and repeated until the background was clear. Images were taken by scanning with the Image Scanner 3.

2.9 THREE-DIMENSIONAL CULTURE OF MCF-10A CELLS ON RECONSTITUTED BASEMENT MEMBRANE (OVERLAY METHOD)

The MCF-10A cell line is able to form polarized and glandular structures in vitro on reconstituted basement membrane derived from Engelbreth-Holm-Swarm mouse tumours (Matrigel) that recapitulate the glandular structure of the epithelium in vivo. It forms polarized growth-arrested acini-like spheroids with hollow lumen, apicobasal polarization of cells and the basal deposition of basement membrane components such as laminin 5 [57,64]. It is therefore an ideal system to study how the regulation of cell differentiation influences the structure and function of glandular epithelium both in early tumour formation and the normal state [64].

In order to form a basement membrane, 50 μL of 100% growth-factor reduced Matrigel was dispensed into an a 8 well chamber coverglass and allowed to set for 15 minutes at 37°C. Cells were prepared for 3D culture by growing to 80% confluency in monolayer. The MCF-10A WT and MCF-10A CDH1−/− cells were resuspended and added to Matrigel and media to achieve a final concentration of 5,000 cells/well in 2% Matrigel. 400 μL of this mixture of cells, media and 2% Matrigel was then overlayed over the basement membrane of 100% Matrigel (Figure 2.1). Once plated, all cultures were incubated at 37°C and 5% CO2 for up to 31 days. Fresh assay media containing 2% Matrigel and 5 ng/ml EGF was added every four days. The 3D cultures were imaged using the BioStation every four days.
Figure 2.1. Schematic of overlay method for 3D culture of MCF-10A cells on Matrigel. Adapted from Debnath et al, [64].
3 RESULTS

The primary aim of my research was to establish if $CDH1^{-/-}$ alone is sufficient to induce an EMT in mammary epithelial cells as a model to determine the role of E-cadherin loss in the disease progression of early lesions in the stomachs of HDGC individuals. The method of approach can be briefly summarised in three steps; characterisation of the gene expression profiles of the MCF-10A cell line with and without $CDH1$ expression, the effect of $CDH1^{-/-}$ on migration and on cell adherence, and a comparison of the differentiation capability of the two cell lines grown in 3D culture.

3.1 CHARACTERISATION OF IN VITRO PHENOTYPES RESULTING FROM $CDH1^{-/-}$

Initial characterisation of the MCF-10A WT and MCF-10A $CDH1^{-/-}$, herein referred to as $CDH1^{-/-}$, cell lines has been undertaken in this paper (Ours is the first group to purchase these isogenic cell lines from Sigma-Aldrich). This was performed before the main aims were addressed in order to have a holistic understanding of the cell lines. This involved verifying E-cadherin expression by Western blot analysis, assessing cell morphology in both two- and three- dimension cultures, doubling time analysis and a characterisation of any major morphological differences between the cell lines.

3.1.1 VERIFICATION OF E-CADHERIN LOSS

In order to confirm the knockout (KO) of E-cadherin in the $CDH1^{-/-}$ cell line a Western blot (Chapter 2.3) was carried out using a specific primary Ab against E-cadherin (SC-7870) and a secondary Ab conjugated to a peroxidase enzyme (when peroxidase is exposed to chemiluminescent substrates such as ECL, it is oxidized and produces light which is detected by the LAS-3000 imaging system [80]).
Western blot analysis showed a 135 kDa E-cadherin protein in the MCF-10A WT cell lysate which was not observed in the CDH1-/– cell line (3.1A), confirming the absence of E-cadherin in the CDH1-/– cell line. There were some other minor low molecular weight bands also observed in the MCF-10A WT cells. These bands could be due to truncated or degraded E-cadherin. β-actin was used as a loading control to show even loading across the wells (Figure 3.1B).

Figure 3.1. Western blot showing E-cadherin is not present in MCF-10A CDH1-/–. (A) The molecular weight of E-cadherin is 135 kDa, (B) the molecular weight of β-actin is 40 kDa. The Western blot was undertaken as described in the methods section. The chemiluminescent exposure time was 40 s and EPI 1/30 s for the marker band.
3.1.2 CELLULAR MORPHOLOGY

The loss of E-cadherin caused visible changes in the appearance of the MCF-10A cells grown in two and three dimensional cultures.

3.1.2.1 TWO-DIMENSIONAL CULTURE

The MCF-10A WT cell line showed a cobblestone-like morphology and dense cell-cell contacts (Figure 3.2A). This was slightly lost in the CDH1/- cell line (Figure 3.2B) where cells had fewer cell-cell contacts compared to the parent cell line. More floating, non-adherent cells were also observed in the CDH1/- line.

![Figure 3.2](image)

Figure 3.2. Differential growth morphology between MCF-10A WT and MCF-10A CDH1/- 95%-100% confluent monolayers of cells after 72 hours, viewed at 10X magnification with phase contrast. (A) MCF-10A WT cells. (B) MCF-10A CDH1/- cells. The scale bar is 400 μm in length.

3.1.2.2 THREE-DIMENSIONAL CULTURE

In three-dimensional cultures, MCF-10A WT cells form distinctly polarised and ordered acini at Day 3 (Figure 3.3A). In contrast to this, CDH1/- cells formed less well ordered and less polarised acini (Figure 3.3B). The sizes of the acini were variable within each cell line; hence no observation was drawn on different acini sizes. Size differences could be attributed to cell seeding differences at
Day 0, or the clumping of cells. At Day 12 in 3D culture the differences in morphology became more apparent (Figure 3.15).

Figure 3.3. Three-dimensional culture of MCF-10A WT and MCF-10A CDH1-/ cells. Acini at Day 3, viewed under 20X magnification with phase contrast on the BioStation. (A) MCF-10A WT cells and (B) MCF-10A CDH1-/ cells were seeded at 2.5 x 10⁴ cells/mL and 3.75 x 10⁴ cells/mL, respectively on day zero in a Matrigel 3D-culture as described in the methods section. Spheroid sizes varied greatly within each cell type. The scale bar is 10 μm in length.

3.1.3 PROLIFERATION RATES OF MCF-10A WT AND MCF-10A CDH1-/

In view of the different growth morphology, the rate of proliferation was investigated. This was achieved by calculating doubling times from cell density assays, based on what was understood from preliminary experiments.
3.1.3.1 CELL DENSITY ASSAY

Cell density assays (Chapter 2.4) were used for two purposes: calculating doubling times and working out seeding densities. Doubling times were used to better characterise the two cell lines. Seeding densities were essential in order to have an equivalent amount of RNA extracted from the two cell lines in RNA expression analysis (Chapter 2.6). It was also important to calculate an appropriate seeding density when passaging the cell lines in T25 and T75 flask setups (Table 2.2).

The total number of cells/well at 0, 24 and 48 hours post-seeding were determined (Figure 3.4). In a preliminary experiment the two cell lines were seeded at different densities because of the different observed growth morphology. Cell numbers were counted at 48 hours post seeding to find the best seeding density for equivalent numbers of cells/well at 48 hours. Hence at 0 hours, cells were seeded at $2.0 \times 10^5$ cells/well for MCF-10A WT and $3.5 \times 10^5$ cells/well for CDH1-/-(Figure 3.4).

![Figure 3.4. Cell density assay of the two isogenic cell lines over 48 hours. MCF-10A WT and MCF-10A CDH1-/ cells were plated in a 6-well format. Cell numbers from each cell line were counted at least twice using an automated cell counter and averaged. Data expressed as mean +/- SEM.](image-url)
At 48 hours post seeding the cell numbers had 98% similarity; there were $9.07 \times 10^5$ cells/well and $9.28 \times 10^5$ cells/well for MCF-10A WT and CDH1-/-, respectively. The growth of MCF-10A WT cells in the first 24 hours was faster compared to the CDH1-/- cells (Figure 3.4). A lag to log phase could be seen for both cell lines between 0 and 24 hours, however CDH1-/- cells had a slower proliferation.

### 3.1.3.2 DOUBLING TIMES

Doubling times for each cell line can be determined to compare proliferation rates between the two cell lines. Doubling times are used to characterise cell lines based on their ability to divide. They can be calculated from two time points using a start and end time point formula, or from multiple time points using an exponential curve fitting method (Chapter 2.4). In this paper, three time points with an exponential curve fitting method were used to calculate doubling times. This was due to the lag to log phase observed in the cell lines, and hence multiple time points meant a more reliable representation of events.

MCF-10A WT cells showed a 1.5 fold higher proliferation rate, with a doubling time of 20.3 hours compared to 31.5 hours for CDH1-/- cells (Figure 3.5).
Figure 3.5. Exponential curve regression for (A) MCF-10A WT and (B) MCF-10A CDH1-/--. Total cells per well were taken from Figure 3.4.

3.1.4 NUCLEOLI NUMBERS

Visual inspection of the morphology of two cell lines suggested that the number of nucleoli per cell may differ between the cell lines. This observation was of interest because high nucleoli numbers have been observed in cancer tissues such as prostate cancer [81]. Nucleoli cellular functions include ribosomal RNA (rRNA) transcription, processing, modification and assembly within the nucleolus [82]. Increased nucleoli number is due to an increased rate of proliferation which is linked to an increased demand for ribosome biogenesis [83]. I sought to see if this was observed in CDH1-/- cells.
After randomised, double blinded analysis, a significant difference in the number of nucleoli in MCF-10A WT and CDH1-/- cells was confirmed (Figure 3.6). MCF-10A WT cells had 12.9% of cells with one nucleolus compared to 43.6% in CDH1-/- cells (P= 0.0012), just over a 3 fold difference. Due to the cut off value of P <0.05, there was no significant difference for the percent of cells with two nucleoli per cell, with 30.4% and 40.3% for MCF-10A WT and CDH1-/-, respectively (p=0.051). However there is still a substantial difference for two-nucleoli per cell between the two cell lines. MCF-10A WT cells had 40.1% of cells with three nucleoli per cell compared to 14.2% in CDH1-/- cells (P= 0.0012). At four nucleoli per cell, MCF-10A WT cells had 15.3% of cells with this phenotype compared to 2.2% in the CDH1-/- cell line (P= 0.0034). None of the CDH1-/- cells had five nucleoli per cell, whereas 1.3% of MCF-10A WT cells had five nucleoli per cell.

Figure 3.6. Randomised, double blinded analysis of nucleoli number in MCF-10A WT and MCF-10A CDH1-/-.

Images taken of both cell lines in the BioStation were used to determine nucleoli numbers. 13 images were taken in total, six for MCF-10A WT and seven for MCF-10A CDH1-/-. A Wilcoxon-Mann-Whitney T-Test was performed to compare each cell line.
3.2 GENE EXPRESSION PROFILING OF MCF-10A WT AND MCF-10A CDH1-/−

The key characterisation step of these two isogenic cell lines, MCF-10A and CDH1-/− involved studying their gene expression. Gene expression profiling allowed a global picture of the changes in cellular function that occur upon E-cadherin loss. It was used to indicate the genes that were upregulated or downregulated between the two isogenic cell lines. Cells were grown in a 6-well plate format over 48 hours at which time RNA was extracted from the two cell lines (Chapter 2.6).

3.2.1 QUANTITATION OF RNA IN SAMPLES

RNA quality had to be assessed as it was important to have high quality of intact RNA for sequencing. To ensure this, the RNA extraction was done in two biological replicates, and the extraction with the highest total RNA and best quality RNA was chosen for RNA sequencing analysis. RNA quantitation was carried out using three methods, Nanodrop, Qubit and Agilent RNA 6000 Bioanalyser (Bioanalyser) (Chapter 2.6.2).

All thee RNA quantification systems gave results in reasonable accordance (Figure 3.7). The average of these three systems was used to quantitate the RNA in samples. With the exception of a variation in RNA quantification of RNA extracted from experiment II (Figure 7B), where total RNA values of MCF-10A CDH1-/−(1) were 26.2 μg and 13.6 μg respectively as measured from Qubit and Bioanalyser systems. A T-test could not be undertaken to compare the significance between these three methods because two or more values are needed for each measurement and each technical replicate was measured only once.
Figure 3.7. Nanodrop, Qubit and Bioanalyser assays were used to quantitate total RNA. (A) RNA extraction Experiment I and (B) Experiment II are biological replicates. Each experiment had two technical replicates shown by (1) and (2). The concentration and volumes of each sample are shown in Appendix B.
3.2.2 QUALITY OF RNA IN SAMPLES

The Agilent RNA 6000 Bioanalyser was also used to assess RNA quality. The Bioanalyser system provides a reliable score of the RNA quality in each sample. This score or RNA integrity number (RIN) is calculated from an algorithm involving the rRNA ratio of 28S and 18S peaks, and the presence or absence of degradation products. The RIN score is largely independent of the amount of RNA used, or the origin of the sample.

RNA extracted for experiment I was aliquoted into two vials in which one vial was subjected to DNase1 treatment. The DNase1 treated samples (Figure 3.8A) however had a multitude of fragments with different lengths and consequently different migration speeds through the liquid-gel matrix of the Bioanalyser. This is reflected in a poor-quality electropherogram with obscured peaks and the peak for 28S rRNA being diminished (Figure 3.8A). The electropherogram for the untreated samples had two clear peaks for 18S and 28S rRNA (Figure 3.8B). The untreated samples had the highest quality RNA with an average RIN value of 9.35, compared to 6.35 for ‘DNase treated’ MCF-10A WT. The DNase1 clean and concentrate kit either degraded the samples or contaminated them. However, the lack of a peak on the far right of the electropherogram for any of the samples is indicative of no genomic DNA, suggesting that DNase1 treatment was not required.
Figure 3.8. DNaseI treated samples have insufficient RNA quality. Samples from RNA extracted from Experiment I were analysed by Agilent RNA 6000 Bioanalyser as described in the methods section. (A) DNaseI treated MCF-10A WT sample. (B) Untreated MCF-10A WT sample. The marker peak is at 25 nt. 18S rRNA and 28S rRNA are shown in the figure as the first and second primary peaks from left to right, respectively. Fluorescence units absorbed= FU.

This was confirmed when RNAs from both cell lines and experiments were again analysed without prior DNaseI treatment (Figure 3.9). They were textbook examples of quality electropherogram results with a flat base line that is peppered with narrow peaks and distinct 18S and 28S rRNA peaks. The RIN values for the various samples were; MCF-10A WT (1) = 9.80, MCF-10A WT (2) = 9.80, MCF-10A CDH1-/ (1) = 9.75, and MCF-10A CDH1-/ (2) = 9.70.
Figure 3.9. RNA extraction Experiments One and Two have similar RNA quality. Agilent RNA 6000 Bioanalyser analysis was performed as described in the methods section. (A) MCF-10A WT. (B) MCF-10A CDH1-/- (1) and (2) are RNA extraction Experiments I and II. Each electropherogram is a representative graph of two technical replicates. The marker peak is at 25 nt. 18S rRNA and 28S rRNA are also shown in the figure as the first and second primary peaks from left to right, respectively. Fluorescence units absorbed= FU.

Experiment I was chosen to be sent off to NZGL at the University of Otago for RNA sequencing analysis due to the high quality of RNA and a higher amount of total RNA in the samples. There was 1.9 fold higher total RNA in experiment I compared to experiment II for MCF-10A WT (1), and 1.2 fold higher total RNA for MCF-10A CDH1-/- (1) (Figure 3.7A-B).
3.2.3 TRANSCRIPTOME SEQUENCING (NZGL)

Quantitated RNA extracted from both isogenic cell lines were sent for RNA sequencing to elucidate the changes associated with E-cadherin loss, with an emphasis on the genes involved in the EMT (increase in mesenchymal markers, decrease in epithelial markers, and identification of EMT inducers).

For robust deduction regarding biological difference, RNA sequencing was performed in two technical replicates for each cell line. The quality and quantity of the cDNA library generated from the RNA sent for RNA sequencing was independently verified by NZGL (Table 3.1). The libraries for RNA sequencing were pooled and prepared using the TruSeq™ RNA Sample Preparation v2.

TABLE 3.1. Quality control on library preparations

<table>
<thead>
<tr>
<th>Sample</th>
<th>Internal quant (ng/μL)</th>
<th>Input</th>
<th>Adapter index</th>
<th>Qubit lib quant (ng/μL)</th>
<th>BA average size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCF-10A WT_A</td>
<td>556</td>
<td>1</td>
<td>A002</td>
<td>27.8</td>
<td>339</td>
</tr>
<tr>
<td>MCF-10A WT_B</td>
<td>426</td>
<td>1</td>
<td>A004</td>
<td>17.4</td>
<td>332</td>
</tr>
<tr>
<td>MCF-10A CDH1-/-_A</td>
<td>814</td>
<td>1</td>
<td>A07</td>
<td>15.1</td>
<td>325</td>
</tr>
<tr>
<td>MCF-10A CDH1-/-_B</td>
<td>850</td>
<td>1</td>
<td>A012</td>
<td>14.1</td>
<td>320</td>
</tr>
</tbody>
</table>

The normalised expression of each gene from the isogenic cell lines were then subjected to edgeR analysis using R (Chapter 2.6.4). We sequenced four cDNA libraries (Table 3.2), two MCF-10A WT and two CDH1-/. One of the MCF-10A WT cDNA libraries generated 8 x 10^8 sequence reads (Table 3.2), with an average for all four libraries of 2.5 x 10^8 sequence reads each of which was 100 base pairs in length. These four libraries in total produced 40.2 Gb of determined sequence after demultiplexing (i.e. the sorting of indexed reads). Q30 indicates an estimated base call accuracy of 1/1000 or 99.9%. 79-80% of determined sequence was >Q30 resulting in a yield >Q30 of 31.8 Gb.
(minimum expected 27 Gbs per lane). The sequence reads were aligned to human reference genome database using biomaRt in R version 2.14.2 [71].

TABLE 3.2. Demultiplex statistics:

<table>
<thead>
<tr>
<th>Lane</th>
<th>Sample</th>
<th>Index</th>
<th>Yield (Mbases)</th>
<th># Reads</th>
<th>% &gt;= Q30</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>MCF-10A WT_A</td>
<td>TGACCA</td>
<td>8053</td>
<td>7.97E+08</td>
<td>79</td>
</tr>
<tr>
<td>6</td>
<td>MCF-10A WT_B</td>
<td>CGATGT</td>
<td>6556</td>
<td>6.49E+07</td>
<td>80</td>
</tr>
<tr>
<td>6</td>
<td>MCF-10A CDH1-/-_A</td>
<td>CTTGTA</td>
<td>6329</td>
<td>6.27E+07</td>
<td>79</td>
</tr>
<tr>
<td>6</td>
<td>MCF-10A CDH1-/-_B</td>
<td>CAGATC</td>
<td>5515</td>
<td>5.46E+07</td>
<td>79</td>
</tr>
</tbody>
</table>

### 3.3 IMPACT OF E-CADHERIN LOSS IN A MAMMARY EPITHELIAL CELL LINE

In order to characterise the two cell lines and determine what changes are occurring with E-cadherin loss, various endeavours were undertaken. First the most differentially expressed genes were investigated. These genes were then analysed via Gene ontology (GO) using GATHER [68]. Intriguing results were then further investigated by a *tour de force* trawl of the relevant literature.

#### 3.3.1 DISTRIBUTION OF DIFFERENTIALLY EXPRESSED GENES FOR MCF-10A WT AND MCF-10A CDH1-/-

A volcano plot is the probability that a gene is differentially expressed vs. log FC (fold change). It is a visual way to show the distribution of gene expression, displaying both magnitude of change and statistical significance at the same time. It also indicates where the significant cut-offs will be on the data set, how many genes are effected by the cut offs and hence allows for quick identification of genes that are significantly dysregulated. In this analysis the cut offs used were log FC +/- 1.00 (or FC +/- 2.00) and an adjusted P value of <0.05.

From the log normalised counts a Volcano plot was created in R [71] (Figure 3.10). A cut off of +/- 1.00 log fold change, and an adjusted P value of 0.05 (-log adjusted P value of 3.00) was used for
this data set. The log fold change is in reference to $CDH1^{-/-}$ and was calculated from normalised counts using edgeR. The cut-offs are symbolised by the blue lines on the volcano plot, the downregulated genes (top left green square) and upregulated genes (top right red square) contain the genes of interest. Of the 1,117 significantly differentially expressed genes from the two cell lines there were 519 genes with a negative log FC and 598 genes with a positive log FC (Figure 3.10).

A -log adjusted P value of just less than 3 equates to a 95% probability that genes are differentially expressed. The volcano plot shows that from 54,000 mapped sequences, around 1,300 lay within the cut off which correlated to 1,117 genes of interest (Figure 3.10).

Figure 3.10. 1,117 genes are significantly differentially expressed between MCF-10A $CDH1^{-/-}$ and MCF-10A WT. The volcano plot for MCF-10A $CDH1^{-/-}$ against MCF-10A WT is shown. Counts were normalized according to the limma manual as explained in the methods section. The graph was creating using R [71] as explained in the methods section.
3.3.2 MOST DIFFERENTIALLY EXPRESSED GENES

Differentially expressed genes were most commonly found on chromosome 1 (153 genes), the largest of the human chromosomes, with chromosome 19 (99 genes) being the second most common and chromosome 2 with 85 genes the third. Of the total genes on chromosome 1, genes differentially expressed on E-cadherin loss made up 4.9% [84]. Human chromosome 19 has more than twice the gene density of other chromosomes [85]. Genes altered on E-cadherin loss make up 6.8% of the genes on chromosome 19 [86].

*CASP14* located on chromosome 19 had the highest positive FC between the two cell lines of 24.1 (P=2.99E-02) (Table 3.3). *NDRG1* located on chromosome 8 had the most significant positive FC with an adjusted P value of 6.82E-07 and a FC of 5.4 (Table 3.3). *PNMAL1*, a poorly annotated gene located on chromosome 19 had the highest negative FC between the two cell lines of -415.9 (P=1.96E-03). This was due to *PNMAL1* having its expression abolished in the *CDH1-*/- cell line (hence a normalised expression of 0.03) and an average normalised expression counts of 12.6 in the MCF-10A WT cell line. *PRSS21, GPR27, B3GALNT1, PDE7B, MAGEE1, ZNF485, PCSK1N, KCNQ3, RBP4*, and *ZNF73* also had their expression abrogated in the *CDH1-*/- cell line (But had normal expression in MCF-10A WT cells). No clearly shared functions were observed with these genes. Zinc factors were abundant on the negative FC list, with four appearing in the top 30; *ZNF485, ZNF239, ZNF737* (Table 3.3), and *ZNF287* (Data not shown). These genes share a common function in regulation of transcription.

E-cadherin (*CDH1*) was 91st on the most differentially expressed list for negative FC, with a FC of -10.0 (P=2.39E-03) (Table 3.3). This reflects the deletion of the *CDH1* gene in *CDH1-*/- cells. There were still counts for *CDH1* in the *CDH1-*/- cell line, with an average of 19.5 normalised counts,
compared to 194.6 average normalised counts for the MCF-10A WT cell line. This suggested the four base pair deletion did not entirely abolish the mRNA expression of CDH1.

TABLE 3.3. Top 20 most differentially expressed genes for positive and negative FC.

<table>
<thead>
<tr>
<th>Positive FC</th>
<th>Chromosome</th>
<th>Fold Change</th>
<th>adj.P.Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CASP14</td>
<td>19</td>
<td>24.08</td>
<td>2.99E-02</td>
</tr>
<tr>
<td>FXYD1</td>
<td>19</td>
<td>21.56</td>
<td>1.19E-02</td>
</tr>
<tr>
<td>ALDH3B2</td>
<td>11</td>
<td>19.70</td>
<td>1.66E-02</td>
</tr>
<tr>
<td>SLC10A6†</td>
<td>4</td>
<td>17.63</td>
<td>3.25E-03</td>
</tr>
<tr>
<td>C6orf223</td>
<td>6</td>
<td>17.51</td>
<td>1.75E-02</td>
</tr>
<tr>
<td>SORCS3</td>
<td>10</td>
<td>16.68</td>
<td>6.34E-03</td>
</tr>
<tr>
<td>SPRR3</td>
<td>1</td>
<td>15.45</td>
<td>3.38E-02</td>
</tr>
<tr>
<td>ATP5A1P7†</td>
<td>9</td>
<td>15.45</td>
<td>4.98E-03</td>
</tr>
<tr>
<td>KCNJ12</td>
<td>17</td>
<td>15.35</td>
<td>1.03E-02</td>
</tr>
<tr>
<td>KRT34†</td>
<td>17</td>
<td>15.03</td>
<td>4.42E-03</td>
</tr>
<tr>
<td>FGF9</td>
<td>13</td>
<td>15.03</td>
<td>2.96E-02</td>
</tr>
<tr>
<td>CD53</td>
<td>1</td>
<td>14.62</td>
<td>2.70E-02</td>
</tr>
<tr>
<td>CA9</td>
<td>9</td>
<td>14.62</td>
<td>7.63E-03</td>
</tr>
<tr>
<td>SEMA5B</td>
<td>3</td>
<td>14.22</td>
<td>5.60E-03</td>
</tr>
<tr>
<td>PPP1R1C</td>
<td>2</td>
<td>13.74</td>
<td>2.70E-02</td>
</tr>
<tr>
<td>SUN3†</td>
<td>7</td>
<td>13.45</td>
<td>2.94E-02</td>
</tr>
<tr>
<td>LINC00330†</td>
<td>13</td>
<td>12.91</td>
<td>5.99E-03</td>
</tr>
<tr>
<td>ZPLD1†</td>
<td>3</td>
<td>12.73</td>
<td>1.29E-02</td>
</tr>
<tr>
<td>MYOM3</td>
<td>1</td>
<td>12.38</td>
<td>3.33E-02</td>
</tr>
<tr>
<td>TSPAN12†</td>
<td>7</td>
<td>12.21</td>
<td>3.54E-02</td>
</tr>
<tr>
<td>NDRG1**</td>
<td>8</td>
<td>5.39</td>
<td>6.82E-07</td>
</tr>
<tr>
<td>Gene</td>
<td>Chromosome</td>
<td>Fold Change</td>
<td>adj.P.Value</td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PNMAL1†</td>
<td>19</td>
<td>-415.87</td>
<td>1.96E-03</td>
</tr>
<tr>
<td>PRSS21</td>
<td>16</td>
<td>-284.05</td>
<td>2.27E-03</td>
</tr>
<tr>
<td>GPR27</td>
<td>3</td>
<td>-282.09</td>
<td>3.07E-03</td>
</tr>
<tr>
<td>B3GALNT1</td>
<td>3</td>
<td>-155.42</td>
<td>1.06E-02</td>
</tr>
<tr>
<td>PDE7B</td>
<td>6</td>
<td>-145.01</td>
<td>2.54E-03</td>
</tr>
<tr>
<td>TPTEP1†</td>
<td>22</td>
<td>-139.10</td>
<td>1.01E-02</td>
</tr>
<tr>
<td>MAGEE1</td>
<td>X</td>
<td>-116.16</td>
<td>5.60E-03</td>
</tr>
<tr>
<td>ZNF485</td>
<td>10</td>
<td>-114.56</td>
<td>3.91E-03</td>
</tr>
<tr>
<td>PCSK1N</td>
<td>X</td>
<td>-106.89</td>
<td>2.87E-03</td>
</tr>
<tr>
<td>SLFN13†</td>
<td>17</td>
<td>-81.01</td>
<td>6.75E-03</td>
</tr>
<tr>
<td>POSTN</td>
<td>13</td>
<td>-81.01</td>
<td>1.92E-02</td>
</tr>
<tr>
<td>CHDH</td>
<td>3</td>
<td>-79.89</td>
<td>3.07E-03</td>
</tr>
<tr>
<td>ZNF239</td>
<td>10</td>
<td>-78.79</td>
<td>3.07E-03</td>
</tr>
<tr>
<td>DARC</td>
<td>1</td>
<td>-76.11</td>
<td>6.96E-03</td>
</tr>
<tr>
<td>PI15</td>
<td>8</td>
<td>-69.55</td>
<td>2.44E-02</td>
</tr>
<tr>
<td>CADM3†</td>
<td>1</td>
<td>-63.12</td>
<td>2.54E-03</td>
</tr>
<tr>
<td>ATOH8</td>
<td>2</td>
<td>-61.39</td>
<td>5.60E-03</td>
</tr>
<tr>
<td>SRGN†</td>
<td>10</td>
<td>-59.30</td>
<td>2.16E-03</td>
</tr>
<tr>
<td>KCNQ3</td>
<td>8</td>
<td>-57.28</td>
<td>3.02E-03</td>
</tr>
<tr>
<td>RBP4</td>
<td>10</td>
<td>-54.57</td>
<td>3.90E-03</td>
</tr>
<tr>
<td>ZNF737</td>
<td>19</td>
<td>-54.57</td>
<td>5.26E-03</td>
</tr>
<tr>
<td>(CDH1)***</td>
<td>16</td>
<td>-10.00</td>
<td>2.39E-03</td>
</tr>
<tr>
<td>FEZ1**</td>
<td>11</td>
<td>-9.99</td>
<td>1.76E-03</td>
</tr>
<tr>
<td>NRCAM**</td>
<td>7</td>
<td>-4.53</td>
<td>1.07E-04</td>
</tr>
<tr>
<td>PF4**</td>
<td>4</td>
<td>-3.84</td>
<td>1.11E-02</td>
</tr>
<tr>
<td>SLIT2**</td>
<td>4</td>
<td>-5.03</td>
<td>6.58E-05</td>
</tr>
</tbody>
</table>

The genes shown had a cut off of FC +/- 2.00 and BH [75] adjusted P value for FDR control of 0.05 or 5.00E-02. Fold changes refer to the CDH1-/- cell line compared to the MCF-10A WT. (CDH1) was deleted by site specific deletion. † Unknown gene according to GATHER. ** Not in chronological order.
As can be expected from such analyses, a large number of diverse genes that were found to be differentially expressed were unknown according to GATHER (Table 3.3). This made analysis of genes based purely on their statistical significance and fold change problematical. As such GO analysis was performed for the top 1,117 significantly differentially expressed genes to reduce complexity and highlight enriched biological processes, cellular components and molecular functions.

3.3.3 GO ANALYSIS

Single genes are not the best indicator of cell wide mechanisms and so gene ontology (GO) analysis was carried out. GO analysis provides a set of organised terms helpful for annotating genes and sequences and is available for public use. GO analysis was split between the 598 significantly upregulated genes and 519 significantly downregulated genes (Table 3.4). A High Bayes factors signifies strong evidence supporting an annotation. The Bayes factor value therefore depends on what the information is intended for. A purely descriptive overview of an association can have a low Bayes factor, however high Bayes factors are needed to definitively prove an association. A cut off of 6 is recommended, as that cut off balances false negatives with false positives [68].

GO terms of highest representation and significance in the genes upregulated related to the CDH1-/ cell line was nucleobase-containing compound metabolic process (35 genes) with the highest Bayes factor of 20 (Table 3.4). This was followed by cell-cell signalling (43 genes) with a Bayes factor of 16. For this study, genes downregulated in the CDH1-/ cell line involving the cell adhesion term (43 genes with a Bayes factor of 17) were of particular interest. Extracellular matrix (ECM)-organisation and biogenesis (7 genes) and extracellular structure-organisation and biogenesis (7 genes) GO terms were also of interest, both of which had Bayes factors of 4 and were
hence more descriptive than solid associations (Table 3.4). Specific genes involved in GO terms can be found in Appendix E.

**TABLE 3.4. Gene Ontology analysis.**

<table>
<thead>
<tr>
<th>GO accession number</th>
<th>Gene Ontology</th>
<th># Genes Upregulated in CDH1-/-</th>
<th>Bayes Factor</th>
<th># Genes Downregulated in CDH1-/-</th>
<th>Bayes Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0006139</td>
<td>Nucleobase-containing compound metabolic process</td>
<td>35</td>
<td>20</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GO:0007267</td>
<td>Cell-cell signaling</td>
<td>43</td>
<td>16</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GO:0044237</td>
<td>Cellular metabolism</td>
<td>132</td>
<td>14</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GO:0044237</td>
<td>Metabolism</td>
<td>149</td>
<td>11</td>
<td>&lt; 2</td>
<td>-</td>
</tr>
<tr>
<td>GO:009887</td>
<td>Organogenesis</td>
<td>54</td>
<td>10</td>
<td>52</td>
<td>12</td>
</tr>
<tr>
<td>GO:0048513</td>
<td>Organ development</td>
<td>54</td>
<td>10</td>
<td>52</td>
<td>12</td>
</tr>
<tr>
<td>GO:009653</td>
<td>Morphogenesis</td>
<td>61</td>
<td>9</td>
<td>68</td>
<td>19</td>
</tr>
<tr>
<td>GO:007275</td>
<td>Development</td>
<td>83</td>
<td>8</td>
<td>89</td>
<td>18</td>
</tr>
<tr>
<td>GO:007154</td>
<td>Cell communication</td>
<td>135</td>
<td>7</td>
<td>129</td>
<td>10</td>
</tr>
<tr>
<td>GO:007155</td>
<td>Cell adhesion</td>
<td>-</td>
<td>&lt; 2</td>
<td>43</td>
<td>17</td>
</tr>
<tr>
<td>GO:030198</td>
<td>Extracellular matrix organization and biogenesis</td>
<td>-</td>
<td>-</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>GO:0043062</td>
<td>Extracellular structure organization and biogenesis</td>
<td>-</td>
<td>-</td>
<td>7</td>
<td>4</td>
</tr>
</tbody>
</table>

GO analysis was done using GATHER [68]. Genes were chosen from the RNA sequencing results using a cut-off of FC +/- 2.00 and a BH [75] adjusted P value of 0.05 using R [71]. All GO results have a P value <0.001.

3.3.4 E-CADHERIN LOSS EFFECTS GENES INVOLVED IN CELL ADHESION

The various genes that encode the proteins involved in the cell-adhesion structures, for example tight junctions, adherens junctions and desmosomes are shown in Table 3.5. Tight junction genes CLDN1, OCLN, and CGN are all significantly upregulated more than 2.00 FC in the CDH1-/- cell line. At the adherens junction, CDH1 and CDH2 were the only genes significantly differentially expressed by more than FC= 2.00. CDH2 was downregulated 2.2 FC (P= 2.37E-04), while, as previously mentioned CDH1 was downregulated 10.0 FC (Table 3.5). L1N7A, although not significant P= 4.20E-01, was downregulated 2.5 FC (Table 3.5). Desmoglein 4 (DSG4) was the only gene significantly differentially expressed by more than FC= 2.00 in the desmosome group, with a FC of 3.0 (P= 9.73E-04) (Table 3.5).
TABLE 3.5. Epithelial cell-cell adhesion proteins.

<table>
<thead>
<tr>
<th>Tight junctions:</th>
<th>Gene name</th>
<th>Fold Change</th>
<th>Adj. P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-catenin</td>
<td>CTNN1A1</td>
<td>1.12</td>
<td>1.66E-02</td>
</tr>
<tr>
<td>Claudin 1</td>
<td>CLDN1</td>
<td>3.08</td>
<td>3.45E-05</td>
</tr>
<tr>
<td>Occludin</td>
<td>OCLN</td>
<td>2.87</td>
<td>2.27E-05</td>
</tr>
<tr>
<td>ZO-1*</td>
<td>TJP1*</td>
<td>-1.20</td>
<td>1.54E-01</td>
</tr>
<tr>
<td>ZO-2*</td>
<td>TJP2*</td>
<td>1.12</td>
<td>5.46E-02</td>
</tr>
<tr>
<td>ZO-3</td>
<td>TJP3</td>
<td>1.87</td>
<td>4.69E-04</td>
</tr>
<tr>
<td>(E-cadherin)</td>
<td>(CDH1)</td>
<td>-10.00</td>
<td>9.10E-06</td>
</tr>
<tr>
<td>Vinculin*</td>
<td>VCL*</td>
<td>-1.11</td>
<td>4.20E-01</td>
</tr>
<tr>
<td>Junctional adhesion molecule A</td>
<td>F11R</td>
<td>1.29</td>
<td>2.04E-04</td>
</tr>
<tr>
<td>Beta-actin</td>
<td>ACTB</td>
<td>1.25</td>
<td>2.66E-02</td>
</tr>
<tr>
<td>Cingulin</td>
<td>CGN</td>
<td>2.99</td>
<td>1.80E-04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Adherens junction:</th>
<th>Gene name</th>
<th>Fold Change</th>
<th>Adj. P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(E-cadherin)</td>
<td>(CDH1)</td>
<td>-10.00</td>
<td>9.10E-06</td>
</tr>
<tr>
<td>N-cadherin</td>
<td>CDH2</td>
<td>-2.18</td>
<td>2.37E-04</td>
</tr>
<tr>
<td>Cadherin 16</td>
<td>CDH16</td>
<td>3.79</td>
<td>6.73E-04</td>
</tr>
<tr>
<td>Cadherin 20</td>
<td>CDH20</td>
<td>3.87</td>
<td>2.89E-02</td>
</tr>
<tr>
<td>p120-catenin</td>
<td>CTNND1</td>
<td>1.19</td>
<td>2.71E-02</td>
</tr>
<tr>
<td>α-catenin</td>
<td>CTNN1A1</td>
<td>1.12</td>
<td>1.66E-02</td>
</tr>
<tr>
<td>β-catenin*</td>
<td>CTNN1B1*</td>
<td>1.03</td>
<td>4.20E-01</td>
</tr>
<tr>
<td>Vinculin*</td>
<td>VCL*</td>
<td>-1.11</td>
<td>4.20E-01</td>
</tr>
<tr>
<td>Alpha-actinin*</td>
<td>ACTN4*</td>
<td>1.07</td>
<td>3.03E-01</td>
</tr>
<tr>
<td>β-actin</td>
<td>ACTB</td>
<td>1.25</td>
<td>2.66E-02</td>
</tr>
<tr>
<td>Formin*</td>
<td>INF2*</td>
<td>1.03</td>
<td>4.20E-01</td>
</tr>
<tr>
<td>Cortactin*</td>
<td>CTTN*</td>
<td>1.06</td>
<td>4.03E-01</td>
</tr>
<tr>
<td>Lin7A*</td>
<td>LIN7A*</td>
<td>-2.45</td>
<td>4.20E-01</td>
</tr>
<tr>
<td>Lin7B*</td>
<td>LIN7B*</td>
<td>-1.59</td>
<td>4.20E-01</td>
</tr>
<tr>
<td>Lin7C*</td>
<td>LIN7C*</td>
<td>1.05</td>
<td>4.20E-01</td>
</tr>
<tr>
<td>Afadin*</td>
<td>MLLT4*</td>
<td>1.19</td>
<td>2.63E-01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Desmosome:</th>
<th>Gene name</th>
<th>Fold Change</th>
<th>Adj. P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desmoglein 2</td>
<td>DSG2</td>
<td>-1.18</td>
<td>4.90E-02</td>
</tr>
<tr>
<td>Desmoglein 3</td>
<td>DSG3</td>
<td>1.66</td>
<td>1.24E-04</td>
</tr>
<tr>
<td>Desmoglein 4</td>
<td>DSG4</td>
<td>3.03</td>
<td>9.73E-04</td>
</tr>
<tr>
<td>Cytokeratin Family</td>
<td>cytokeratins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cadherin Family</td>
<td>cadherins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Desmoplakin*</td>
<td>DSP*</td>
<td>1.00</td>
<td>9.87E-01</td>
</tr>
<tr>
<td>Junction plakoglobin*</td>
<td>JUP*</td>
<td>1.29</td>
<td>4.20E-01</td>
</tr>
</tbody>
</table>

*Not significant P value. (CDH1) was deleted by site specific deletion. Fold changes refer to the CDH1-/- cell line compared to the MCF-10A WT. A full list of claudins, cytokeratins and cadherins is shown in Appendix D.
3.4 GENES ASSOCIATED WITH CANCER

From those genes with marked differential expression, specific groups of genes that were found to be associated with cancers in published literature were highlighted for discussion. These groups included S100-calcium binding proteins, matrix metallopeptidases (MMPs), and mucins. These three groups were chosen as they are involved in gastric and breast cancer and have the possibility of an involvement in the EMT.

S100 proteins are non-ubiquitous calcium-binding proteins expressed in vertebrates exclusively, with 22 known members to date [87]. In the present study, numerous S100 proteins were significantly induced on CDH1 downregulation (Table 3.6), including S100A2, S100A3, S100A7 and S100A10. These four genes have previously been shown to be highly expressed in GC [87]. Furthermore S100A6, S100A7, S100A8 and S100A9 (Table 3.6) have been demonstrated to be upregulated in breast cancer [88,89]. S100A8 and S100A9 form a heterodimer complex, and the two proteins are overexpressed in numerous human cancers including GC [90]. S100A5, S100A7-S100A9, and S100A14 were all significantly upregulated more than 2.0 FC.
MMPs are zinc-dependent endopeptidases, that have been implicated in the degradation of ECM during metastasis [91,92]. There a currently 26 known MMPs, the large majority of which are secreted as inactive zymogens and are activated extacellularly. An additional feature of MMPs is their inhibition by tissue inhibitors of metalloproteinases (TIMPs) [93].

The differential expression of MMPs is shown in Table 3.7. *MMP1* and *MMP9* are both significantly upregulated by 5.8 and 3.3 FC, respectively. *MMP14* was below the FC= 2.00 threshold, but was still significantly upregulated (P= 4.48E-05). *MMP-3* was significantly downregulated by 6.1 FC (Table 3.7). Tissue inhibitors of metalloproteinases (TIMPs) are all downregulated by around 1.2 FC, but *TIMP1* did not have a significant P value (P= 1.43E-01).

---

**TABLE 3.6. S100-calcium binding proteins**

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Fold Change</th>
<th>Adj. P Value</th>
<th>Cancer Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>S100A2</td>
<td>1.49</td>
<td>1.02E-02</td>
<td>Gastric cancer [87]</td>
</tr>
<tr>
<td>S100A3</td>
<td>1.35</td>
<td>7.82E-03</td>
<td>Gastric cancer [87]</td>
</tr>
<tr>
<td>S100A4*</td>
<td>0.93</td>
<td>4.20E-01</td>
<td>-</td>
</tr>
<tr>
<td>S100A5</td>
<td>6.12</td>
<td>2.58E-02</td>
<td>-</td>
</tr>
<tr>
<td>S100A6</td>
<td>1.41</td>
<td>3.62E-02</td>
<td>Breast cancer [88][89]</td>
</tr>
<tr>
<td>S100A7</td>
<td>7.12</td>
<td>8.37E-04</td>
<td>Gastric cancer [87]</td>
</tr>
<tr>
<td>S100A7</td>
<td>7.12</td>
<td>8.37E-04</td>
<td>Breast cancer [88][89]</td>
</tr>
<tr>
<td>S100A8</td>
<td>4.11</td>
<td>1.11E-04</td>
<td>Gastric cancer [90]</td>
</tr>
<tr>
<td>S100A9</td>
<td>2.41</td>
<td>1.19E-03</td>
<td>Gastric cancer [90]</td>
</tr>
<tr>
<td>S100A10</td>
<td>1.61</td>
<td>1.11E-03</td>
<td>Breast cancer [88][89]</td>
</tr>
<tr>
<td>S100A11</td>
<td>1.65</td>
<td>2.60E-03</td>
<td>Breast cancer [89]</td>
</tr>
<tr>
<td>S100A13</td>
<td>1.25</td>
<td>1.65E-03</td>
<td>-</td>
</tr>
<tr>
<td>S100A14</td>
<td>2.01</td>
<td>6.73E-04</td>
<td>-</td>
</tr>
<tr>
<td>S100A16</td>
<td>1.40</td>
<td>1.80E-03</td>
<td>-</td>
</tr>
</tbody>
</table>

*Not significant P value. Fold changes refer to the CDH1-/- cell line compared to the MCF-10A WT.
MMP1, MMP2, MMP3, and MMP9 have been previously shown to be dysregulated in GC [94,95,96]. The levels of MMP1, MMP9 and MMP14 were found to be significantly upregulated on E-cadherin loss (Table 3.7). It is interesting to note that MMP-2 (MMP2), a key protein associated with carcinoma progression [92,96] is not upregulated in the present study (Table 3.7).

TABLE 3.7. Matrix Metallopeptidases

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Fold Change</th>
<th>Adj. P Value</th>
<th>Cancer Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMP1</td>
<td>5.80</td>
<td>1.45E-02</td>
<td>Gastric cancer [94] [95] [96] [97]</td>
</tr>
<tr>
<td>MMP2*</td>
<td>-1.21</td>
<td>8.78E-02</td>
<td>Gastric cancer [81] [94] [95] [96]</td>
</tr>
<tr>
<td>MMP3</td>
<td>-6.08</td>
<td>9.86E-03</td>
<td>Gastric cancer [94] [95] [96]</td>
</tr>
<tr>
<td>MMP9</td>
<td>3.31</td>
<td>9.08E-04</td>
<td>Gastric cancer [94] [95] [96][98] [99] [100]</td>
</tr>
<tr>
<td>MMP14</td>
<td>1.49</td>
<td>4.48E-05</td>
<td></td>
</tr>
<tr>
<td>TIMP1*</td>
<td>1.18</td>
<td>1.53E-01</td>
<td></td>
</tr>
<tr>
<td>TIMP2</td>
<td>1.21</td>
<td>5.12E-03</td>
<td></td>
</tr>
<tr>
<td>TIMP3</td>
<td>1.20</td>
<td>1.01E-03</td>
<td></td>
</tr>
</tbody>
</table>

*Not significant P value. Fold changes refer to the CDH1-/- cell line compared to the MCF-10A WT.

Mucins are transmembrane glycoproteins with multiple functions including cancer cell migration and immune response [101]. MUC1 and MUC2 were significantly upregulated, with MUC2 having the highest differential expression of 2.6 FC (Table 3.8). MUC1 and MUC2 are involved in gastric [102,103] and colon cancer [102,104,105], whereas only MUC1 is involved in breast cancer [102,106]. Additionally, MUC1 and MUC2 are specifically expressed in gastric and breast SRCC [102].

TABLE 3.8. Mucins

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Log Fold Change</th>
<th>Adj. P Value</th>
<th>Cancer Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUC1</td>
<td>1.72</td>
<td>8.34E-05</td>
<td>Gastric cancer [102,103] Breast cancer [102,106] Colon cancer [102,104,105]</td>
</tr>
<tr>
<td>MUC2</td>
<td>2.58</td>
<td>2.66E-02</td>
<td>Gastric cancer [102,103] Colon cancer [102,104,105]</td>
</tr>
</tbody>
</table>

Fold changes refer to the CDH1-/- cell line compared to the MCF-10A WT.
3.5 GENE CRITERIA FOR AN EMT

The primary intention of this paper was to address whether E-cadherin loss alone can cause an EMT in MCF-10A cells. This can be achieved by profiling specific biomarkers to help identify whether an EMT has occurred. The criteria for an EMT requires changes in genes involved in cell-surface proteins, cytoskeletal markers, extracellular matrix proteins, and transcription factors. A further characteristic of an EMT is the downregulation of epithelial markers and the upregulation of mesenchymal markers. These markers have been reviewed in Table 3.9.

For cell-surface proteins, nine genes were chosen from the established literature as criteria for an EMT. These genes were expected to be either upregulated or downregulated in the CDH1-/- cell line (Table 3.9). CDH1 was included in this category, even though it contained the target deletion. Only two genes followed the expected EMT profile, SDC1 and ITGB6. SDC1 was upregulated 1.6 fold in the CDH1-/- cell line (P= 3.11E-05), while ITGB6 was upregulated 1.3 FC (P= 1.40E-02). The other seven cell-surface proteins EMT markers either had no change in expression levels or were downregulated, contrary to that expected for EMT events.

Four genes involved in cytoskeletal markers were also considered. Of these, none showed significant change in expression levels between the two cell lines (Table 3.9). Vimentin (VIM), a key upregulation marker for EMT was not observed in CDH1-/- cells.

In the category of ECM proteins, four of six genes chosen had significant differential gene expressions. Of these, Collagen 4 alpha 1 (COL4A1) and Laminin 1 (LAMA1) followed the expected gene profile of cells undergoing an EMT. However Fibronectin (FN1) and Laminin 5 (LAMA5) were downregulated (not characteristic of EMT) by 7.2 and 1.6 fold respectively (P= 1.78E-04, P= 7.29E-
Collagen 1 alpha 1 (COL1A1) and Collagen 3 alpha 1 (COL3A1) had no significant change in expression.

There were 12 genes selected for the transcription factors or EMT inducers category, none of which followed the expected gene expression profile for an EMT in CDH1−/−. Six of these genes did not change in expression levels; Slug (SNAI2), TWIST1, TWIST2, E47 (TCF3), ETS1 and Goosecoid (GSC). As for those genes that did have altered expression, the highest change was observed in ZEB2 with a fold change of -7.3 (P= 5.02E-05); and second, Snail (SNAI1) with a fold change of -2.49 (P=1.38E-02) (Table 3.9) which is opposite to an EMT event.
### TABLE 3.9. Biomarkers for EMT.

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Expected Regulation</th>
<th>Observed Fold Change</th>
<th>Adj. P Value</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cell-surface proteins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(CDH1)</td>
<td>Down</td>
<td>-10.00</td>
<td>9.10E-06</td>
<td>[107][62]</td>
</tr>
<tr>
<td>CDH2</td>
<td>Up</td>
<td>-2.18</td>
<td>2.37E-04</td>
<td>[107][62][108]</td>
</tr>
<tr>
<td>TIP1*</td>
<td>Down</td>
<td>-1.20</td>
<td>1.54E-01</td>
<td>[49]</td>
</tr>
<tr>
<td>ITGA5</td>
<td>Up</td>
<td>-1.19</td>
<td>8.75E-03</td>
<td>[109][110]</td>
</tr>
<tr>
<td>ITGB1</td>
<td>Up</td>
<td>-1.38</td>
<td>2.91E-04</td>
<td>[109][110]</td>
</tr>
<tr>
<td>ITGB6</td>
<td>Up</td>
<td>1.31</td>
<td>1.40E-02</td>
<td>[109][111]</td>
</tr>
<tr>
<td>SDC1</td>
<td>Up</td>
<td>1.63</td>
<td>3.11E-05</td>
<td>[112,113]</td>
</tr>
<tr>
<td>DDR2</td>
<td>Up</td>
<td>-4.98</td>
<td>2.46E-02</td>
<td>[114][115]</td>
</tr>
<tr>
<td>CDH11*</td>
<td>Up</td>
<td>1.40</td>
<td>3.11E-01</td>
<td>[62]</td>
</tr>
<tr>
<td><strong>Cytoskeletal markers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S100A4*</td>
<td>Up</td>
<td>-0.93</td>
<td>4.20E-01</td>
<td>[116]</td>
</tr>
<tr>
<td>ACTA2*</td>
<td>Up</td>
<td>1.01</td>
<td>9.12E-01</td>
<td>[62][117]</td>
</tr>
<tr>
<td>VIM*</td>
<td>Up</td>
<td>1.09</td>
<td>6.12E-02</td>
<td>[62][118][119]</td>
</tr>
<tr>
<td>CTNNB1*</td>
<td>Up</td>
<td>1.03</td>
<td>4.20E-01</td>
<td>[120]</td>
</tr>
<tr>
<td><strong>Extracellular matrix proteins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COL1A1*</td>
<td>Up</td>
<td>-1.40</td>
<td>1.36E-01</td>
<td>[45]</td>
</tr>
<tr>
<td>COL3A1*</td>
<td>Up</td>
<td>-1.35</td>
<td>4.85E-01</td>
<td>[45]</td>
</tr>
<tr>
<td>COL4A1</td>
<td>Down</td>
<td>-2.62</td>
<td>3.45E-05</td>
<td>[45]</td>
</tr>
<tr>
<td>FN1</td>
<td>Up</td>
<td>-7.24</td>
<td>1.78E-04</td>
<td>[121][122]</td>
</tr>
<tr>
<td>LAMA5</td>
<td>Up</td>
<td>-1.61</td>
<td>7.29E-03</td>
<td>[123]</td>
</tr>
<tr>
<td>LAMA1</td>
<td>Down</td>
<td>-2.49</td>
<td>8.51E-04</td>
<td>[62][124]</td>
</tr>
<tr>
<td><strong>Transcription factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNAI1</td>
<td>Up</td>
<td>-2.49</td>
<td>1.38E-02</td>
<td>[61][125]</td>
</tr>
<tr>
<td>SNAI2*</td>
<td>Up</td>
<td>1.02</td>
<td>5.83E-01</td>
<td>[107][126]</td>
</tr>
<tr>
<td>TWIST1*</td>
<td>Up</td>
<td>-1.14</td>
<td>3.58E-01</td>
<td>[29][127]</td>
</tr>
<tr>
<td>TWIST2*</td>
<td>Up</td>
<td>-1.15</td>
<td>4.20E-01</td>
<td>[107]</td>
</tr>
<tr>
<td>ZEB1</td>
<td>Up</td>
<td>-1.68</td>
<td>1.23E-02</td>
<td>[128][129]</td>
</tr>
<tr>
<td>ZEB2</td>
<td>Up</td>
<td>-7.28</td>
<td>5.02E-05</td>
<td>[130][131]</td>
</tr>
<tr>
<td>TP53*</td>
<td>Down</td>
<td>-1.04</td>
<td>4.20E-01</td>
<td>[132][133]</td>
</tr>
<tr>
<td>SRC</td>
<td>Up</td>
<td>-1.25</td>
<td>2.29E-02</td>
<td>[20]</td>
</tr>
<tr>
<td>TCF3*</td>
<td>Up</td>
<td>1.05</td>
<td>2.08E-01</td>
<td>[134]</td>
</tr>
<tr>
<td>LEF1</td>
<td>Up</td>
<td>-1.76</td>
<td>1.02E-02</td>
<td>[45]</td>
</tr>
<tr>
<td>ETS1*</td>
<td>Up</td>
<td>-1.18</td>
<td>3.02E-01</td>
<td>[45]</td>
</tr>
<tr>
<td>FOXC2</td>
<td>Up</td>
<td>-1.26</td>
<td>2.38E-02</td>
<td>[135]</td>
</tr>
<tr>
<td>GSC*</td>
<td>Up</td>
<td>-1.70</td>
<td>5.58E-01</td>
<td>[136]</td>
</tr>
<tr>
<td>SPARC</td>
<td>Up</td>
<td>-5.41</td>
<td>2.46E-05</td>
<td>[62]</td>
</tr>
</tbody>
</table>

*Not significant P value. (CDH1) was deleted by site specific deletion. Fold changes refer to the CDH1-/- cell line compared to the MCF-10A WT. Genes in bold follow the expected expression profile for an EMT.
Other characteristics of an EMT are the downregulation of epithelial markers and the upregulation of mesenchymal markers. Table 3.10 shows these gene criteria. With the exception of E-cadherin (CDH1), epithelial markers were on the whole upregulated or did not change. Keratin 4 (KRT4) and keratin 10 (KRT10) followed the expected gene criteria for an EMT and were downregulated by 5.9 (P= 1.57E-02) and -2.3 fold (P= 1.95E-03), respectively. Other keratins that showed upregulation in CDH1-/− cells were KRT13, KRT14, KRT15, KRT23, KRT34, and KRT75 (Appendix H). Claudin 1 (CLDN1), Occludin (OCLN) and keratin 9 (KRTN9) were upregulated by more than 2.00 fold. The greatest positive fold change was observed in Claudin 1, with a 3.1 FC (P= 3.45E-05). Numerous claudins were differentially expressed with; CLDN4, CLDN7 and CLDN10 upregulated and CLDN15 and CLDN20 downregulated (Appendix H).

Of the mesenchymal markers chosen to be evaluated, none followed the expected criteria for an EMT. Vimentin, fibroblast-specific protein 1 (S100A4), alpha-SMA (ACAT2), and vitronectin (VTN) had no significant change in gene expression. Fibronectin and N-cadherin (CDH2) were in fact downregulated more than 2.00 fold change as opposed to being upregulated as expected for cells undergoing an EMT (Table 3.10).
TABLE 3.10. Epithelial and Mesenchymal Markers.

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Expected Regulation</th>
<th>Observed</th>
<th>Adj. P Value</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Epithelial Markers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(CDH1)</td>
<td>Down</td>
<td>-10.00</td>
<td>9.10E-06</td>
<td>[107][62][48]</td>
</tr>
<tr>
<td>CTNNB1*</td>
<td>Down</td>
<td>1.03</td>
<td>4.20E-01</td>
<td>[29]</td>
</tr>
<tr>
<td>TJP1*</td>
<td>Down</td>
<td>-1.20</td>
<td>1.54E-01</td>
<td>[49]</td>
</tr>
<tr>
<td>ATP1A1*</td>
<td>Down</td>
<td>1.07</td>
<td>3.47E-01</td>
<td>[49]</td>
</tr>
<tr>
<td>CLDN1</td>
<td>Down</td>
<td>3.08</td>
<td>3.45E-05</td>
<td>[137]</td>
</tr>
<tr>
<td>OCLN</td>
<td>Down</td>
<td>2.87</td>
<td>2.27E-05</td>
<td>[137]</td>
</tr>
<tr>
<td>DSP*</td>
<td>Down</td>
<td>1.00</td>
<td>9.87E-01</td>
<td>[137]</td>
</tr>
<tr>
<td>KRT4</td>
<td>Down</td>
<td>-5.88</td>
<td>1.57E-02</td>
<td>[62]</td>
</tr>
<tr>
<td>KRT8</td>
<td>Down</td>
<td>1.42</td>
<td>7.67E-04</td>
<td>[107][62][137]</td>
</tr>
<tr>
<td>KRT9</td>
<td>Down</td>
<td>2.48</td>
<td>2.89E-02</td>
<td>[62][137]</td>
</tr>
<tr>
<td>KRT10</td>
<td>Down</td>
<td>-2.31</td>
<td>1.95E-03</td>
<td>[62]</td>
</tr>
<tr>
<td>KRT18</td>
<td>Down</td>
<td>1.36</td>
<td>3.86E-04</td>
<td>[107][62][137]</td>
</tr>
<tr>
<td>MUC1</td>
<td>Down</td>
<td>1.72</td>
<td>8.34E-05</td>
<td>[137]</td>
</tr>
<tr>
<td><strong>Mesenchymal Markers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDH2</td>
<td>Up</td>
<td>-2.18</td>
<td>2.37E-04</td>
<td>[107]</td>
</tr>
<tr>
<td>VIM*</td>
<td>Up</td>
<td>1.09</td>
<td>6.12E-02</td>
<td>[107][62][45][137]</td>
</tr>
<tr>
<td>FN1</td>
<td>Up</td>
<td>-7.24</td>
<td>1.78E-04</td>
<td>[107][137]</td>
</tr>
<tr>
<td>S100A4*</td>
<td>Up</td>
<td>-1.07</td>
<td>4.20E-01</td>
<td>[45]</td>
</tr>
<tr>
<td>ACTA2*</td>
<td>Up</td>
<td>1.01</td>
<td>9.12E-01</td>
<td>[62][137]</td>
</tr>
<tr>
<td>VTN*</td>
<td>Up</td>
<td>-1.33</td>
<td>6.07E-01</td>
<td>[137]</td>
</tr>
</tbody>
</table>

*Not significant P value. (CDH1) was deleted by site specific deletion. Fold changes refer to the CDH1-/cell line compared to the MCF-10A WT. Genes in bold follow the expected genotype for an EMT.
3.6 CELL MIGRATION IS AFFECTED BY E-CADHERIN LOSS

Cell wounding assays are a common method to observe cell migration [77,138]. Increased cell motility is a characteristic of an EMT. To test whether CDH1-/- could increase the migration of MCF-10A cells in this experimental setting, a cell wounding assay was undertaken. A thin strip was scraped on 90% confluent monolayers of MCF-10A and MCF-10A CDH1-/- cells using a sterile 200 µL gel loading pipette tip. An example of how scratches were made is depicted in Figure 3.11. Wells B and C for CDH1-/- cells had perfect wounding, however wells A and D for MCF-10A WT were not perfect. However, this was not important for determining the decrease in acellular area. When wounding the confluent monolayer of cells, sometimes an additional scratch was created when pulling the pipette tip back, on the way out. As the cells are adherent, occasionally a small sheet of cells that remained attached to cells in the line of the wound would also be removed from the monolayer, creating a jagged wound.
Figure 3.11. Representative ‘wide view’ of a HiQ4 dish in the BioStation under phase contrast. The BioStation uses 20X magnification to put together a grid of images to produce a ‘wide view’ image. The green lines show the outline of where a wound was made. (A) MCF-10A WT replicate 1, (B) MCF-10A CDH1-/- replicate 1, (C) MCF-10A CDH1-/- replicate 2, (D) MCF-10A WT replicate 2.

Figure 3.12 outlines the sequence of events involved in the cell wounding assay. Cell migration was quantified from the wound edge every 30 minutes over 48 hours. The time scale mentioned in Figure 3.13 is in reference to post-wounding, which is itself 24 hours post-seeding.

Figure 3.12. Diagram of the wounding timeline.
Four hours post-wounding, the $CDH1$-/ cell line had a 12.1% decrease in acellular area, whereas the MCF-10A WT cell line had a 69.8% decrease, a 5.8 fold difference (Figure 3.13A). By 12 hours, MCF-10A WT cells had 0% acellular area, compared to $CDH1$-/ cells at the same time point, which had 50.1% remaining. The $CDH1$-/ cells did not reach 0% acellular area until 32 hours post-wounding, 20 hours after the WT cell line (Figure 3.13A). Microscopy images of the cell wound at 0-48 hours post wounding are shown in Appendix E. In order to verify the migration across the wound was not due solely to proliferation, cell counts over 48 hours were quantified (Figure 3.13B). The cell counts for MCF-10A WT cell line increased by 48.3 cells over the 48 hours in a $3 \times 10^4 \ \mu m^2$ area (Figure 3.13B). In the same area the cell numbers increased by 29.5 cells over 48 hours for the $CDH1$-/ cell line (Figure 3.13B). MCF-10A WT cells migrated across the wound faster, but had a faster proliferation rate. Determining the difference in migration across a wound between cell lines with similar proliferation rates is reasonable, but altered proliferation rates can account in part for differing migration rates.
Figure 3.13. MCF-10A WT cells migrate faster than MCF-10A CDH1-/- . A confluent monolayer of cells was wounded as explained in the methods section and transferred to the BioStation after wounding. (A) Three representative technical replicates for each of MCF-10A WT and MCF-10A CDH1-/- are shown +/- SEM. FIJI [66] was used in order to quantify the percent of acellular area. The zero hour time point is 24 hours post seeding, at which time the isogenic cell lines were seeded at 3.0 x 10^4 cells/well and 4.5 x 10^4 cells/well for MCF-10A WT and MCF-10A CDH1-/-, respectively. (B) Cell numbers (using 0.1 μg/ml Hoechst dye) were counted in FIJI [66] in a 3 x 10^4 μm^2 area at 0, 24 and 48 hours post-wounding.
3.7 THE EFFECT OF CDH1-/− ON CELL ANCHORAGE INDEPENDENT GROWTH

Growth of normal epithelial cells are supported by basement membranes conferring survival and proliferative signals. In the absence of such substratum, normal epithelial cells will undergo apoptosis, a phenomenon termed anoikis [139]. The capacity of cells to evade anoikis and to survive and proliferate without correct positional signals is thought to be a requirement for invasion and metastasis to distant places and a key step in cancer progression. Loss of cell anchorage is also a hallmark characteristic of an EMT.

Cells were grown suspended in a 0.3% agar layer (Figure 3.14), without access to the cell culture plasticware for support. Cancerous cells are able to form colonies in this setting, while non-cancerous cell are not.

Dimethyl thiazolyl diphenyl tetrazolium salt (MTT) was used to assess cell viability as evidence of anchorage independent growth. The two isogenic cell lines were unable to grow unattached to a surface and had no colonies form in soft agar after 24 days (Figure 3.14B-C). The MCF-7 cell line (Figure 3.14A), was used as a positive control as their growth in soft agar is well documented [140]. The MCF-7 cell line had an average of 43 colonies per well. One of the three technical replicates for the positive control had developed a contamination, and was removed prior to staining of the well, hence the well appears blank (Figure 3.14A).

The experimental conditions were adequate for growth unattached to a surface, demonstrated by the ability of 2 x 10³ cells per well of MCF-7 cells to grow in soft agar (Figure 3.14A) [140]. 2 x 10⁴ cells per well were added for each of the two isogenic cell lines in order to increase their chances of forming colonies. This experiment clearly shows that E-cadherin loss does not equip the MCF-10A cell line to evade anoikis.
Figure 3.14. Soft agar assay to determine anchorage-independent growth as a result of E-cadherin loss. Soft agar was prepared as explained in the methods section. Cells were stained with MTT after growing for 24 days in growth media. Three technical replicates for each cell line were used. (A) For the MCF-7 cell line, 2 x 10^3 cells per well were added. (B) For the MCF-10A WT and (C) MCF-10A CDH1 -/- cell lines, 2 x 10^4 cells per well were added. A, B and C are representative images of two biological replicates. The first technical replicate in (A) had a contamination and was removed. This did not affect the other wells.
3.8 THE DIFFERENTIATION CAPABILITY OF THE TWO CELL LINES

Acini or spheroids formed from MCF-10A WT cells grown in three-dimensional culture *in vitro* on Matrigel resemble interlobular ducts of normal breast tissue [141]. This contrasts with most human breast tumour cell lines which do not form acini when grown in 3D culture and instead form non-polarized clusters with limited differentiation [142]. Hence cancerous and non-malignant cell lines can be distinguished by the ability of the non-malignant cells to form structurally and functionally differentiated phenotypes within the basement membrane. Comparing the differentiation capability of the MCF-10A WT and *CDH1*-/− cell lines grown in 3D culture may provide evidence for an EMT which was not obvious from standard 2D cultures.

The ability of the two isogenic cell lines to grow in 3D culture is illustrated in Figure 3.15. Cells from both lines are able to form acini, as has been described earlier in this report (Figure 3.3). *CDH1*-/− cells grown in Matrigel have emerged from the spheroids and formed a monolayer of cells by Day 12 (Figure 3.15C-D), a phenotypic change typical of EMT induction in 3D culture [143]. However MCF-10A WT acini also have cells emerging from acini to form monolayers (Figure 3.15A). The cell clusters at 40X magnification are shown in Appendix F. The most feasible explanation for this observed phenotype is that the *CDH1*-/− cells prefer to grow attached to a surface and grew directly on the 100% matrigel in monolayer, rather than forming polarised 3D structures.
Figure 3.15. Day 12 ‘wide view’ of MCF-10A WT and MCF-10A CDH1/-/ cells grown in matrigel in an 8 well chamber coverglass as described in the methods section. The BioStation uses 20X magnification under phase contrast to put together a grid of images to produce a ‘wide view’ image. (A) MCF-10A WT replicate one and (B) MCF-10A WT replicate two. (C) MCF-10A CDH1/-/ replicate one and (D) MCF-10A CDH1/-/ replicate two.
It’s a waste of time discovering something new if there’s nothing meaningful that can be achieved with this new knowledge- Henry Beetham
4 DISCUSSION

4.1 CELL LINE CHARACTERISATION

A detailed characterisation of MCF-10A WT and CDH1-/- allowed a better understanding of the cell lines before any experiments addressing the main aims were carried out.

First, we confirmed that E-cadherin was knocked-out in the CDH1-/- cell line (Appendix G). This was demonstrated by Western blot analysis (Figure 3.1) where there was a total loss of the E-cadherin protein in the CDH1-/- cell line. The transcriptome analysis also confirmed that CDH1 mRNA was significantly downregulated by 10.0 fold in the CDH1-/- cell line (Tables 3.3 and 3.9).

In 2-D (Figure 3.2) and 3-D (Figure 3.3), CDH1-/- is less ordered, less polarised, less adherent and has less cell-cell contacts compared to the WT. These distinct differences in cellular morphology underlie the beginnings of a phenotype produced by the loss of E-cadherin. Presumably these characteristics observed in the CDH1-/- cells are indicative of the start of an EMT event, this idea will be further established in the following sections.

MCF-10A WT cells consistently demonstrated a faster proliferation rate compared to CDH1-/- cells (Figure 3.5). This may be attributed to cell adherence, whereby there was a lag to log growth phase for the CDH1-/- cells of around 24 hours (Figure 3.4). The slower observed proliferation rate is supported by the transcriptome analysis; upregulation of S100A9 and S100A11, as observed in the CDH1-/- cell line (Table 3.6), has been previously shown to inhibit cell proliferation in myoepithelial cells [144,145].

This lower proliferation rate could also be reflected by the lower nucleoli count in CDH1-/-.

The phase contrast and Hoechst stained images allowed clear visualisation for the presence or absence of nucleoli in cells (Figure 3.6), although an improvement would have been to use an
immunofluorescence nucleoli specific antibody such as anti-fibrillarin [146]. Multiple nucleoli per cell were statistically more common in MCF-10A WT as compared with CDH1-/- with an average of 2.25 and 1.5 nucleoli per cell, respectively (P=0.0001). An increase in nucleoli number commonly observed in cancerous tissues such as prostate cancer [81], is due to an increased rate of proliferation which is linked to an increased demand for ribosome biogenesis [83][147]. Hence increased nucleoli number is associated with certain cancer cells, but this is a mere consequence of the increased metabolic necessities of proliferating cells. The higher proliferation rate in MCF-10A WT as compared to CDH1-/- could therefore be associated with a higher number of nucleoli per cell.

It is highly unlikely the observed differing proliferation rates were an error associated with using the automated cell counter. For each biological replicate, three technical replicates were used and each technical replicate counted at least twice and averaged. Cells were also counted after trypsinization, but without centrifuging cells in order to lose as few cells as possible. Other options to quantify proliferation include BrdU incorporation into DNA of proliferating cells [78], or using a MTT Cell proliferation Assay [148].

### 4.2 CELLULAR IMPACT OF E-CADHERIN LOSS

Understanding the function of genes differentially expressed on E-cadherin loss will help the universal understanding of the effects of E-cadherin on the cell as a whole. The highest positive fold change was observed for Caspase-14 (CASP14) (Table 3.3). It is involved with apoptosis [149] and may contribute to the increase in number of floating cells observed in CDH1-/- cell culture (Figure 3.2). Seven significantly downregulated genes were associated with ECM-organisation and extracellular structure-organisation from GO analysis (Table 3.4), these included ADAMTS3, COL12A1 and PCDHB4 (Appendix C). Additionally, cytoskeleton genes, KRT4 and KT10 and components of the ECM, Collagen type-4 alpha-1 (COL4A1), Laminin-1 (LAMA1) and Fibronectin
(FN1) were significantly downregulated (Table 3.9). This lack of cell-matrix interaction caused by E-cadherin loss could explain the decreased proliferation rate (Figure 3.5), and also further contribute to the increased number of floating cells in the CDH1-/- cultures.

From the negative fold change list, B3GALT3, CHDH, KCNQ3, PDE7B, PRSS21, RBP4, ZNF239 and ZNF485 (Table 3.3) showed the greatest change in expression. Their effects are cell-wide from the mitochondrion (CHDH) [150], potassium voltage-gated channel (KCNQ3), the functioning of second messengers (PDE7B) and cell-surface anchored serine protease (PRSS2).

4.2.1 CELL ADHESION

There are three major modes of cell-cell adhesion in cells: tight junctions, adherens junctions, and desmosomes. E-cadherin KO disturbed cell-cell adhesion stability and polarity, as is demonstrated by Figures 3.2 and 3.3. As a result, CDH1-/- cells appear to have attempted to compensate for the loss of this key regulator of adhesion by upregulating numerous genes involved in these three adhesion complexes (Table 3.5).

Adherens junctions initiate cell-cell contacts and enable maturation and maintenance of epithelial integrity. They consist of E-cadherin, α-actinin, vinculin and intracellular components p120-catenin, β-catenin and α-catenin (Figure 4.1) [151,152]. E-cadherin is required for the binding and localisation of catenins which provide anchorage to the actin cytoskeleton to form stable cell-cell contacts (Figure 4.1) [153,154]. E-cadherin loss caused a significant downregulation for adherens junction component, N-cadherin (CDH2) (Table 3.5), however there was also a significant upregulation of intracellular components p120-catenin, β-actin and α-catenin, as well as cadherins 16 and 20 (Table 3.5). As adherens junction formation leads to the assembly of the tight junction, tight junction formation would also be predicted to be affected by E-cadherin loss (Figure 4.1) [155].
E-cadherin depletion has been shown to disrupt the establishment but not the maintenance of tight junctions [155]. Tight junctions regulate paracellular permeability [156] and establish the boundary between apical and basolateral membranes [157], and so are required for the correct functioning of a cell. The main components of tight junctions (Figure 4.1) were upregulated in E-cadherin negative cells (Table 3.5). These included the transmembrane proteins claudin and occludin, the cytoplasmic scaffolding proteins ZO-3 and cingulin, and the junctional adhesion molecule A (Figure 4.1). Additionally, at the desmosome, desmoglein 4 (DSG4) was significantly upregulated (Table 3.5).

![Figure 4.1.](image)

Figure 4.1. The various proteins associated with tight junctions, adherens junctions and desmosomes are shown. Adapted from Thiery and Sleeman [137].
4.2.2 **CDH1 LOSS CASUES REDUCED MIGRATION**

Cell adhesion and migration are two processes linked to each other. Since cell adhesion was disrupted by E-cadherin loss, it was proposed it would also alter the migration of CDH1/-/ cells. Indeed, the CDH1/-/ cell line displayed reduced migration; after four hours the CDH1/-/ cell line reduced the acellular area by 12%, whereas the MCF-10A WT cell line decreased the acellular area by 70% (Figure 3.13A). Differences in migration rates across the wound were partially affected by proliferation rates (Figure 3.13B). The difference in proliferation over 24 hours post-seeding is 1.7 FC lower in CDH1/-/ cells compared to MCF-10A WT (Figure 3.13B). This is in accordance to the FC calculated for the doubling times of 1.5 (Figure 3.5). However the increased proliferation rate cannot explain the 2.7 FC difference between times to close the wound (Figure 3.14A), hence the MCF-10A have decreased migration capability on E-cadherin loss. Transcriptome analysis showed numerous genes involved with cell migration were significantly downregulated by more than 2.00 FC in the CDH1/-/ cell line; FEZ1, FN1, LAMA1, NRCAM, PF4 and SLIT2 (Tables 3.3 and 3.9). Furthermore, specific downregulation of vimentin (VIM) or N-cadherin (CDH2) expression (Table 3.9) in MCF-10A CDH1/-/ cells effectively decreases cellular motility and invasiveness [108,158]. While creation of integrin-mediated adhesions is required to generate traction for cells to move, excessively robust adhesion to the ECM can impede migration [148]. Expression levels of integrins (ITGB6) and cell-surface proteins involved with cell attachment to the ECM (CD53 (Table 3.3) and SDC1 (Table 3.9)), were significantly upregulated in the CDH1/-/ cell line in this study. Therefore, increased or dysregulated cell-ECM adhesion in the CDH1/-/ cells may contribute to the diminished migration.
S100 proteins, MMPs and mucins were upregulated on E-cadherin loss (Tables 3.6-3.8). There is a wealth of evidence that these specific groups of genes have a role to play in gastric [87] and breast cancers [88,89]. S100A14 may be a novel gene involved in cancer progression. The fact they are upregulated on E-cadherin loss (Tables 3.6-3.8) supports the idea that these CDH1-/- cells have progressed down the pathway to malignancy. S100-proteins can enhance the activity of MMPs [159] which have been correlated with GC stage [96]. As E-cadherin loss occurs early in HDGC, this suggests the induction of MMP1 in particular is involved with carcinoma progression. In this study, expression of MMP9 was increased 3.3 fold on E-cadherin loss (Table 3.7), consistent with previous studies involving downregulation of E-cadherin [160]. The upregulation of MMP9 could be considered a predictive marker of early stage HDGC as it has been shown to have a substantial hazard ratio [98,99] and is associated with poor survival in GC patients [100]. Presumably its ability to degrade the basement membrane [91,92] enables the invasion of cells and is the reason for these detrimental effects.

In renal fibrosis (type 2 EMT), disruption of the basement membrane facilitates EMT [124]. The exact amount of ECM degradation that must occur in order for tumour cell invasion to transpire is unknown, but the breakdown of connective tissue proteins and disease progression is without a doubt correlated with an upregulation of MMPs by tumour cells [93,161]. There has also been a direct correlation shown between high MMP expression and the invasive phenotype of tumour cells, demonstrated in vivo [94] and in vitro [95] in numerous cancers including breast, gastric and colon cancers.
Mucin-1 (MUC1) and mucin-2 (MUC2) are upregulated on CDH1 KO (Table 3.8). A strong association of MUC1 overexpression with prognosis has previously been reported in breast cancer [106]; MUC1 and MUC2 in colon cancer [104,105]; and in GC [103].

Despite these correlations in expression, we must keep in mind that correlation does not equal causation. The differential expression of S100 proteins, MMPs and mucins in the CDH1-/- cell line does provide however, evidence that the threshold for malignant progression is lowered in these cells. This begins to paint a picture of a cell with the loss of E-cadherin being poised for carcinoma progression.

### 4.3 E-CADHERIN DOWNREGULATION ALONE DOES NOT CAUSE AN EMT

The primary aim of this research was to determine if E-cadherin KO alone is sufficient to induce an EMT in MCF-10A cells and consequently gain insight into the potential role of E-cadherin loss in the disease progression of early lesions in the stomachs of HDGC individuals.

In this study, biomarkers were used to identify the occurrence of a bona fide EMT in the CDH1-/- cell line. The expression profile of the majority of the EMT biomarkers in CDH1-/- cells did not follow their expected differential expression expected of an EMT event (Table 3.9 and 3.10). In order for an EMT to occur, there must be a differential expression of genes involved in cell-surface proteins, cytoskeletal markers, extracellular matrix proteins, transcription factors, epithelial markers and mesenchymal markers. However of these gene criteria, loss of epithelial markers CDH1, KRT4 and KRT10, and ECM proteins COL4A1 and LAMA1 and the upregulation of cell surface proteins ITGB6 and SDC1, were the only notable changes in CDH1-/- (Tables 3.9 and 3.10). There was a distinct lack of upregulation of well-documented EMT regulators such as vimentin [62,118,119], fibronectin [121,122], Twist1 [29,127] and Snail [61,125]. Hence E-cadherin loss alone does not cause
an EMT in mammary epithelial cells [20,52,56,132]. It should also be recognised that although an EMT was not observed, the cells appear primed for one.

During an EMT, the epithelial phenotype which is characterised by strong cell-cell junctions and cell polarity, is replaced by a mesenchymal phenotype, with reduced cell-cell interactions. Cancer progression also depends on the loss of cell adhesion. 43 genes involved with cell-adhesion (Bayes factor= 17) were downregulated on E-cadherin loss (Table 3.4). These included cadherins, collagens, and protocadherins. A full list of these genes is shown in Appendix C. However, there was an upregulation of tight junction and desmosome genes to compensate for the loss of E-cadherin (Chapter 4.2.1). The dynamic formation and termination of junctional complexes is a key process during EMT. The loss of cell contacts and the secretion of MMPs from the surface of CDH1-/- cells in the stomachs of HDGC individuals may be sufficient to allow the cells to penetrate the basement membrane and lead to the ‘pagetoid’ spread of in situ signet ring cells and the development of the stage T1a foci.

The cell migration results (Figure 3.13) also support the conclusion that CDH1 loss is insufficient to produce an EMT. One of the key characteristics of an EMT is increased migration capacity which CDH1-/- cells clearly lack.

4.3.1 MCF-10A CDH1-/- CELLS DO NOT DISPLAY ANCHORAGE INDEPENDENT GROWTH

Neither isogenic cell lines were able to grow in soft agar (Figure 3.15B-C). This was expected for MCF-10A WT cells, as they are highly normalised epithelial cells and the ability of cells to grow unattached to a surface is a phenotype of cancerous cells. The CDH1-/- cell line however, had also not acquired anoikis resistance on E-cadherin loss. The fact that CDH1-/- cells show an inability to survive and proliferate unattached to a membrane (in soft agar) has important implications for
patients. It reinforces the idea that E-cadherin KO alone is unlikely to be sufficient for an EMT and the subsequent disease progression.

---

### 4.3.2 INDICATIVE EVIDENCE OF AN EMT FROM 3D CULTURE

The ability of both cell lines to form acini by Day 3 when grown in 3D culture has been demonstrated in the current study (Figure 3.3). By Day 12, phenotypic changes suggestive of an EMT were observed in MCF-10A CDH1-/- cells (Figure 3.15C-D). This provides possible evidence of an EMT in three-dimensional culture.

There are however a number of different explanations for the observed outgrowth of CDH1-/- cells, indicative of an EMT. First of all it should be noted that as this phenotype was only observed in one experiment it is difficult to address the cause. Further biological replicates are required before compelling conclusions can be drawn. MCF-10A WT cells have been previously shown to have spontaneous morphological and phenotypic EMT-like changes in response to a low cell density [62]. A lower cell density in both wells could consequently explain the EMT-like changes observed in CDH1-/- cells.

---

### 4.3.3 GENES NECESSARY FOR E-CADHERIN INDUCED EMT

The transcriptome profiling along with wounding and soft agar experiments demonstrate that mammary epithelial cells with E-cadherin loss alone is not sufficient for an EMT to occur. Other events additional to E-cadherin loss are required to push these cells towards an EMT. After all the penetrance of HDGC is incomplete with the estimated cumulative risk of 80% (95% CI) for men and women by age 80 years [12]. Clearly there are other factors involved with progression of these early lesions.
There are various theories for the likely mechanisms behind the EMT with the prerequisite of E-cadherin downregulation. These include another EMT inducer such as kinase c-Src (SRC) and its downstream targets FAK and Stat-3 [20]; dysfunction of other tumour suppressor genes such as p53 [57]; DNA damage [44] or over expression of oncogenes such as c-myc [58]. Some proteins that regulate the EMT such as p120 [59]; Twist [60]; LOXL2; and Snail [61] may also independently contribute to a malignant phenotype.

Twist upregulation has been shown to be necessary for E-cadherin-loss induced metastasis [29]. In the present study, Twist and other TFs such as SNAI1 and ZEB2 were downregulated 2.5 fold and 7.3 fold, respectively in the present study (Table 3.9). Background mutations in immortalised cell lines such as those used by Onder et al [29] may facilitate EMTs. A previous human breast epithelial cell line used to demonstrate the role of E-cadherin KO in EMT was immortalised using SV40 large T-antigen [29]. SV40 large T-antigen has been shown to inhibit various tumour suppressor genes including p53 (TP53) [162,163,164]. In the present study, TP53 was found to have insignificant differential expression (Table 3.9). Therefore perhaps the combined inactivation of E-cadherin and p53 are necessary for a complete EMT [132,133]. p53 inactivation alone may be sufficient to induce an EMT and lead to GC [165,166]. Additionally p53 inactivation has been described in 0-21% of diffuse gastric cancers [167], thus there is a high possibility of p53 inactivation being of key importance to E-cadherin-loss induced EMT and they almost certainly have synergistic effects in the progression of HDGC.

One MMP that has been shown to directly cause an EMT in vitro in mammary epithelial cells [168], and in vivo in mice [169] is MMP-3 (MMP3). However, it is significantly downregulated in the CDH1-/- cell line (Table 3.7). MMP-3 upregulation could be a further change required for E-cadherin-loss induced EMT.
4.4 CLINICAL RELEVANCE

The variety of possible mechanisms behind the EMT explain why there is such unpredictability with HDGC, as it is not known what causes the sudden shift from indolent signet ring cells to metastatic cells. This is a major gap in the current knowledge of HDGC. Better characterisation of the EMT and E-cadherin’s role within this context is vital to gain a clearer understanding of the HDGC disease process. This may allow pathologists to ascertain the likelihood of progression of T1a foci from benign to malignant and provide a better prediction of HDGC risk for CDH1 mutation carriers.

There is much controversy in the literature about the optimum clinical management of HDGC. This is because the value of surveillance endoscopy regimes are unproven and the detrimental consequences of prophylactic gastrectomy. It is recommended that genetic counselling and testing is started from the age of consent in families diagnosed with HDGC [24]; in New Zealand genetic testing consequently begins around 16 years of age. Prophylactic total gastrectomy is the single option to abolish an inherited risk of GC. Mutation carriers are advised to undergo prophylactic gastrectomy after about 20 years of age [14] based upon the understanding that at the age of 20 years the mortality rate associated with a prophylactic gastrectomy is the same as the risk of advanced HDGC [9,17].

The penetrance of HDGC is incomplete with the estimated cumulative risk of 80% (95% CI) for men and women by age 80 years [12] (analysis updated in 2008, unpublished data). In women, the estimated cumulative risk of LBC was 39% by age 80 years [12]. Hence it would appear E-cadherin negative cells can remain indolent for many years before GC or LBC develop. Evidence from the present study provides a piece to the puzzle; It is now clear that it is not a fait accompli that the early lesions will progress by means of an EMT and further ‘hits’ such as DNA damage are
likely to be required. The fact that there is possibly more time to act between E-cadherin loss and the development of cancer means some gastrectomies could be delayed.

The identification of additional EMT triggers behind the E-cadherin-loss induced EMT such as p53 KO and c-Src kinase could lead to better markers of risk for predictive testing and better patient stratification. Such classifications would allow for the targeting of preventative or more intensive surveillance strategies to those at higher risk. They could also produce possible drug targets such as inhibitors of c-Src, and MMP-9 in order to enable chemo-preventive steps to impede the progress of HDGC.

4.5 FUTURE DIRECTIONS

This research has provided the foundation for additional studies to refine the impact of CDH1 loss on epithelial cell behaviour.

- As previously mentioned, additional factors are behind the E-cadherin-loss induced EMT in MCF-10A cells. One possible factor is p53 (TP53) inhibition. To test this, shRNA could be used to knock-down TP53 in CDH1/- cells. The invasion capability of MCF-10A CDH1 KO and p53 KO cells could then be determined using Matrigel-coated Transwell invasion assays [29]. Identifying the presence of mutations in p53 in SRCC from HDGC patients might indicate early stage tumours with an increased risk of carcinoma progression.

- To further characterise the progression of an EMT in E-cadherin KO cells, well-characterised inducers of EMT such as TGF-β [122], Snail [61,125], Twist [29,127], or Goosecoid [136] could be added in low levels to both cell lines. Similar experiments to this study could then be carried out to determine if there is a pronounced effect on CDH1/- compared to MCF-10A WT cells.
• To test the impact of DNA damage, Cisplatin treatment could be used on both isogenic cell lines [170,171].

• In order to determine the effect of MMP upregulation in invasion, a MMP inhibitor such as galardin (GM6001) [172] could be used.

• To verify migration rates and determine the capacity of CDH1-/- cells to invade, a Transwell migration assay could be undertaken, as previously described [148].

• An improved 3D modelling system could be explored. For example Alvetex, a highly porous polystyrene scaffold, has the ability to achieve greater differentiation and hence improved formation of 3D structures resembling in vivo ducts.

• A further improvement would be to fix the cells in 3D-culture and undertake immunofluorescence using antibodies against E-cadherin, β-catenin, vimentin, and laminin-5. These antibodies would be used because they possess location-specific expression in acini. This would allow us to demonstrate that the MCF-10A WT acini have a normal architecture, and possibly if the locations of these antibodies are irregular, that MCF-10A CDH1-/- cells have abnormal structure.

• In order to further characterise the two isogenic cell lines, an Annexin V assay could be performed to detect apoptosis in the two cell lines [173]. The

• Finally, the effect of CDH1 loss on the adhesion to the ECM could be quantified using an adhesion assay as has been performed by Alexopoulou et al [148].
4.6 SUMMARY

This study undertook the first characterisation of MCF-10A WT and MCF-10A CDH1-/− cells. E-cadherin loss in this mammary epithelial cell line produced various observable morphological and proliferative changes.

Although loss of E-cadherin may be a necessary event for an EMT, its loss does not appear to be sufficient. This was highlighted by the transcriptome analysis of CDH1-/− cells in comparison to the WT which revealed that the loss of E-cadherin alone in MCF-10A cells does not cause a complete epithelial-mesenchymal conversion. E-cadherin loss slowed the migration rate of cells, but did not enable growth in soft agar, nor cause increased invasion.

These observations may have important implications for the patients and families of those diagnosed with HDGC; the data suggests that it is not a fait accompli that early lesions will undergo an EMT and progress aggressively and it underscores the apparent indolence of most early stage SRCCs observed in HDGC.

This may provide the opportunity for interventions that can block the drivers of the EMT.
Phosphate Buffered Saline (PBS)

One PBS tablet was added to 100 mL deionized water and then autoclaved.

1.5 M Tris-HCl pH 8.8

18.7 g Tris in 25 mL distilled water, pH balanced to 8.8 with HCL using the Hi 9321 Microprocessor pH meter and made to a total volume of 100 mL with distilled water.

0.5M Tris-HCL pH 6.8

6.057 g Tris in 25 mL distilled water, pH balanced to 6.8 with HCL using the Hi 9321 Microprocessor pH meter and made to a total volume of 100 mL with distilled water.

Components for 2X 10% SDS polyacrylamide gel

<table>
<thead>
<tr>
<th>Component</th>
<th>Resolving gel (10%)</th>
<th>Stacking gel (~4%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30% acrylamide</td>
<td>6.7 mL</td>
<td>1.3 mL</td>
</tr>
<tr>
<td>1.5 M TrisHCl (pH 8.8)</td>
<td>5.0 mL</td>
<td></td>
</tr>
<tr>
<td>0.5M Tris-Cl (pH 6.8)</td>
<td></td>
<td>1.95 mL</td>
</tr>
<tr>
<td>10% SDS</td>
<td>0.2 mL</td>
<td>50 μL</td>
</tr>
<tr>
<td>TEMED</td>
<td>10 μL</td>
<td>10 μL</td>
</tr>
<tr>
<td>10% APS</td>
<td>100 μL</td>
<td>40 μL</td>
</tr>
<tr>
<td>Water</td>
<td>8.0 mL</td>
<td>4.7 mL</td>
</tr>
<tr>
<td>Total volume</td>
<td>20.01 mL</td>
<td>8.05 mL</td>
</tr>
</tbody>
</table>

10X Running buffer 1 Litre (Stock)

Reagents:

1X : 25 mM Tris-HCL, 190 mM Glycine, 0.1% (w/v) SDS

10X 1 Litres:

Tris 30.3 g, Glycine 142.6 g, SDS 10 g

Make up to 1 L with MQ. Stir with magnetic flea. Make 1X Running Buffer (Working) to 1 L with MQ.
**20X Transfer Buffer 1 Litre (Stock)**

Reagents:

1X : 48 mM Tris-HCL, 39 mM Glycine

20X 1 Litres:

Tris 116.29 g, Glycine 150.1 g

Then make to 1 L with MQ, stir with magnetic flea to get all reagent into solution.

**1X Transfer buffer (Working)**

20X stock 50 mL, Methanol 100 mL (to make 10%)

Make this to 1 L with MQ.

**PBS 0.2% Tween (PBS-T)**

200 μL Tween20 added to 1 Litre of PBS

**Blocking Buffer** (5% skim in PBS 0.2% Tween)

Add 50 g skim milk powder (Pams) to 1 Litre PBS-T

**1X Protein loading buffer**

2% SDS, 5% 2-mercaptoehtanol, 10% glycerol, 0.1% bromophenol blue, 0.05 M Tris HCl pH 6.8.

**Ice-cold cell culture freezing media**

90% Growth media + 10% DMSO.
5.2 APPENDIX B - CONCENTRATIONS OF SAMPLES FROM RNA EXTRACTION EXPERIMENTS I AND II

<table>
<thead>
<tr>
<th>Experiment One</th>
<th>Nanodrop (ng/μL)</th>
<th>Qubit (ng/μL)</th>
<th>Bioanalyser (ng/μL)</th>
<th>Volume (μL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCF-10A WT (1)</td>
<td>562</td>
<td>360</td>
<td>503</td>
<td>45</td>
</tr>
<tr>
<td>MCF-10A WT (2)</td>
<td>560</td>
<td>583.2</td>
<td>441</td>
<td>45</td>
</tr>
<tr>
<td>MCF-10A CDH1-/- (1)</td>
<td>856</td>
<td>650.16</td>
<td>789</td>
<td>45</td>
</tr>
<tr>
<td>MCF-10A CDH1-/- (2)</td>
<td>877</td>
<td>876.6</td>
<td>838</td>
<td>45</td>
</tr>
<tr>
<td>Experiment Two</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCF-10A WT (1)</td>
<td>576</td>
<td>346.19</td>
<td>484</td>
<td>24.42</td>
</tr>
<tr>
<td>MCF-10A WT (2)</td>
<td>545</td>
<td>379.22</td>
<td>557</td>
<td>27.5</td>
</tr>
<tr>
<td>MCF-10A CDH1-/- (1)</td>
<td>961</td>
<td>1049.65</td>
<td>544</td>
<td>25</td>
</tr>
<tr>
<td>MCF-10A CDH1-/- (2)</td>
<td>1036</td>
<td>1222</td>
<td>932</td>
<td>25</td>
</tr>
</tbody>
</table>

Nanodrop, Qubit and Bioanalyser assays were used to quantitate total RNA. RNA extraction Experiment I and Experiment II are biological replicates. Each experiment had two technical replicates shown by (1) and (2). This table is supplementary to Figure 3.7.

Experiment I RNA samples were given to NZGL.

5.3 APPENDIX C- GO RESULTS

GO_logFC= +1.00, adj.P.value= 0.05

Total genes= 598
### GO logFC = -1.00, adj.P.value = 0.05

Total genes = 519

### GO logFC = +/- 1.00, adj.P.value = 0.05

Total genes = 1117
5.4 APPENDIX D - EPITHELIAL CELL-CELL ADHESION GENES

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Fold Change</th>
<th>Adj. P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Claudin Family</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLDN1</td>
<td>3.08</td>
<td>3.45E-05</td>
</tr>
<tr>
<td>CLDN3*</td>
<td>3.46</td>
<td>3.03E-01</td>
</tr>
<tr>
<td>CLDN4</td>
<td>3.78</td>
<td>4.87E-05</td>
</tr>
<tr>
<td>CLDN7</td>
<td>2.73</td>
<td>1.51E-05</td>
</tr>
<tr>
<td>CLDN10</td>
<td>9.04</td>
<td>2.61E-02</td>
</tr>
<tr>
<td>CLDN14*</td>
<td>2.22</td>
<td>4.20E-01</td>
</tr>
<tr>
<td>CLDN15</td>
<td>-2.02</td>
<td>1.48E-02</td>
</tr>
<tr>
<td>CLDN20</td>
<td>-3.86</td>
<td>2.65E-02</td>
</tr>
<tr>
<td><strong>Cadherin Family</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(CDH1)</td>
<td>-10.00</td>
<td>9.10E-06</td>
</tr>
<tr>
<td>CDH2</td>
<td>-2.18</td>
<td>2.37E-04</td>
</tr>
<tr>
<td>CDH3</td>
<td>1.77</td>
<td>1.51E-05</td>
</tr>
<tr>
<td>CDH16</td>
<td>3.79</td>
<td>6.73E-04</td>
</tr>
<tr>
<td>CDH20</td>
<td>3.87</td>
<td>2.89E-02</td>
</tr>
<tr>
<td><strong>Cytokeratin Family</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KRT4</td>
<td>-5.88</td>
<td>1.57E-02</td>
</tr>
<tr>
<td>KRT6B</td>
<td>4.66</td>
<td>4.21E-03</td>
</tr>
<tr>
<td>KRT10</td>
<td>-2.31</td>
<td>1.95E-03</td>
</tr>
<tr>
<td>KRT13</td>
<td>8.01</td>
<td>3.33E-03</td>
</tr>
<tr>
<td>KRT14</td>
<td>2.22</td>
<td>3.98E-03</td>
</tr>
<tr>
<td>KRT15</td>
<td>2.51</td>
<td>2.26E-03</td>
</tr>
<tr>
<td>KRT23</td>
<td>3.69</td>
<td>3.13E-02</td>
</tr>
<tr>
<td>KRT34</td>
<td>15.05</td>
<td>4.42E-03</td>
</tr>
<tr>
<td>KRT75</td>
<td>3.07</td>
<td>4.30E-02</td>
</tr>
</tbody>
</table>

*Not significant P value. (CDH1) was deleted by site specific deletion. Fold changes refer to the CDH1-/- cell line compared to the MCF-10A WT. These lists are supplementary to Table 3.3.
### 5.5 APPENDIX E: CELL WOUNDING PICTURES

<table>
<thead>
<tr>
<th>MCF-10A WT</th>
<th>MCF-10A CDH1-/-</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
</tbody>
</table>

- **0 Hours**
- **4 Hours**
One of the three representative technical replicates for each cell line are shown. Pictures were taken in the BioStation under phase contrast at 20X magnification.
5.6 APPENDIX F- CELLS GROWN IN MATRIGEL AT DAY 12

Day 12 ‘wide view’ of MCF-10A WT and MCF-10A CDH1-/- cells grown in Matrigel in an 8 well chamber coverglass as described in the methods section. Images were taken using the BioStation at 40X magnification under phase contrast. (A) MCF-10A WT replicate I and (B) MCF-10A WT replicate II. (C) MCF-10A CDH1-/- replicate I and (D) MCF-10A CDH1-/- replicate II.
5.7 APPENDIX G - CREATION OF CDH1 KNOCKOUT IN MCF-10A CELLS

Site-specific deletion at the CDH1 Locus in MCF10A cell line, Alleles 1 and 2 – 4 base pair deletion:

ATAACTGAAGAAGCGCTTAAGCCGTTTTTCAGCTACA
TGTTGTTTGCTGCTCTTATATCATTAAAGCCAGAGCT
TGTCCCGTTCAGATATCGGATTTGGAGAGACACT
GCCAACTGCGTGAGATTAATCCGGACACTGGTGC
CATTTCCACCTCGGGCTGAGTTGACAGGGAGGATT
TTGAGGACGTGAAGAAGACGACGTACTACACAGCCTA
ACTATAgctacaGACAATGGTAAGGGGCCTCATCT
GAGCCTTTGCTGCCCTCGACCTCTAGCTAGTCAG
TTCTTGCCCCTCCCTCTTTTGGAGGGAAGAGTT
CATTCTTTTCTTTATCCTTTTTGCTGATTG
ATTTGTTATAATGTATGGAGTACAAGGGTAATTTTG
TTACATGCATAGTGTAGTGGTAAAGTC

Schematic of the genomic sequence at the target region (exon 11) recognized by the ZFN pair; the resulting deletion, and the CEL-I primer sequences:

CEL-I Primers - **Bolded and underlined**
ZFN binding site - **UPPER CASE, BOLDED RED**
ZFN cut site - **lower case red**
Deletion - **yellow highlighted**

(Adapted from Sigma-Aldrich data sheet)
6 REFERENCES


[67] GraphPad Prism version 5.00 for Windows. GraphPad Prism version 5.00 for Windows. GraphPad Software 2007.


[70] Papers, version 1.0.3. Papers, version 1.0.3. Mekentosj, Gerberastraat, Aalsmeer 2012.


