THE UTILISATION OF MICROFLUIDIC qPCR TECHNOLOGY FOR THE IDENTIFICATION OF NOVEL CANCER BIOMARKERS

A thesis submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy at the Faculty of Science and Engineering, University of Limerick, Ireland

Christopher J Hayes, BEng (Hons) Mech.Eng

Under the supervision of

Dr. Tara Dalton
Stokes Laboratories
University of Limerick

Dr. Patrick Kiely
General Entry Medical School
University of Limerick

Submitted to the University of Limerick, November 2017
Declaration

The substance of this thesis is the original work of the author and due reference and acknowledgement has been made, where necessary, to the work of others. No part of this thesis has been submitted in canditure for any degree.

___________________________________________

Christopher J Hayes (Candidate)

___________________________________________

Dr. Tara M Dalton (Supervisor)

___________________________________________

Dr. Patrick A Kiely (Supervisor)
Abstract

Cancer is a major cause of death worldwide causing 1 in 6 deaths globally with approximately 14 million new cases and 8.8 million cancer-related deaths in 2015. This figure is expected to increase to 21.7 million cases and 13 million deaths by 2030. In Ireland, an average of 30,000 new cases of cancer are diagnosed each year. The number is also expected to rise to over 40,000 per year by 2020. A key focus to reverse this trend is the identification of new prognostic markers and therapeutic targets that can be used to predict the incidence or outcome of the disease. These targets can also be used to screen populations to identify symptomatic patients, identify differential diagnosis and for the clinical staging of cancer.

The quest for a better understanding of the carcinogenesis process has led to impressive growth in the field of large-scale and high-throughput biology which has become a driving force behind the emergence of new technologies. Large-scale gene expression profiling is an essential tool for many biological and medical investigations. Higher throughputs and increased levels of specificity coupled with reduced volumes of samples and reagents is a dominant factor. The development and application of a microfluidic platform to analyse biological samples at a rate of approximately one thousand data points per hour is demonstrated in this thesis. Utilisation of such a platform gives an advantage over many traditional systems and allows for the manipulation of smaller quantities to achieve corresponding biological outcomes.

This thesis investigates the application of microfluidic gene expression profiling to identify biomarkers in colorectal and breast carcinoma, envisioning that future validation tests using a microfluidic platform could produce an expression assay to aid the stratification of the disease. The focus is on candidate genetic targets that are regulating extracellular matrices production which are central components of a tumour’s microenvironment, possessing key roles for tumour formation, growth and potential metastatic spread. In two genetic studies of colorectal and breast cancer malignancies, a number of genes from the extracellular matrix family emerge as being statistically differentially expressed. Correlations of the genes with the histopathological parameters along with heterogeneity of the tumours are investigated also. Genetic targets identified could potentially serve as molecular markers in screening, diagnosis and treatment strategies for the individualised care of patients in the future.
Publications

Some of the findings reported in this thesis have been published in the following articles.

Peer-Reviewed Journal Articles


Conference Proceedings

Acknowledgements

The final task in this journey is to acknowledge all the people who have helped me along the way. Firstly, I would like to thank my supervisors Dr. Tara Dalton and Dr. Patrick Kiely; you have been tremendous mentors for me. Tara, your guidance and support on both research topics as well as my personal development has been essential. I can honestly say that every time I came out from a meeting with you, not only did I not only get the answer to my questions but also acquired a focused and renewed drive for my work. Inspirational! And to Pat, the belief and encouragement that you have has been a driving force throughout the completion of this thesis. I would like to thank you both for encouraging my research and for allowing me to grow as a research scientist.

I would like to thank Professor Mark Davies for granting me an amazing opportunity back in 2010 to join Stokes Bio, which was where it all began for me. To the team there including Barrett, Mike Sayers, Noel, Chawky, Biff, Damo, Dave K, Xiaona, Katie - you guys are great and true leaders. In Stokes Laboratories, thanks to Fionnuala for her wonderful assistance with anything and everything from day one. Thanks to Paddy for his trojan work and great chats in the workshop and all the best in your retirement. Thanks to my fellow postgrads, past and present, for making my time there so enjoyable especially the ‘ten am tea’ team of Conor Mc, Marie, Donovan, Paudie and to the rest of the gang Mary, Susan, Seanie, the three Mikes (Mike K, Mike G, Mike C), Anne, Finola, Emer, Lisa, Niamh, Tim and Jane. Thanks to Jeff, Pat, Vanessa, Eric, Maurice and Ronan for sound advice over the years. Thank you to Professor J.C Coffey and the surgical team at University Hospital Limerick. I did not get the chance to meet all of you face-to-face but without your invaluable work and dedication, this research would not be possible so thank you.

A special thanks to my wonderful family. Words cannot express how grateful I am to my mother, father, brother and two sisters for all of the sacrifices that you have made on my behalf throughout my life. Finally, I would like express appreciation to Aisling for her support, encouragement and for never being more than a phone call away at any hour.
To my parents Theresa and James for their continuous support and encouragement
Contents

Abstract ........................................................................................................................................ i

Publications................................................................................................................................... ii

Dedication ...................................................................................................................................... ii

Acknowledgements .................................................................................................................. iii

List of Tables .............................................................................................................................. ix

List of Figures ............................................................................................................................. x

Abbreviations .............................................................................................................................. xiii

1 Introduction............................................................................................................................... 1

2 Literature Review .................................................................................................................... 18

2.1 Cancer - An Autonomous and Evolving Disease ............................................................. 18

2.1.1 Colorectal Cancer .............................................................................................................. 20

2.1.2 Breast Cancer .................................................................................................................... 22

2.1.3 Associated Risks ............................................................................................................... 24

2.1.4 Metastasis .......................................................................................................................... 30

2.1.5 Biomarker Discovery ......................................................................................................... 34

2.1.6 Genetic signatures ............................................................................................................ 41

2.2 The Extracellular Matrix .................................................................................................... 43

2.2.1 Overview .......................................................................................................................... 43

2.2.2 Fibrous proteins ............................................................................................................... 46

2.2.3 Proteoglycans and Glycoproteins .................................................................................... 48

2.2.4 Integrins ............................................................................................................................ 51

2.2.5 Growth factors .................................................................................................................. 54

2.2.6 Matrix Metalloproteinases ............................................................................................... 55

2.2.7 Other ECM Molecules ...................................................................................................... 56
5 Instrument Validation ................................................................. 112
  5.1 Introduction ........................................................................... 112
  5.2 Applied Biosystems AB7900HT PCR Platform .................... 113
  5.3 Validation Tests ................................................................. 114
    5.3.1 Assay Efficiency Validation ........................................... 114
    5.3.2 Line to Line Variance Validation ................................. 116
    5.3.3 Expression Knockdown Validation ............................... 118
    5.3.4 Dual Expression Validation ........................................... 119
    5.3.5 Contamination Testing Validation ............................... 121
  5.4 Chapter Close ..................................................................... 122
6 Experimental Results ............................................................... 124
  6.1 Abstract ............................................................................... 124
  6.2 Colorectal Profiling ECM study ......................................... 124
    6.2.1 Differential gene expression in matched samples .......... 124
    6.2.2 Dysregulation Discussion .............................................. 127
    6.2.3 Analysis of expression coordination and gene networking 131
    6.2.4 Histopathological parameters associations ................. 135
    6.2.5 Verification study using qPCR Instrument ................... 139
  6.3 Breast Profiling ECM study ................................................ 142
    6.3.1 Abstract ....................................................................... 142
    6.3.2 Breast Profiling Findings .............................................. 143
    6.3.3 Dysregulation Discussion .............................................. 145
    6.3.4 Verification study using commercial qPCR instrument .... 155
  6.4 Chapter Closure .................................................................. 159
7 Conclusions and Recommendations ............................................. 162
  7.1 Conclusions ........................................................................ 162
  7.2 Recommendations for future work ..................................... 164
    7.2.1 Instrumentation Recommendations ............................ 164
    7.2.2 Biological Recommendations ....................................... 166
7.2.3 Pathological Recommendations .................................................. 167

Bibliography ................................................................................................. 168

Appendix A .................................................................................................. 215
List of Tables
Table 1- List of FDA-approved tumour markers currently used in clinical practice. 38
Table 2- Candidate cancer-related ECM genes .......................................................... 60
Table 3 - Reference gene stability in colon tissue samples ........................................ 75
Table 4- Reference genes used for normalisation ...................................................... 76
Table 5- RACK1 and PGK1 expression in HCT116 colon cancer cell cDNA ..... 121
Table 6 - Expression fold change between normal and cancerous samples .......... 125
Table 7 - Genes with dysregulated expression ........................................................ 131
Table 8 - Clinicopathological data of patients selected for the analysis study. ...... 136
Table 9 – Sample and Reagent Volumes for ABI PCR verification studies ...... 140
Table 10- Results from parallel verification experiments on the GEI and ABI platforms, .......................................................... 141
Table 11 – Differential expression fold change patterns determined by REST ..... 144
Table 12- Cq Values for verification study .............................................................. 156
Table 13 – Expression fold changes between normal sample and corresponding malignant samples .............................................................................................................. 157
List of Figures

Figure 1 – Cancer cells differentiate and proliferate at an increased rate in comparison to normal cells.................................................................2
Figure 2 - DNA Structure- The nucleotides pair with complimentary bases and hydrogen bonds form between the bases to create stable “base pairs”......................3
Figure 3- Central Dogma of Molecular Biology..............................................4
Figure 4- mRNA is synthesised within the nucleus of the cell and moves into the cytoplasm. .........................................................................................5
Figure 5- Key constituents of the ECM. ..........................................................7
Figure 6 - Six hallmarks identified during the development of cancer ..........19
Figure 7- The digestive system .....................................................................20
Figure 8 - World incidence age-standardised rates for colorectal cancer for both sexes in 2012...............................................................21
Figure 9 - Current Worldwide Breast Cancer Incidence Rate. .......................22
Figure 10- A stage II breast cancer .................................................................23
Figure 11- Some potential risk factors of cancer formation............................25
Figure 12- The process of metastasis ..............................................................30
Figure 13- Abnormal cells which have developed into a cancer ....................31
Figure 14- Abbreviated biomarker discovery process flow ..............................35
Figure 15 - ECM functions in development and microenvironmental balance in the body.................................................................................44
Figure 16 - Schematic of connective tissue underlying an epithelium ............47
Figure 17- Proteoglycan and glycoprotein interactions in the extracellular space. ....49
Figure 18- Integrins are transmembrane receptors that bridge between the intracellular and extracellular space.................................................51
Figure 19 – Integrin mediated cell adhesion to the ECM is enhanced through focal adhesions. ..............................................................................52
Figure 20- Growth factors bind to receptors on the surface of the cell............54
Figure 21- MMPs function to alter and shape the basement membrane .........55
Figure 22- Determination of epithelial and stromal content of colorectal samples. 67
Figure 23 - Variation in cellular content between the cancerous breast tissue and the matched normal tissue sample. .................................67
Figure 24 – Modified 384-well plate............................................................68
Figure 25- ds-DNA is heated to 90-95°C to denature the hydrogen bonds that bind the double helical structure and allow for the primers to attach at the specific sequence.

Figure 26- The extension phase.

Figure 27- PCR s-curve showing the accumulation of fluorescent dye.

Figure 28- The QuantStudio 3D digital PCR ® System and BioMark HD® System

Figure 29- Wafergen four quadrant SmartChip®

Figure 30- Publications indexed within databases for the search term “droplet microfluidics”

Figure 31- Schematic illustration droplet formation and unit length.

Figure 32- The GEI (Gene Expression Instrument) design.

Figure 33 -Microfluidic dipping process using PTFE tubing.

Figure 34- a) Single tip (individual sample dipping) and b) Multilumen (bundled reagent dipping) heads

Figure 35- Liquid Bridge Concept for combining samples and reagents and inset, droplet passing between capillaries.

Figure 36- Individual droplets of DNA template, gene assay and polymerase mastermix are combined to form one microreactor

Figure 37- Pumping System

Figure 38-Fluidic manifold system

Figure 39- Sensirion flow sensors s.

Figure 40- Phases of amplification of the DNA target.

Figure 41- 3D model of thermal module.

Figure 42- Temperature control schematic

Figure 43- Description of the Taqman Probe.

Figure 44- Fluorescence excitation and emission spectrum for ROX and FAM.

Figure 45 Illustration of the optical detection platform.

Figure 46 - An example of an s-curve generated by the instrumentation.

Figure 47- - Photograph of complete instrument

Figure 48- Internal components of the Applied Biosystems 7900HT

Figure 49- Standard curves for a number of expression assays.

Figure 50- Determination of efficiency of B2M assay.

Figure 51-HCT116 cDNA vs. β2M expression tests.

Figure 52- Reproducibility test performed on GEI Instrument.
Figure 53-Ct values for PPIA reference gene and RACK1 gene of interest expression.................................................................................................................119
Figure 54- AB7900HT Real-Time PCR amplification plot .................................................120
Figure 55-Contamination tests performed on instrument. ..................................................122
Figure 56- Genetic fold change levels for the patients included in the study. ........126
Figure 57- Gene Co-Expression Network........................................................................132
Figure 58- Expression fold-difference for the gene panel between normal and cancer samples when samples are categorised according to UICC staging. .........133
Figure 59-Potential genetic biomarkers for CRC diagnosis based on UICC staging.. .................................................................134
Figure 60- Association between gene expression and histopathological parameters. ..................................................................................................................................139
Figure 61 -Plate Layout for ABI Instrument. .................................................................140
Figure 62-Fold change for 5 genes (3 upregulated and 2 downregulated) as observed in a parallel experiment.................................................................142
Figure 63- Location of the carcinoma within the upper region of the left breast of the patient.....................................................................................................................143
Figure 64 - Matrix Metalloproteinase differential expression. ........................................146
Figure 65 -Structural Constituents differential expression .................................................147
Figure 66 - Integrin differential expression.......................................................................148
Figure 67 – Growth factor differential expression in ..........................................................150
Figure 68 – Other ECM constituents differential expression........................................152
Figure 69-Well plate layout for reactions for ABI PCR verification study ........155
Figure 70 –Amplification S-curve example comparison......................................................157
Figure 71 – Two platform comparison experiment............................................................158
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB7900HT</td>
<td>Applied BioSystems 7900HT Real-time PCR Instrumentation</td>
</tr>
<tr>
<td>AKT</td>
<td>Alpha threonine-protein kinase</td>
</tr>
<tr>
<td>APC</td>
<td>Adenomatous polyposis coli</td>
</tr>
<tr>
<td>β2M</td>
<td>Beta-2-Microglobulin</td>
</tr>
<tr>
<td>BTA</td>
<td>Bladder tumour antigen</td>
</tr>
<tr>
<td>CA</td>
<td>Cancer antigen</td>
</tr>
<tr>
<td>CCD</td>
<td>Charge coupled device</td>
</tr>
<tr>
<td>CD</td>
<td>Cluster of differentiation</td>
</tr>
<tr>
<td>CD9</td>
<td>CD9 molecule</td>
</tr>
<tr>
<td>CDH2</td>
<td>Cadherin 2</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complimentary deoxyribonucleic acid</td>
</tr>
<tr>
<td>CEA</td>
<td>Carcinoembryonic antigen</td>
</tr>
<tr>
<td>CEACAM1</td>
<td>Carcinoembryonic antigen-related cell adhesion molecule 1</td>
</tr>
<tr>
<td>COL1A1</td>
<td>Collagen, Type 1, Alpha 1</td>
</tr>
<tr>
<td>COL3A1</td>
<td>Collagen, Type 3, Alpha 1</td>
</tr>
<tr>
<td>CRC</td>
<td>Colorectal Cancer</td>
</tr>
<tr>
<td>CTC</td>
<td>Circulating tumour cells</td>
</tr>
<tr>
<td>CXCR2</td>
<td>Chemokine (C-X-C motif) receptor 2</td>
</tr>
<tr>
<td>DEP</td>
<td>Dielectrophoresis</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco's Modified Eagle's medium</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxynucleotide triphosphate</td>
</tr>
<tr>
<td>E</td>
<td>Efficiency</td>
</tr>
<tr>
<td>ECM</td>
<td>Extracellular Matrix</td>
</tr>
<tr>
<td>EGF</td>
<td>Epidermal growth factor</td>
</tr>
<tr>
<td>EGFR</td>
<td>Epidermal growth factor receptor</td>
</tr>
<tr>
<td>ER</td>
<td>Estrogen Receptor</td>
</tr>
<tr>
<td>EWOD</td>
<td>Electrowetting-on-dielectric</td>
</tr>
<tr>
<td>FAM</td>
<td>6-carboxyfluorescein</td>
</tr>
<tr>
<td>FAP</td>
<td>Familial adenomatous polyposis</td>
</tr>
<tr>
<td>FBS</td>
<td>Fetal bovine serum</td>
</tr>
<tr>
<td>FOBT</td>
<td>Faecal occult blood test</td>
</tr>
<tr>
<td>GF</td>
<td>Growth Factor</td>
</tr>
<tr>
<td>GUSB</td>
<td>Glucuronidase Beta</td>
</tr>
<tr>
<td>HE4</td>
<td>Human epididymis secretory protein 4</td>
</tr>
<tr>
<td>HER2</td>
<td>Human epidermal growth factor receptor 2</td>
</tr>
<tr>
<td>hGC</td>
<td>Human chorionic gonadotropin</td>
</tr>
<tr>
<td>HNPCC</td>
<td>Hereditary nonpolyposis colorectal cancer</td>
</tr>
<tr>
<td>HPRT1</td>
<td>Hypoxanthine Phosphoribosyltransferase 1</td>
</tr>
<tr>
<td>IBD</td>
<td>Inflammatory bowel disease</td>
</tr>
<tr>
<td>IGF1</td>
<td>Insulin-like growth factor 1</td>
</tr>
</tbody>
</table>
IGF1-R  Insulin-like growth factor 1 receptor
IGF2  Insulin-like growth factor 2
IGF2-R  Insulin-like growth factor 2 receptor
IGFBP2  Insulin-like growth factor binding protein 2
IGFBP3  Insulin-like growth factor binding protein 3
IGFBP4  Insulin-like growth factor binding protein 4
IL8  Interleukin-8
ITGA2  Integrin alpha 2
ITGA5  Integrin alpha 5
ITGA8  Integrin alpha 8
ITGAV  Integrin alpha type V
ITGB1  Integrin beta 1
ITGB4  Integrin beta 4
ITGB5  Integrin beta 5
ITGBL1  Integrin beta-like 1
KER18  Keratin 18
KRAS  Kirsten rat sarcoma viral oncogene homolog
LAMA1  Laminin alpha 1
LOC  Lab-on-a-chip
MAPK  Mitogen-activated protein kinases
mCRC  Metastatic colorectal cancer
MMP2  Matrix-metalloproteinase 2
MMP3  Matrix-metalloproteinase 3
MMP7  Matrix-metalloproteinase 7
MMP9  Matrix-metalloproteinase 9
mRNA  Messenger-ribonucleic acid
NGS  Next-Generation Sequencing
NMP  Nuclear Matrix Protein
NTC  No-template -control
p63  (Tumour) protein 63
PCR  Polymerase chain reaction
PGK1  Phosphoglycerate Kinase 1
PI3K  Phosphoinositide 3-kinase
POSTN  Periostin
PR  Progesterone Receptor
PPIA  Peptidylprolyl Isomerase A
PSA  Prostate-specific antigen
PTEN  Phosphatase and tensin homolog
PTFE  Polytetrafluoroethylene
PTK2  Protein Tyrosine Kinase 2
qPCR  Quantitative polymerase chain reaction
RACK1  Receptor for activated C-kinase 1
RNA  Ribonucleic acid
ROMA1  Risk for ovarian malignancy algorithm
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROX</td>
<td>6-Carboxyl-X-Rhodamine</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>SPARC</td>
<td>Secreted protein acidic and rich in cysteine</td>
</tr>
<tr>
<td>TGFβ</td>
<td>Transforming growth factor beta 1</td>
</tr>
<tr>
<td>TNBC</td>
<td>Triple-negative breast cancer</td>
</tr>
<tr>
<td>µTAS</td>
<td>Micro total analysis systems</td>
</tr>
<tr>
<td>VEGF</td>
<td>Vascular endothelial growth factor</td>
</tr>
<tr>
<td>VIM</td>
<td>Vimentin</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION
1 Introduction

“Knowledge is not enough; we must apply. Willing is not enough; we must do”

The above quotation by von Goethe encapsulates the predominant goal of cancer epidemiology which is to apply knowledge, gained through scientific discovery, to improve population health. Recent advances in this knowledge field have improved the molecular understanding of cancers and led to the identification of possible therapeutic interventions but ultimately, cancer remains a prodigious threat to global populations. A targeted focus to combat this challenge is the combination of multiple disciplines such as biology, engineering, genomics and chemical sciences into efficient diagnostic technologies that can be used to improve identification, progression and treatment of the disease. Utilising state of the art technologies to improve patient outcomes is a primary goal. It is only with a multi-dimensional, multi-disciplinary approach, that cancer can, one day, become a fully curable disease.

Described in literature as a ‘multi-headed dragon’ (Mäbert et al., 2014), cancer can ultimately be defined as a group of diseases characterised by the uncontrolled growth and spread of abnormal cells (Peedell, 2005) (Figure 1). These cellular malformations are ultimately controlled by genetic factors which can become dysregulated. Cancer initiation can also be associated with specific mutations that affect proto-oncogenes, tumour suppressor genes and DNA repair genes which can lead to unrestrained growth of primary neoplasms (Gout and Huot, 2008). The abnormal cells gain new capabilities, become dominant over normal cells and often reduce the function of the affected organ. Further to this, new blood supplies which can form help these cells propagate more forming a solid tumour and even use the vessels as ‘highways’ to move to secondary sites in the body. These developmental stages can make the targeting of cancer formation, development and progression very challenging for scientists and creates an ever-evolving research environment. Applying a growing knowledge of the biology of the disease can gradually lead to better treatments of the disease (Alberts B, 2002b) and potential discovery of novel treatments and therapies.
In recent years, a wide variety of cancer targets have been identified and brought from benchtop to bedside but this process is extremely slow and challenging. Novel biomarkers need to undergo years of development, tests and clinical trials before reaching a stage where they may be used in human treatment. Despite this, many ultimately turn out to be of no medical use, because they do not work in humans, have bad side effects or are simply too difficult to implement (Alberts B, 2002b). To address this challenge, researchers need to identify novel areas which expedite experiments and develop new translational tools which allow for better research outcomes. New technologies can help in this regards and the development of new methods to do biological research more efficiently with less operator involvement can be key to this. Advances in computational, analytical and optical fields could help to decipher the genetic changes which are occurring in people who develop cancer but ultimately, fundamental variations of disease states need to be fully understood (Dalton and Friend, 2006). Research undertaken in this thesis, aims to build on advances in microfluidic and genetics to develop medical diagnostic

**Figure 1** –Cancer cells differentiate and proliferate at an increased rate in comparison to normal cells. Their potential to invade through basement membrane and mucosa is increased and can develop into a benign mass of dysregulated cells which can form into a tumour shown in the above image (Cancer Research UK, 2014b). Preventing and/or controlling the growth of this tumour is key in cancer management strategies.
introduction for the analysis and classification of cancers and other genetic disorders.

**Molecular Biology of the Cell**

The body consists of trillions of living cells, which are constantly growing, dividing to make new cells and dying in a controlled manner. The cell cycle is a very finely tuned process and responds to the specific needs of any specific tissue or cell (Hartwell and Kastan, 1994). Each cell in the body contains the exact same set of approximately 20,000-25,000 genes (International Human Genome Sequencing Consortium, 2004) and these genes control how cells in the body work by encoding for proteins which dictate cell function (Venter et al., 2001).

Within each nucleated cell, discrete sections of genetic information called genes are carried by deoxyribonucleic acid (DNA). DNA is the “blueprint” of each individual and is made up of a double-helical spiral of chemical bases known as nucleotides (Nussbaum et al., 2015). The nucleotides consist of nucleobases (Figure 2) of either adenine (A), cytosine (C), guanine (G) and thymine (T) in addition to a monosaccharide sugar called deoxyribose, and a phosphate group which fuse to form a sugar-phosphate backbone (Puri, 2014). To get from DNA to a functional protein,
which becomes the phenotypic product of the cellular genetic response, the DNA is then transcribed into messenger ribonucleic acid (mRNA) and further translated into a protein for cellular use (U.S National Library of Medicine, 2016). RNA is synthesised in the nucleus and is very similar to DNA except that the sugar is ribose and a thymine base is replaced by a uracil base (Egli and Saenger, 2013). The actual information in the DNA of chromosomes is decoded in a process called transcription. During the transcription process (where the DNA transcribed into RNA), the double helical strand of DNA is temporarily split by an enzyme into two individual strands (Biotopics, 2016). The free bases of the DNA allow for complimentary RNA to attach.

Once the RNA nucleotides have paired with the free DNA strand, the newly synthesised mRNA moves from the cell nucleus to the cytoplasm where it attaches to a small subunit of a cellular structure called the ribosome. It is at this point that the translation process (RNA translated into a protein) begins. Protein function is defined by their molecular function, localisation within cell and involvement in a particular biological process (ProteinSynthesis.org, 2016). The production of a protein begins when a cell receives a signal or multiple signals from either an internal or external stimuli. The signals can be physiochemical or mechanical in

Figure 3- Central Dogma of Molecular Biology- through the production of mRNA (transcription) and the synthesis of proteins (translation), the information contained in DNA is expressed.
nature and lead to diverse biological processes (Shivashankar, 2011).

During translation, one amino acid is added to the protein strand for every three bases of the RNA, as opposed to a one-one DNA-RNA nucleotide pairing relationship during transcription. With the activation of transfer RNA (tRNA), the first amino acid for the protein synthesis is brought to the ribosome where the tRNA also reads the triplet code in the mRNA (Clancy and Brown, 2008). The tRNA binds to the mRNA strand when it’s three nucleotides match the first three coding bases on the mRNA sequence. This process of tRNA transportation of amino acids repeats itself and coupled with ribosomal activity and cellular energy, the amino acids are bound to each other to form a polypeptide chain of amino acids which eventually constitutes a complete three dimensional, functioning structure called a protein.

**Figure 4** - mRNA is synthesised within the nucleus of the cell and moves into the cytoplasm. Here, it attaches to the ribosome where tRNA brings nucleotide bases with an attached amino acid. As this cycle continues, the amino acids form a sequenced chain known as a protein.

The protein production process is tightly managed by regulatory control points where interactive pathways are formed between the nucleus and the external environment. However, if mechanisms in parts of the cell begin to malfunction, the cell undergoes a series of transformations that allow it to begin a transition from a normal cell to one which is abnormal. These transformations include but are not
limited to the development of an abnormal phenotype, irregular differentiation, increased rates of proliferation and can lead to the development of a dysplastic environment by creating irregular replicates. In a normal cell, if the DNA is damaged, the cellular machinery either repairs the DNA or the entire cell undergoes apoptosis, or programmable cell death, but a cancer cell overrides this function and the damaged DNA is not repaired and therefore the cell keeps replicating with the same damaged DNA. The replication process can be expedited in cancer cells giving it a further advantage over its normal counterpart. Combining these transformations allows these cells obtain a superior advantage over the normal cells and in most cases, lead to the formation of an abnormal cellular environment which, if left untreated, may develop into a tumour.

**Microenvironmental Influence**

Coupled with genetic alterations which occur in the DNA of the cell, the microenvironment surrounding the tumour is also inherently linked to how a tumour develops and progresses (Gout and Huot, 2008, Comen, 2012, Li et al., 2016). A major component of this niche is the extracellular matrix (ECM) which is a network of macromolecules with distinctive physical, biochemical and biomechanical properties (Lu et al., 2012, Ozbek et al., 2010). The diverse function of the ECM promotes an environment of tissue regularity which is synonymous with normal tissue order from early embryonic development through to adulthood. Our knowledge of these functions, the composition and overall structure of the ECM is constantly evolving from a previous hypothesis of being a dormant supportive constituent for cells, to having a dynamic and versatile influence in the fundamental aspects of cell biology (Lu et al., 2011, Hynes, 2009). It is this dynamic-reciprocity theory which was first proposed in 1982 (Bissell et al., 1982) which highlighted prominent features of the ECM and its role in directing gene expression via transmembrane proteins and cytoskeletal components. This concept rejected the hypothesis that the ECM was simply an inert supportive scaffold role with limited roles in cellular behaviour or tumour development. ECM composition and dynamics became a key focus of research groups and key fundamental roles of multiple regulatory systems within the niche were outlined (Bissell and Aggeler, 1987, Weaver et al., 1996, Lochter and Bissell, 1995) which highlighted the ability of the
ECM to provide the contextual information responsible for controlling both individual and collective cellular behaviour (Cox and Erler, 2011).

![Diagram of ECM](image)

**Figure 5-** Key constituents of the ECM. The ECM plays a key role in the microenvironment of both normal and abnormal cells. The versatile functions of the ECM are dependent on its diverse physical and biochemical components which create a dynamic environment of communication between the inside and outside of the cell. (Punjabia et al., 2014.)

At a fundamental level, as cells are responsible for the production of ECM they are constantly realigning and remodelling its components to change the ECM composition and topography (Lu et al, 2011). Conversely, ECM modifications which occur because of cellular influences on the organisation of the ECM, in turn, stimulate the behaviour of adjacent cells to adapt to the altered state (Butcher et al., 2009). Cellular transformations which occur inside the cell are transformed through
the cytoplasm and induce a micro-environmental reaction such as synthesis, degradation or remodelling of the ECM, which in-turn contributes to an increase in cell activity by increasing gene regulation (Figure 5). This mechano-transduction of a physical ECM force to a cytoskeletal-dependant cell generated response contributes to an environment which is ever changing (Bissell and Radisky, 2001). The structural architecture in which cells reside is also regulated by key ECM protein interactions. The diverse processes by which these interactions occur, dictate tissue specific cell behaviour. It is the disruption of the ECM dynamics that become the hallmark of cancer and key regulatory functions that, in normal tissue are tightly controlled, become dysregulated. The main contributors of altered ECM metabolism are stromal cells and immune cells. Subtle changes in ECM characteristics and dysregulation of its components can lead to development of disease.

Fibrous proteins, proteoglycans and glycoproteins are main matrix components that assemble into three-dimensional structures which form the basement membrane for mammalian organ structure and function (Schaefer and Schaefer, 2010a, Järveläinen et al., 2009). Fibrous proteins such as collagen and elastin compose a large area of the ECM and give structural strength to the matrix (Kim et al., 2011). The collagen family contains 28 members and they can be divided into several subgroups based on their structural and functional properties (Heino, 2007). Interspersed between these collagen fibres are proteoglycans composed of one or more glycosaminoglycan (GAG) chains attached to a core protein (Hynes and Naba, 2012). These function to regulate the movement of molecules through the matrix. The third major component of the ECM is glycoproteins which there are approximately 200 of in the mammalian matrisome (Naba et al., 2012). The most commonly studied glycoproteins are laminin and fibronectin. These proteins interact with cells through crucial connections that in turn regulate cell mediated processes such as adhesion, differentiation and migration.

In addition to the structural support and mechanical influence on cells, the ECM actively participates in the development and maintenance of the tissue structure through the regulation of the abundance of integrins, growth factors and proteases on the cell surface which further propagate downstream signalling activities (Hynes, 2009). Furthermore, mutations in genes which code for functionality, architecture and composition of the ECM can greatly influence cancer initiation and progression.
Increased deposition of structural fibril networks such as collagen and laminin by mechanisms alter the natural balance of the environment and affect the delicate balance between stability and remodelling of the ECM (Lu et al., 2011). Conversely to this, reduced deposition and increased ECM breakdown by proteases can cause tissue destruction which can increase the migratory capability of the cancer cell. The structural fibrils of the matrix can also expedite this process by linearising and stiffening cross-linked collagen fibre bundles to provide an aligned migratory tract for the cell which potentiates cell migration (Lu et al., 2012). These mechanisms of microenvironmental influence will be expanded on in later chapters.

**Biomarkers**

By definition, a biomarker is a measurable, reliable indicator used to assess a disease process or an outcome, or to estimate whether a drug used in the treatment was effective or not (Biomarkers Definitions Working Group, 2001). Miniaturisation, automated and highly specific technologies can set the stage for robust advancements in biomarker discovery research. Since the emergence of novel technologies, or -omic technologies, thousands of putative biomarkers have been identified and published, which have dramatically increased the opportunities for developing more effective therapeutics (Drucker and Krapfenbauer, 2013). Cancer biomarkers can be cells (circulating tumour cells), DNA, (e.g single-nucleotide polymorphisms (SNPs), copy number variations) mRNA (over/under expressed transcripts, miRNAs), proteins (cell surface receptors, tumour antigens), peptides (released by tumours into urine or other bodily fluids) or metabolites (compounds such as cholesterol and tyrosine (Hayes et al., 1996, Hassanein et al., 2012, Ludwig and Weinstein, 2005b, Harris, 2003) and can be produced either by the tumour itself or by other tissues (Kulasingam and Diamandis, 2008). They are usually classified into three categories: prognostic, predictive, and pharmacodynamics (Madu and Lu, 2010); based on the ability to predict the progression of the disease and ultimately the patients outcome, the ability to predict a patient's response to treatment and the treatment effects of a drug on a tumour. Biomarkers can help to improve the efficacy and safety of cancer treatment by enabling physicians to tailor treatment for individual patients (Madu and Lu, 2010). However, there is still a large gap between initial biomarker discovery studies and their clinical translation due to the challenges
in the process of cancer biomarker development (Goossens et al., 2015). In the last two decades alone, fewer than 12 biomarkers have been approved by the US Food and Drug Administration (FDA) for monitoring response, surveillance or recurrence of cancer (Anderson and Anderson, 2002). This failure in finding highly-sensitive and high-specificity biomarkers may be attributed to the following factors (Issaq et al., 2011):

- Small number of samples that are analysed;
- Lack of information on the history of the samples;
- Non-standard operating procedures for sample selection, collection, storage, handling, analysis.

- Data interpretation and validation disparities between research groups

These key obstacles slow the discovery of new biomarkers and prevent researchers from making scientific advances in the field. However, using a more cross-collaborative approach, new technologies and new development techniques, biomarkers have the potential to allow for a more exact and efficient assessment of the disease diagnosis and prognosis and lead to more patient specific therapies.

Outcomes from the discovery of clinically validated biomarkers can have a major influence on several factors of disease diagnosis and prognosis for patients. Biological therapies such as monoclonal antibodies and inhibitors function to destroy or restrict certain types of cellular activities. Primarily these therapies are protein targeted biomarkers which ‘lock’ to certain influential proteins. Examples of these molecularly targeted biomarker therapies are: human epidermal growth factor receptor 2 (HER2)/neu blockade in HER2/neu-positive breast cancer (Sevcikova et al., 2013, Mayer, 2009), estrogen receptor expression targeting in breast cancer (Rochefort et al., 2003), and KRAS mutation and anti-EGFR antibody therapies (Kalia, 2015, Harari, 2004, Paik et al., 2006). HER2 treatment strategy has been validated to be not only a prognostic factor, but also a predictor of response to HER2 targeting therapy (Weigel and Dowsett, 2010). This treatment targets the overexpression of the HER2 gene, leading to increased proliferation of breast cancer cells, by attaching to certain proteins on the surface of particular cancer cell receptors. HER2 overexpression affects approximately 20% of breast cancer patients and the identification of this overexpression in a wide cohort of patients and in a...
wide variety of tissue types has allowed for the discovery of targeted therapy drugs which ultimately can benefit survival rates for patients (McLeod, 2013). Anticancer therapies, like HER2 targeting, take advantage of a property of the cancer cell that distinguishes them from normal cells. This allows for successful outcomes from utilising new technology-driven methods to develop novel strategies to combat the disease. As we become better able to determine which genes are amplified, which are deleted, and which are mutated in the cells of any given tumour, we can begin to tailor treatments more accurately to each individual patient (Alberts B, 2002b).

**Existing Technologies**

It was not until the mid-1970’s that two families of genes- proto-oncogenes and tumour suppressor genes- that normally regulate the natural process of cell growth and death in healthy tissues and organs were identified as playing crucial roles in cancer formation and development (Biemar and Foti, 2013). This realisation of genetic mutations and changes paved the way for therapies to target specific defects. In the past decade, the sheer volume and access to scientific research and health data has allowed for remarkable progress to be seen in cancer research (Abernethy et al., 2014). However, despite advancements in understanding development, progress and treatment of cancers, the disease continues to affect millions of people worldwide (Cancer Research UK, 2016b) and the occurrence rates continue to rise (due, in part, to improved screening for early detection). According the World Health Organisation (WHO) estimates, by the year 2030 cancer deaths could increase globally by as much as 80% (Dizon et al., 2016), which puts further urgency on the need to reduce the gap between initial discovery studies and their clinical translation. To aid this, public gene repositories such as Gene Expression Omnibus, Array Express and CIBEX have been developed to give a pan-genomic appraisal of the transcriptome. However, variations in experimental approaches coupled with diverse workflows in the interpretation and standardisation procedures used in these studies have ensured that a consensus platform has not been established (Hogan, 2012). This therefore, can limit the number of significant advances that can be made in the field and combating the diseases’ advance remains challenging.

Many new technologies have emerged in recent years to expedite various processes within the fields of cancer etiology and pathogenesis. From sample preparation to
improved real time data analysis, the combination of biological and technological advances has allowed researchers to use cross-scientific platforms to become more efficient and precise and are now enabling researchers to investigate living systems on an unprecedented scale by studying genomes, proteomes, or molecular networks in their entirety (Baginsky et al., 2010). The wave of knowledge has been made possible from advancements in the fields of genomics, transcriptomics and proteomics such as next generation sequencing, DNA microarrays, high-throughput PCR, and mass spectrometry.

- DNA microarrays or DNA biochips are widely established research tools to establish RNA levels in cells and tissues. Picomoles of a specific DNA sequence are attached to individual spots on a chip and these spots are used to hybridise to a target sample under high-stringency conditions (Hayes and Kruger, 2014). The technology can be utilised to screen for multiple potential gene expression biomarkers and drug targets and has been used in a wide range of cancer classification studies (Perou et al., 2000, Welsh et al., 2001, Hedenfalk et al., 2001).

- Next-generation sequencing (NGS) is a primary analytical method to give researchers a global view of the transcriptome. NGS quantifies discrete, digital sequencing read counts allowing for sequence-based gene expression where individual genetic transcripts are digitally analysed. The field has rapidly expanded in recent years with the aid of new technologies, rethinking of experimental design and new approaches to data-analysis (Shendure and Ji, 2008). This has seen the process become the leading method to decipher the genetic changes involved in disease formation and progression. Challenges regarding the ability to fully analyse “big-data” outputs however can obstruct its use in many laboratory applications.

- Targeting more specific regions of mRNA levels using high throughput PCR can sometimes be a more appropriate option for researchers to analyse samples for transcriptional biomarkers. Quantitative polymerase reaction (qPCR) is a powerful technique that enables fast, quantitative and reliable results. The target DNA is exponentially amplified repeating the thermal cycling between the temperatures for a number of cycles which can yield thousands of target DNA
molecules that can optically monitored at the end of each temperature cycle to allow quantification of the target DNA.

The comprehensive aim to reduce sample, gene assay and chemistry quantities is motivating researchers to attain increased levels of genetic information that have implications for biomarker discovery. Novel technologies focus on increasing throughput to allow more rapid and efficient processing of samples in the effort to gain a clearer understanding of the molecular dynamics of diseases. Scaling reactions from the microliter to the nanolitre and picoliter range has been a key goal of technological advances. More sensitive detection limits have become achievable with new technologies and have revealed an unprecedented level of information on genetic and epigenetic levels. New instrumentation also focuses on combining entire process flows on single instruments and it is not uncommon for a single high throughput instrument to carry out thousands of experiments per day with minimal involvement for the user. Coupled with these engineering and scientific breakthroughs, advancements in bioinformatic approaches have allowed vital statistical information to be deciphered more easily and efficiently in comparison to traditional methods. By increasing the number of reactions through the use of replicates in a single run, increases the statistical power of an experiment (Fay and Gerow, 2013). However, translating the data into patient relevant information, such as disease marker identification, has remained a challenge that is yet to be overcome. It is only with a multidisciplinary approach utilising a range of diverse technologies that can truly maximise the diagnostic impact of studies of this type.
Objectives

This thesis aims to characterise the use of microfluidic, microscale, droplet-based experimentation methods to utilise the technology to perform biologically relevant experimentation on matched normal and cancerous tissue extracted from patients who are affected by colorectal and breast carcinomas. Genetic targets will be from a family of extracellular matrix genes that play key roles in normal cellular activity and tissue homeostasis. A theoretical design evaluation for the study will be performed to gain an understanding from both a mechanical and biological viewpoint. Key focus will be put on the following objectives:

1. To optimise the design of a microfluidic qPCR platform including the microfluidic network, the devices amplification capabilities and the development and characterisation of the optical detection system.

2. To validate the instrument potential for the determination of RNA levels in biological specimens using the microfluidic qPCR technology which has been designed to be ultra-sensitive, high throughput and reproducible.

3. To identify clinically relevant, RNA biomarkers for the stratification of colorectal carcinoma and identify a panel/or panels of ECM genes which emerge as being most dysregulated in their expression patterns in malignant and matched-normal patient samples.

4. The dysregulation will then be further investigated to determine expression correlations that can be used to develop a consensus profile. It is anticipated that co-expressed genes and correlations with histopathological and clinical parameters will form a key part of this.

5. Genetic targets that emerge from the original subset will be analysed in a validation test on a commercial platform with a goal to validate the gene expression profile generated from the microfluidics platform.

6. Further microfluidic profiling to identify gene expression variations from a separate malignant breast tissue specimen will be performed to examine tumour heterogeneity in differing subsections of the tumour, with validation tests again being performed.
The differential expression pattern which is expected to emerge from the genetic profiling of the samples will highlight the potential for using novel microfluidic methods for biological experimentation. It will also indicate the major implications of dysregulation in these genes and the potential for expanding the profile further to include more ECM genes and more tissue types in larger cohorts. The targets which emerge can then be further validated to possibly be analysed for their potential to be used as clinically relevant biomarkers in carcinoma classification.

**Chapter Closure**

This chapter presented a brief biological background on cancer formation and some of the molecular characteristics and microenvironmental influences which can influence its progression. Current novel technologies which focus on reducing volumes and increasing data output have also been mentioned, revealing some benefits and possible limitations to each technology. Advancing some of these technologies can allow for more precise, targeted analysis of samples and is a key driving factor for this research. Finally, understanding biomarker discovery and what treatments can be developed from biomarker research are highlighted. These objectives have been laid out which summates the motivation for this research. The remainder of this thesis is structured as follows;

**Chapter 2: Literature Review (Cancer and Extracellular Matrix)** discusses colorectal and breast carcinoma, incidence rates, risk factors, target discovery, metastasis, biomarker identification, current molecular tools and extracellular matrices role in cancer development and progression. The key regulatory functions that govern ECM/cellular interactions in the body are outlined including intracellular signalling and their role in cancer progression.

**Chapter 3: Materials and Methods** details sample retrieval processes, sample preparation, reaction chemistries and plate preparation along with selection of stable reference genes and components necessary for the PCR experimentation.

**Chapter 4: High Throughput u Flu qPCR Instrumentation** examines current technologies for the PCR reaction and new and emerging technologies. Following this, the developed microfluidic instrumentation is discussed with the device amplification and detection capabilities examined.
Chapter 5: Instrument Validation describes the validation process to verify the instrument is capable of carrying out high throughput biological experiments and outlines a number of tests which are performed in parallel with a commercial platform.

Chapter 6: Colorectal and Breast Experimental Results chapter discusses gene expression profiling of matched colorectal carcinoma patient samples giving a detailed description of the quantitative PCR experiment performed on the instrument. The results obtained from comparing gene expression profiles between the matched patient samples are presented. Gene expression correlations with histopathological parameters are outlined identifying dysregulated genes which correlate with specific parameters recorded during histological examination. Part two of the chapter explores tumour heterogeneity in triple-negative breast cancer. ECM genetic expression is profiled with a focus on differential expression of certain genes in various sections of the tumour. Validation tests on a commercial platform are performed to further support the findings.

Chapter 7: Conclusions and Recommendations present conclusions from this work from both an engineering and biological viewpoint and provide recommendations for future work in this field of research. References and appendices bookend the thesis.
CHAPTER 2
LITERATURE REVIEW
2 Literature Review

In the previous chapter, the primary objective of this thesis was outlined. This objective is to examine the expression pattern of extracellular matrix genes which may aid the stratification of carcinomas, provide new biomarkers and act as a supporting clinical prognostic indicator. This identification/discovery process will be achieved using microfluidic, miniature instrumentation which will focus on performing experiments in micro-reactor droplets. A large section of this chapter focuses on the biological processes relevant to this research.

The literature is examined with regard to the mechanisms of cancer formation and progression, along with incidence rates and risk factors such as diet, lifestyle and pre-existing conditions. A focus is placed on colorectal and breast cancer as these are the two cancers which are make up the experimental results in this thesis. The metastatic potential of a tumour is described with emphasis on angiogenesis and the increase of invasive tumour cell properties. Furthermore, advancements in biomarker discovery and clinical therapies that have emerged from biomarker studies are also elucidated. Following this, the influence of microenvironment and extracellular interactions in normal tissue homeostasis and cancer is presented along with the dynamicity of signalling molecules and pathways which function to progress cancer from a benign mass to metastatic tumour. From this, the genetic changes relating to the ECM which are most influential in the disease progression are discoursed and the dysregulation which can occur is highlighted. Finally, the selection process for a small panel of ECM genes for a pilot study is then outlined.

2.1 Cancer - An Autonomous and Evolving Disease

Cells are constantly dividing and proliferating to increase their presence in the human body and replace cells which are dying. Normal cells divide to replace damaged cells in a tightly controlled manner regulated by complex interplay between the cells and the local microenvironment. However, if this process malfunctions, an abnormal multiplying of irregular cells can occur and propagate to form systematic abnormalities in the cells and irregular neoplastic growth of tissue. The irregular cells can acquire genetic mutations leading to uncontrollable proliferation coupled with growth into adjacent tissues and spread to distant sites in the body. Coupled
with abnormal cellular phenotype, cancer formation can also be initiated by the activation of proto-oncogenes and/or inactivation of tumour suppressor genes. Proto-oncogenes stem from normal genes that acquire a DNA mutation producing a modified protein. These give rise to oncogenes of which there are currently over 400 currently listed (70 genes associated with germline mutations and 342 genes associated with somatic mutations) (Chial, 2008). In contrast, tumour suppressor genes are normal genes which repair damaged DNA or slow down cell division thus controlling the propagation of defective genes.

![Figure 6](image)

**Figure 6** - The illustration highlights six hallmarks identified during the development of cancer (Hanahan and Weinberg, 2000) and demonstrates that a key set of functional capabilities are acquired through various mechanisms.

In addition to the genetic modifications which can initiate the formation of cancer, other regulatory signals also become aberrantly altered which further disrupt the physiology of the cell from normal to malignant. The signals can activate cell functions allowing it to acquire the capability to move from a previous quiescent
state to an active proliferative state. These cell functions each represent intrinsic modifications of the cell replicative ability, depending in part on successive somatic mutations or epigenetic modifications in the evolving tumours and exemplify the acquired ability of a successful tumour to allow growth and dissemination of the malignant cells (Hanahan and Weinberg, 2000).

2.1.1 Colorectal Cancer

![Image](https://example.com/image.png)

Figure 7- The body’s digestive system showing the small intestine, large intestine (bowel) rectum and anus. Food passes down the oesophagus, into the stomach to be digested by stomach acids, through the small intestine and into the large intestine where water and salts are absorbed before reaching the rectum and anus Image: (NIH Medline Plus, 2009)

Colorectal cancer is cancer of the large intestine or the last part of the digestive tract. The large bowel is made up of five primary sections namely the ascending colon, the transverse colon, the descending colon, the sigmoid colon and the rectum. When cancer occurs in the colon and rectum, combined, they are referred to as colorectal cancer. Like most cancers, colorectal cancer is caused by the abnormal growth of cells which occurs in the lining of the digestive tract. If left untreated, this abnormal cell growth can progress to develop inflammations and greater tissue abnormalities in the mucosal wall. This can lead to a build-up of irregular cells and tissue structures to form polyps which can protrude into the lumen. Polyps can be
described as lobular, fleshy outgrowths of tissue mass from the bowel wall. Over time polyps can become more necrotic and some eventually may mutate from a benign mass to a cancerous one if they are not detected early and removed. In addition to outgrowths, pocketings or pouches in the colonic wall can also occur when abnormal cellular behaviour allows for weaknesses to develop in the bowel wall and crevasses to occur. This is known as diverticular disease or diverticulosis and can further irritate the lining of the colon by inflaming the surrounding tissue. Despite this, the link between diverticulosis and colorectal cancer formation or progression of the disease still remains unclear (Granlund et al 2011).

Globally, colorectal cancer is the third most common cancer in men (746,000 cases, 10.0% of the total number) and the second in women (614,000 cases, 9.2% of the total number) (Ferlay et al., 2014) (Figure 8). Predominantly, it is a disease which is most frequent in developed countries with a western culture such as the United States, Australia and European countries (Boyle and Langman, 2000). In contrast to this, countries such as China, India and Africa have significantly lower levels. In some cases, this difference is 10-fold although under reporting in some of these countries may introduce bias into these figures. In Ireland, almost 2,500 people are diagnosed with colorectal cancer each year. It is the second most common cause

![Figure 8](image-url) - World incidence age-standardised rates for colorectal cancer for both sexes in 2012 (Ferlay et al., 2014). Incidence rates are highest in developed countries and could be influenced by improved detection methods in comparison to developing countries.
of cancer death in Ireland. with the number expected to rise to over 3,500 per year by 2020 (National Cancer Registry of Ireland, 2012). Similar to global patterns, incidence rates are higher for males as opposed to females.

2.1.2 Breast Cancer

Despite progress in breast carcinoma diagnosis and treatment, the molecular factors which dominate its development and progression are poorly understood. These molecular influences are further complicated by variations in cell populations within tumours coupled with diverse expression patterns of critical genes which make the tumour significantly more difficult to treat. However, emerging technologies in genomic research allow for a ‘molecular diagnosis’ to be established which could lead to individual types of therapy based on a patient’s specific gene expression profile.

Breast cancer is the second most common cancer in the world and has the highest mortality of any cancer in women worldwide. It is, by far, the most frequent cancer among women with an estimated 142,980 European women dying from the disease in 2012 (Ferlay et al., 2014). Similar to colorectal cancer, there are numerous factors shown to have a contributory effect towards the increasing number of breast cancer cases including changes in lifestyle habits, increase in sedentary lifestyle, weight gain and obesity (Stewart BW, 2014). However, improved screening rates to aid early detection and more targeted therapies of subtypes has led to an increased
survival rate of 81.8% in Europe (2005-2007) (average 5-year survival rate), up from 78.4%. (1999-2001). Without a fundamental deciphering of the molecular mechanisms influencing this disease, the survival rate for breast cancer would not have improved or indeed may have declined in recent years.

Specifically, breast carcinomas can be classified into three distinct subtypes: luminal, human epidermal growth factor receptor 2 positive (HER2+) and basal-like or triple negative breast cancer (TNBC). Luminal tumours are evaluated for the presence of estrogen and progesterone receptors which respond to hormonal treatment therapies. An upregulation or overexpression of the growth promoting ERBB2 oncogene can lead to a HER2+ classification. This subset group can also be effectively treated using anti-HER2 treatments and have been shown to be a successful way of managing the disease. A TNBC diagnosis makes the cancer more difficult to treat as the ER, PR and HER2 targets cannot be used as biologic targets. Therefore it has become critical to examine certain malignancies in greater detail to identify new targets by correlating numerous expression profiles with that of a corresponding matched normal sample taken from the same patient. Therefore, an increased focus

Figure 10- The first noticeable symptom of breast cancer is typically a lump that feels different from the rest of the breast tissue. The cancer is staged to assist in determining suitable therapies for the patient. A stage II breast cancer is shown where the tumour is more than 2 centimetres, but no more than 5 centimetres across. Image: (Cancer Research UK, 2015)
has been put on genomic analysis of tumours in greater detail. A one size fits all approach which was synonymous with previous analytics of tumour classification but does not substantially identify the overall heterogeneity within the tumour itself. Tumour heterogeneity is described as tumours which have a high degree of morphological and phenotypical variations which can occur between cells within a single tumour (intra-tumour heterogeneity) and variations in the same tumour type in different patients (inter-tumour heterogeneity). Both classifications can lead to different therapeutic responses. To account for this, traditional pathology-driven classifications of tumours have been refined to include genetic and molecular classifications due to a greater understanding of the staggering complexity of tumours. Also, the emergence of new technologies (Liu et al., 2003, Hayes et al., 2015, Kinugasa et al., 2015) which allow for the identification of variations in gene expression and protein expression between subpopulations of tumour cells to be investigated has permitted a more patient specific approach to diagnosis. These variations in subsections of tumours play key roles in fundamental cellular and microenvironment changes that can greatly alter its progression from a benign cellular mass to an aggressive malignancy. Understanding the heterogeneous nature of a tumour and how it contributes to various molecular processes and disease progression is vital for future breast cancer classification and management. Recent studies examining the heterogeneity of breast cancer through comprehensive gene expression profiling (Sørlie et al., 2001) and next generation sequencing (Russnes et al.) has identified significant genetic diversity within tumours. These findings led to a hypothesis that levels of expression of genes which function to regulate cellular and microenvironmental networks (such as ECM genes) may not only be differentially expressed in comparison to a corresponding normal sample but also significantly differentially expressed within different subsections of the tumour itself. Performing analyses on individual tumours may reveal genes which could potentially serve as valid biomarkers for prognosis of the carcinoma and its metastatic potential.

2.1.3 Associated Risks

Cancer in some instances has been described as preventable diseases through increased surveillance and lifestyle changes (Rennert, 2007, Anand et al., 2008).
While there is no specific cause for different types of cancer, certain attributes have been shown to increase a person’s lifetime risk such as age, history of the disease in the immediate family, genetics, diet, physical activity and lifestyle (Marchand et al., 1997, Haggar and Boushey, 2009)(Figure 11).

**Age**

The biggest single risk factor for a person to develop both colorectal and breast cancer is age. The older people get, the more likely it is that abnormal changes will occur in our cells. **Colorectal:** More than 90% of colorectal cancer cases occur in people aged 50 or older (Cancer Research UK, 2016a). The incidence rate is more than 50 times higher in persons aged 60 to 79 (Haggar and Boushey, 2009) with the average age the disease is found is 73. However, over the last twenty years, colorectal cancer rates in people ages 20 to 49 have increased significantly(Cleveland Clinic, 2015) which has become a growing concern although a primary reason for this may be earlier and improved screening of populations. **Breast:** Age is also a strong risk factor for the development of breast cancer. The risk of developing the disease rises in the female population after the age of 40 and is highest in women over 70. In the UK, almost 50% of invasive breast cancers are diagnosed in people over 65.

**Familial/ Genetic**

A familial condition of genetic abnormalities can propagate a person’s risk of developing colorectal and breast cancer. **Colorectal:** Up to 70-80% of CRCs are sporadic with the remaining 20 and 30 per cent of patients at presentation having a
first-degree relative with colorectal cancer (Brandão and Lage, 2015). The association between family history of colorectal cancer in first-degree relatives and risk of developing colon cancer has been well defined and findings have shown that a family history of colorectal cancer is associated with the greatest risk among those diagnosed at age 50 or younger (Slattery et al., 2003). Furthermore, factors such as not having a sigmoidoscopy, diet and smoking significantly associate with cancer risk among those with a family history of colorectal cancer. **Breast:** Women who have a close blood relative with breast cancer have a higher risk of the disease although it is important to note that 8/10 of women who develop the disease do not have a family history of the disease. Only around 5-10% of breast cancers are thought to be the result of genetic mutations passed on from a parent. The most common cause of hereditary breast cancer is an inherited mutation in the BRCA1 and BRCA2 genes which when function normally, prevent the production of proteins which control cell growth. If the BRCA1 and BRCA2 genes are not functioning correctly then abnormal cell growth can occur, leading to cancer diagnosis. Other genetic mutations can be an inherited risk such as TP53 and PTEN mutations but are less frequently identified.

**Diet and Lifestyle**

**Colorectal:** Additional factors such as dietary components, lifestyle, levels of physical activity, sleep, alcohol and tobacco consumption have been shown to greatly influence the risk of developing colorectal cancer (Anand et al., 2008, Correa Lima and Gomes-da-Silva, 2005, Slattery et al., 1999, Slattery, 2004, Terry, 2001, Zhang et al., 2013b). As the colon is a key part of the digestive tract, continuously functioning to process food which we consume, it is therefore understandable that diet is one of the most important factors when it comes to lifestyle changes which we can control. The main function of the colon is the absorption of remaining fluids and salts in the stool before elimination (Barron, 2010) which maintains a hydration balance in the body. The liquids removed are absorbed through the surrounding colonic tissue. Therefore, the contents of the stool are a good pre-indicator to how healthy the colorectal tissue remains. Consuming a diet which is low in fruits, vegetables, fibre, plant rich foods (Chan and Giovannucci, 2010) and vitamin D (Jenab et al., 2010) can all lead to a reduction in the amount of toxins that can be
present in the colon at the digestion stage. In contrary to this, a diet with high quantities of fats (Reddy, 2002), red meat (Santarelli et al., 2008), sugars (Slattery et al., 1997) and salts can all lead to irregularities and increased risk of inflammation in the digestive tract. Decreased levels of physical activity (Hardman, 2001), increased body mass index (BMI) (Murphy et al., 2000) and obesity (Frezza et al., 2006) are also known to elevate levels of gut inflammation which further propagate causal and risk factors of CRC. Furthermore, the increase in CRC incidence and mortality due to alcohol and tobacco use is well known and it has been shown that people that smoke cigarettes and consume large quantities of alcohol are at a significantly increased risk of developing the disease (Martínez et al., 1995). It is these factors which play a major role in a person’s susceptibility to the development of the disease but also are the factors which a person has most control over. Breast: Being overweight or obese after menopause increases breast cancer risk. This increased risk stems from variations in estrogen levels in the body. Estrogen is mainly made in the ovaries in pre-menopausal women and fat tissue makes a reduced amount. However, post menopause, having increased fat tissue can raise estrogen levels and increase the risk of getting breast cancer (American Cancer Society, 2016). Physical activity in the form of exercise can reduce this risk (Steindorf et al., 2013). With regard to diet and the risk of breast cancer, the results have been conflicting although studies have shown that a diet which is low in fat reduces breast cancer risk (Chlebowski et al., 2006).

Diabetes

Colorectal: People with diabetes and insulin resistance may have an increased risk of colon cancer (Peeters et al., 2015). This can be predominantly due to increased levels of insulin in the blood of a person who has diabetes in contrast to normal blood insulin levels (Giovannucci et al., 2010). Insulin is the body’s way of regulating sugar or glucose metabolism in the blood (Saltiel and Kahn, 2001). Therefore, increased levels of insulin and glucose promote cellular survival and growth (Kiely et al., 2002) which potentially could aid a colorectal tumour in the quest to grow. Breast: Breast cancer has been shown to be more common in women with type II diabetes (Diabetes UK, 2015) with research showing that diabetic women
Literature Review

are 20% more likely to develop post-menopausal breast cancer than non-diabetic women (Larsson et al., 2007).

Colorectal Specific Risks

FAP/HNPCC: Two molecular disorders of the disease that can be inherited from close family members are familial adenomatous polyposis (FAP) or hereditary non-polyposis colorectal cancer (HNPCC). FAP and HNPCC syndromes are inherited in an autosomal dominant fashion and account for about 1% and 5-7% of all colorectal cancer cases, respectively (Soravia et al., 1997). These conditions occur due to alterations which occur at a genetic level. FAP arises from inherited mutations in tumour suppressor genes, primarily the adenomatous polyposis coli (APC) gene, which results in the development of hundreds of polyps in the colon (Jasperson KW, 2014). The flawed tumour suppressor genes do not promote the formation of cancer but more-so, prevent the body’s ability to protect against the risk of aged cells becoming cancerous. The polyps develop in the intestinal wall of the colon and can develop into carcinomas if left untreated (Reviews, 2015). Similarly, HNPCC can cause polyps to develop albeit at a lower rate but with a higher risk of becoming cancerous. Defects in DNA mismatch repair (MMR) genes, which fail to make corrections to errors in DNA replication due to microsatellite instability (MSI) (Wheeler et al., 2000), are the hallmark of HNPCC (Jass, 2000). The DNA replication errors accumulate to form abnormal cells (Kheirelseid et al., 2013) and can allow for the development of benign masses and possible development of CRC.

KRAS: Other genetic abnormalities can lead to an increased risk of developing CRC. KRAS mutations are present in approximately 30-40% of CRCs (Andreiev et al., 1998, Gnanasampanthan et al., 2001, Imamura et al., 2012). The KRAS gene acts a molecular on/off switch, encoding for a binding protein downstream of the epidermal growth factor receptor (EGFR) (Phipps et al., 2013). However, when specific mutations in KRAS occur, the resulting KRAS protein can be constitutively active (it can then function independently of upstream growth factor receptor driven signals and remain active) (Kashyap et al., 2014) which directs cells to grow and divide in an uncontrolled way. The altered signalling interferes with the normal development of many organs and can propagate cancer formation in the body. Studies showing its frequent mutation in CRC (Karapetis et al., 2008, Vogelstein
and Kinzler, 2004, Misale et al., 2012) have allowed it to emerge as a predictive tool for responses to certain drug treatments and combination therapies (Lievre et al., 2006, Van Cutsem et al., 2009) such as anti-EGFR and chemotherapy.

**Inflammatory bowel disease (IBD):** Patients with long-standing inflammatory bowel disease have an increased risk of developing colorectal cancer (Triantafillidis et al., 2009, Kim and Chang, 2014). IBD consists primarily of two diseases, ulcerative colitis (UC) which causes inflammation of the mucosa of the colon only (Lasher, 2013) and Crohn disease (CD) which involves an inflammation of the full thickness of the bowel wall and can affect any part of the digestive tract (Gohil and Carramusa, 2014). Both conditions can be causes by an abnormal response by the body's immune system by increasing the number of white blood cells it sends to the lining of the intestines, where they produce chronic inflammation and ulcerations. This can lead to the formation of pre-malignant tissue which can further increase a persons’ risk of developing CRC. Identifying those individuals who could be at increased risk of developing CRC through hereditary conditions and IBD along with early detection are the key clinical management strategies for CRC. Early detection of the disease can greatly increase patient survival with Stage I carcinomas having a five-year survival rate of 90% (Haggar and Boushey, 2009). The survival rate drops to 10% for Stage IV cancers (Rodriguez-Bigas MA, 2003). The current screening options for colorectal cancer are limited to endoscopic procedures such as sigmoidoscopy/colonoscopy where the colon and rectum are physically examined using a lighted scope. More recently, faecal occult blood tests (FOBT) have been employed as an early detection method which is a simplistic test which checks for blood in the faecal matter of high risk patients.

**Breast Specific Risks**

**Gender:** A prominent increase in risk of developing breast cancer is gender. Breast cancer is approximately 100 times more common in women than in men and may be due to the fact of increased levels of estrogen and proestrogen, which can promote breast cancer cell growth (American Cancer Society, 2016). Further female links include age at first childbirth, age at first period, birth control pills and IGF-1 hormone levels before menopause.
**Tissue density:** Breasts are made up of lobules, ducts and fatty and fibrous connective tissue. Dense breasts have less fatty tissue compared to breasts that are not dense. Studies have found that women with dense tissue in more than 60-75% of the breast are at four to six times greater risk of breast cancer than those with no densities (Boyd et al., 1998). Factors which affect breast tissue density are age, menopausal status, pregnancy and genetics (American Cancer Society, 2015a).

2.1.4 Metastasis

![Figure 12](image_url)

**Figure 12**—The process of metastasis is demonstrated in the above simplified image of cancer cells moving through the basement membrane, entering the blood vessel, exiting at a new site and developing a metastatic tumour (Lee and Lim, 2007).

Metastases has been a primary focus of cancer researchers for more than 100 years but it is still the cause of 90% of human cancer deaths (Weigel and Dowsett, 2010). At the early stages of primary tumour expansion, the cells are not invasive or metastatic but this ability appears as a result of further accumulation of genetic alterations in the cells (Yokota, 2000) which can change both physiological and biological characteristics. The process of how a cancer cell metastasises is not yet fully understood but consists of a long series of sequential, interrelated steps including cellular transformation, vascularisation, cellular detachment and extravasation (Fidler, 2003). This stepwise progression makes the treatment of a
malignancy much more difficult when it is found that the cancer has spread to secondary sites such as the liver, lungs, bones and lymph nodes than a localised tumour. Both the circulatory and lymphatic systems can facilitate the metastasis of a cancer. The lymphatic vascular system can often be the first location for where cancer can spread to. The system is composed of a dense network of thin-walled capillaries that drain protein-rich lymph from the extracellular space (Oliver and Detmar, 2002). The lymph fluid is released into the tissue which in turn absorbs any toxins or harmful macromolecules within the tissue and cellular environment (Science Clarified, 2016). This fluid is then filtered in lymph nodes (small, circular collections of immune cells) (Figure 13) which purify the fluid and eradicate the toxins (American Cancer Society, 2015b). Cancer cells in the tissue can be apprehended by the lymph fluid before being transported back to the lymph nodes for elimination. If they are not eradicated here successfully then a secondary tumour can form at the location or the cells can continue to spread through the lymph vessels (Cancer Research UK, 2014a).

Figure 13-Abnormal cells which have developed into a cancer can invade through various tissue and muscle layers in the colonic wall with the aim to reach a more hospitable environment for growth or to reach surrounding lymph nodes and lymphatic vessels. Image Reference: (National Cancer Institute, 2013)

However, before a cancer becomes metastatic, it is possible to treat it successfully or manage it in a controlled way where the growth and invasion properties can be halted or reversed (Klement and Kämmerer, 2011, Priebe et al., 2011, Fujiwara et al., 1994). An enhanced understanding of epigenetic modifications and epithelial-mesenchymal transition of cells has led to some defining changes in the way some
cancers are treated. Furthermore, successfully reaching another location in the body does not guarantee that a metastatic tumour will form as cells can lie dormant at a distant site for many years before they begin to grow again, if at all (National Cancer Institute, 2013). Although the migration of cancer cells is well-orchestrated and not a random process, the identification of the basic cellular and molecular processes that regulate their movement and subsequent arrival and survival at distant sites remain elusive (Amano et al., 2009). Once a cancer does start to metastasize, the location of the primary tumour often determines where the cancer will spread to (American Cancer Society, 2014). Nearby suitable organs and lymph nodes are primary locations for cancer cells that break away from the original location and transported via capillaries. The most common sites for cancer metastasis are the liver, lungs and bone. Because the blood supply is filtered by the liver, it is a very common site for the spread of cancer cells that have developed. Often, there are no symptoms of liver metastases until enough of the liver is affected to alter liver enzymes (Mayer, 2014). Blood tests and scans are used to diagnose spreading of cancer to the liver. Furthermore, the heart pumps blood from the rest of the body through the lungs’ blood vessels before sending it elsewhere making it too a common metastatic site. Following the lung and liver, bone is the third most common site of metastasis in cancer patients (Ibrahim et al., 2013), in particular breast cancer patients. These highly specialised cancer cells have unique characteristics that allow for the invasion of the bone marrow cavity in preparation for progression to a bone metastasis. The specific properties of the cells such as adhesive molecules, angiogenic factors and bone reabsorbing factors enhance tumour growth in the bone (Suva et al., 2011). The dissemination of cancer cells throughout the lining of the abdominal cavity or peritoneum is second only to the liver as a site for colon cancer distant metastasis (Aoyagi et al., 2014). Cancer cells can break off from the main tumour and escape into the abdomen, implanting in the peritoneum on the surface of the organs and tissues that are contained there (Beating Bowel Cancer, 2016) and thus lead to metastatic invasion.

Metastasis of the cancer cell usually involves the following key steps:

- **Invasion** of the cancer cells through nearby normal tissue, ECM, and basement membrane.
• **Angiogenesis** activity occurs stimulating the growth of new blood vessels to obtain a blood supply providing the new metastatic cells with oxygen and nutrients.

• **Intravasation** of the cancer cells into nearby blood and/or lymphatic vessels.

• **Circulation** through the capillaries and survival while withstanding increased haemodynamic pressures.

• **Extravasation** of the carcinoma cell out through the vessel wall and colonisation at the distal site.

As the tumour grows, it demands more space in the body and can grow directly into other bodily structures in a process known as local invasion (Cancer Research UK, 2014b). Softer tissue in the milieu of the tumour is particularly susceptible to this process of invasion, and with basement membrane remodelling and degradation, the tumour is allowed to propagate. The ECM can provide a substrate, as well as a barrier towards the advancing cell body (Friedl and Wolf, 2003) and changes in cell-matrix interaction enable the cells to invade the surrounding stroma (Martin TA, 2013). Cellular motility can also become a defining factor of the tumour at this stage and the ability of a cell to move more, using elongated morphologies (Madsen et al., 2015), to squeeze through cellular junctions (Alexander, 2011) allowing the cancer cells to break off from the primary tumour and independently move through the tissue. Enhanced motility and migratory factors of the tumour cell can coincide with a dysfunctional degradation of the tissue. As it grows and the centre gets further away from local blood vessels it can result in reduced amounts of oxygen and nutrients. This causes demand for more bodily nutrients, all while the cells become more and more dysregulated (Cooper, 2000). To combat this, the tumour sends out signals called angiogenic factors that encourage the formation of new blood vessels where there was no pre-existing ones, known as vasculogenesis, or encourages blood vessel growth from pre-existing vessels, in a process known as **angiogenesis**. New blood vessels will nourish the growth of a tumour and increase the probability that additional harmful mutations will occur (Lodish H, 2000a) but smaller tumours have equal metastatic potential despite their size. This renewed blood supply becomes the sustenance that the tumour needs to continue to grow. Nutrients are transported to
and waste is removed from the tumour via these new vessels allowing for further proliferation and growth (Ziyad and Iruela-Arispe, 2011). A cancer cell actively moving towards blood or lymphatic vessels has been suggested to be in response to nutrient or chemokine gradients (van Zijl et al., 2011) though early cancers simply may have pre-established vasculature nearby already that supports the process. Entry of tumour cells into the circulatory system is a critical step in cancer metastasis. To facilitate the process, cancer cells may attached to a vessel wall at a desirable location and invade through the wall in a process known as **intravasation** (Reymond et al., 2013). Cancer cells entering the blood stream obey a similar guise to cells entering the lymphatic system, where they can masquerade themselves and become known as **circulating** tumour cells or CTC’s (Plaks et al., 2013). Once this potential to break away from the primary tumour has been established the newly formed vasculature and lymphatic channels provide a highway for spread throughout the body to their new destination (i.e. a secondary favourable site). Once in the circulatory system, they must endure the haemodynamic stresses, evade the natural host immune response, and arrest at the secondary site (Bockhorn et al., 2007) through **extravasation** (exiting the capillary) to establish themselves in the hospitable secondary site. Concluding all these steps means that the cells can begin to proliferate again and form new tumours. Additionally, having executed all these evasive and agile processes, the cells may not be the same as the cells which originated in the primary site which can sometimes make them more difficult to treat (American Cancer Socieity, 2014) increasing the challenges posed by metastatic cancers.

### 2.1.5 Biomarker Discovery

Innovation in the field of biomarker research is effectuated by a decrease in the rate of deaths from the disease and supported by an increase in early screening programs which allow for earlier disease intervention. Additionally, increases in knowledge of the molecular etiology of cancer have led to profound advances in understanding the development and prevention of the disease (Gwynne, 2001). Underpinning the physiological alterations in the normal genetic mechanisms of the cell, which modify key functional components such as gene expression, cell motility, cell-cell adhesion and growth factor signalling, have enormous potential to improve human health.
These alterations are identified in the form of biomarkers which are defined as “any substance, structure or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease” (World Health Organisation, 2001).

A biomarker of disease, such as cancer, can be categorised as a *prognostic marker* where the aim is to objectively evaluate the patient’s overall outcome, such as the probability of cancer recurrence after standard treatment or as a *predictive marker*, which aims to objectively evaluate the likelihood of benefit from a specific clinical intervention, or the differential outcomes of two or more interventions (Mehta et al., 2010). Biomarker research has advanced dramatically within the past decade as a result of developments and improvements in genomic technologies but paradoxically this has not been translated to the clinic where only a small number of markers are incorporated. However, the advantages of biomarker discovery, coupled with early screening diagnoses, are evident by the increase in cancer survival rates globally (Hashim et al., 2016). In addition, due to increased early screening programs, the death rate has further declined and allowed for earlier disease intervention. At the forefront of this innovation are advanced molecular techniques such as gene expression profiling. Gene expression profiling is a commonly used method in the quest to find genetic dysregulation and potential biomarkers. It can be described as the process of taking a snapshot of the activity of a number of genes at once as the transcriptional activity occurs in bursts (Evanko, 2006). It is a key method for determining how the molecular hallmarks of cancer or other disease may progress.

![Figure 14- Abbreviated biomarker discovery process flow](image_url)
Genetic activity in the body is a tightly regulated process where genes respond to biochemical, mechanical or physical environmental stimuli. The process of identification of genes that are dysregulated can allow researchers to build on past physiological understanding of carcinomas and expand this to include the molecular mechanisms that influence cancers. Furthermore, 21st century technological advances, such as sequencing, allow for the ability to identify deeper mutations linked to different forms of cancers and to identify genetic transcript variants that increase a person’s risk of developing cancer. This development has led to the progressive screening of a greater amount of biomolecules that may play a significant role in cancer. The increased throughput has allowed for the identification of hundreds of potential biomarkers, which can then be advanced further to determine the effect of their substance in patient care.

Disease-associated biomarkers are consistently described in scientific literature such as biomarkers for cardiovascular disease (Vasan, 2006), neurodegenerative diseases such as Alzheimer’s (Humpel, 2011) and Parkinson’s (Sharma et al., 2013), and cancers such as breast (Weigel and Dowsett, 2010, Yadav et al., 2015), prostate (Crawford et al., 2014) and colorectal (Newton et al., 2012, Langan et al., 2013). Yet, because of limitations in validation and regulatory boundaries which exist, there is still a large gap between initial biomarker discovery studies and their clinical translation and utilisation by oncology practitioners. Relatively few cancer biomarkers have successfully advanced beyond the discovery phase to become clinical diagnostic biomarkers. Currently less than 30 are approved by the Food and Drug Administration (FDA)(Wagner and Srivastava, 2012), seen in Table 1 of which nine are protein biomarkers identifiable in the blood (Ludwig and Weinstein, 2005b).
<table>
<thead>
<tr>
<th>Biomarker</th>
<th>Gene Name</th>
<th>Type</th>
<th>Cancer</th>
<th>Source</th>
<th>Clinical Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEA</td>
<td>PSG2</td>
<td>Protein</td>
<td>Colorectal</td>
<td>Serum</td>
<td>Management and prognosis</td>
</tr>
<tr>
<td>EGFR</td>
<td>EGFR</td>
<td>Protein</td>
<td>Colorectal</td>
<td>Colon</td>
<td>Therapy Selection</td>
</tr>
<tr>
<td>Faecal occult blood</td>
<td></td>
<td>Haemoglobin</td>
<td>Colorectal</td>
<td>Faeces</td>
<td>Detection of faecal occult blood</td>
</tr>
<tr>
<td>CD117</td>
<td>KIT</td>
<td>Protein</td>
<td>GIST</td>
<td>FFPE tissue</td>
<td>Diagnosis &amp; Therapy Selection</td>
</tr>
<tr>
<td>Estrogen Receptor</td>
<td>ESR1</td>
<td>Protein</td>
<td>Breast</td>
<td>FFPE tissue</td>
<td>Hormone treatment suitability</td>
</tr>
<tr>
<td>Progesterone Receptor</td>
<td>PSG2</td>
<td>Protein</td>
<td>Breast</td>
<td>FFPE tissue</td>
<td>Hormone treatment suitability</td>
</tr>
<tr>
<td>CTCs</td>
<td>(CD45,cytokeratins)</td>
<td></td>
<td>Breast</td>
<td>Whole blood</td>
<td>Progression and Survival</td>
</tr>
<tr>
<td>CA15-3</td>
<td>MUC1</td>
<td>Glycoprotein</td>
<td>Breast</td>
<td>Serum</td>
<td>Monitoring therapy response</td>
</tr>
<tr>
<td>CA27.29</td>
<td>MUC1</td>
<td>Glycoprotein</td>
<td>Breast</td>
<td>Serum</td>
<td>Monitoring therapy response</td>
</tr>
<tr>
<td>HER2/NEU</td>
<td>ERBB2</td>
<td>Protein</td>
<td>Breast</td>
<td>FFPE tissue</td>
<td>Prognosis &amp; Therapy Selection</td>
</tr>
<tr>
<td>p63</td>
<td>TP63</td>
<td>Protein</td>
<td>Prostate</td>
<td>FFPE tissue</td>
<td>Aid in differential diagnosis</td>
</tr>
<tr>
<td>PSA(total)</td>
<td>KLK3</td>
<td>Protein</td>
<td>Prostate</td>
<td>Serum</td>
<td>Screening and Monitoring</td>
</tr>
<tr>
<td>Pro2PSA</td>
<td>KLK3</td>
<td>Protein</td>
<td>Prostate</td>
<td>Serum</td>
<td>Discriminating benign from cancer</td>
</tr>
<tr>
<td>Free PSA(%)</td>
<td>KLK3</td>
<td>Protein</td>
<td>Prostate</td>
<td>Serum</td>
<td>Discriminating benign from cancer</td>
</tr>
<tr>
<td>Thyroglobulin</td>
<td>TG</td>
<td>Protein</td>
<td>Thyroid</td>
<td>Serum</td>
<td>Aid in monitoring</td>
</tr>
<tr>
<td>NMP22</td>
<td></td>
<td>Protein</td>
<td>Bladder</td>
<td>Urine</td>
<td>Diagnosis and monitoring</td>
</tr>
<tr>
<td>Fibrin/FDP</td>
<td></td>
<td>Protein</td>
<td>Bladder</td>
<td>Urine</td>
<td>Monitoring</td>
</tr>
<tr>
<td>BTA</td>
<td></td>
<td>Protein</td>
<td>Bladder</td>
<td>Urine</td>
<td>Monitoring</td>
</tr>
<tr>
<td>Alpha-Fetoprotein</td>
<td>AFP</td>
<td>Glycoprotein</td>
<td>Testicular</td>
<td>Serum</td>
<td>Staging &amp; management</td>
</tr>
<tr>
<td>Beta-hGC</td>
<td>CGB</td>
<td>Protein</td>
<td>Testicular</td>
<td>Serum</td>
<td>Diagnosis &amp; Staging</td>
</tr>
<tr>
<td>CA19-9</td>
<td></td>
<td>Carbohydrate</td>
<td>Pancreatic</td>
<td>Serum</td>
<td>Monitoring</td>
</tr>
<tr>
<td>CA125</td>
<td>MUC16</td>
<td>Glycoprotein</td>
<td>Ovarian</td>
<td>Serum</td>
<td>Monitoring</td>
</tr>
</tbody>
</table>
Table 1 - List of FDA-approved tumour markers currently used in clinical practice adapted from (Füzéry et al., 2013, Ludwig and Weinstein, 2005a, Pavlou et al., 2013).

<table>
<thead>
<tr>
<th>HE4</th>
<th>WFDC2</th>
<th>Protein</th>
<th>Ovarian</th>
<th>Serum</th>
<th>Reoccurrence and progression</th>
</tr>
</thead>
<tbody>
<tr>
<td>OVA1</td>
<td></td>
<td>Multiple proteins</td>
<td>Ovarian</td>
<td>Blood</td>
<td>Prediction of malignancy</td>
</tr>
<tr>
<td>ROMA1</td>
<td></td>
<td>(HE4+CA-125)</td>
<td>Ovarian</td>
<td>Serum</td>
<td>Prediction of malignancy</td>
</tr>
<tr>
<td>Pap Smear</td>
<td></td>
<td>Cervical Smear</td>
<td>Cervical</td>
<td>Cervix</td>
<td>Screening</td>
</tr>
</tbody>
</table>

Ideally, biomarkers of cancer are blood-based molecules that indicate malignant properties, allowing for minimally invasive testing. Advances in biomarker discovery are focused on all cancers but due to increasing occurrence rates, carcinoma of the gastrointestinal tract and breast are of key focus in modern biomarker studies. Two prognostic biomarkers and one predictive biomarker are described here indicating how individual targets can become key tools in combating diseases such as cancer.

**Carcinoembryonic Antigen**: Elevated preoperative CEA levels in resectable colorectal cancer are associated with poor prognosis (Mehta et al., 2010, Wolmark et al., 1984)

An example of a prognostic biomarker for cancer medicine that has been validated and approved by FDA is the evaluation of CEA levels which has been studied extensively for patients with gastrointestinal cancers, especially in colorectal malignancy (Fakih and Padmanabhan, 2006, Duffy, 2001, Macdonald, 1999). CEA is a glycoprotein molecule which is present in normal mucosal cells but increased amounts have been shown to be associated with adenocarcinoma, especially colorectal cancer (Perkins et al., 2003). Patient CEA levels are measured pre-surgical intervention and checked again post-operatively. Monitoring of CEA levels are most beneficial in post-operative follow-up of patients where elevated levels may indicate
reduced disease-free survival and potential disease reoccurrence (Polat et al., 2014). However, CEA level measurement is not an effective screening test as levels are not elevated in all cases of the cancer and therefore CEA detection is not beneficial in all patients. Normal levels of CEA are deemed to be in the 2.5-5µg per litre of blood however, factors such as smoking may increase this causing false positives. Levels exceeding 10µg/L are rarely due to benign disease or the moderate elevation that may occur due to smoking (Fakih and Padmanabhan, 2006). Generally, levels >100µg/L signify metastatic disease. This makes CEA an effective prognostic marker for detecting the risk of disease reoccurrence but due to specificity shortfalls, further tests are needed to support the findings.

**Epidermal Growth Factor Receptor (EGFR):** EGFR gene amplification appears to be a predictive factor for response to anti-EGFR antibody treatment in CRC (Amado et al., 2008)

An example of a predictive biomarker is EGFR which has been studied extensively for patients with colorectal cancer (Markman et al., 2010, Spano et al., 2005a). EGFR is a member of ErbB family of receptors that exists on the cell surface and activated by binding to specific ligands including EGF and TGFα. Like CEA, EGFR is expressed by normal cells but is recognised as playing a significant role in colorectal cancer initiation and during tumour progression (Tong et al., 1998) where it appears in higher amounts in malignant zones of colorectal cancer specimens than in the surrounding mucosa (Messa et al., 1998). Elevated EGFR expression has been documented in 60% to 80% of patients with mCRC (Cunningham et al., 2004). It has also been shown that an increased expression of EGFR correlated with a more aggressive disease an increased metastatic risk(Gross et al., 1991). The identification of EGFR as a potential biomarker in numerous studies led to EGFR therapeutic targeting (anti-EGFR) with the development of Cetuximab and Panitumumab. These therapies targeting EGFR use monoclonal antibodies to bind to the extracellular domain of EGFR, preventing ligand binding and interrupting the signalling cascades downstream (Vokes and Chu, 2006) and thus effecting all aspects of carcinogenesis, including cell growth and invasion, angiogenesis, and metastasis (Citri and Yarden, 2006). The addition these treatments to the management plan of a patient as a combination therapy, in conjunction with chemotherapy, indicated a progression-free
survival benefit to the patient by blocking EGFR (Kloth et al., 2010). However, like with CEA level monitoring as a molecular marker, the marker has potential drawbacks which does not make it suitable as a prognosis tool for all EGFR-overexpressing patients. Clinical indications showed overall patient survival of patients was increased only a subset of patients (Khambata-Ford et al., 2007) which led to further evaluation of the efficacy of Cetuximab. On further analysis, the benefit of disease-free progression was only found in patients with wild-type Kirsten-RAS (KRAS) with no mutations (Van Cutsem et al., 2009). Mutations in the KRAS gene led to the subsequent activation of EGFR receptors which annulled any earlier intervention to block EGFR activity with Cetuximab further highlighting the challenges with biomarker translation. However, this has progressed to the development of KRAS as a potential biomarker where patients without any KRAS mutation in the primary tumour were noted to potentially benefit from anti-EGFR therapy (Soulières et al., 2010).

HER2/neu: Patients with HER2/neu-positive breast tumours are more aggressive and have a worse prognosis compared to HER2/neu-negative tumours (Mass et al., 2005)

HER2 (Human Epidermal Growth factor receptor 2) protein is further example of a predictive biomarker used in the treatment of cancer. It is a member of the ErbB family of receptors, like EGFR mentioned previously, and functions as a receptor that can interact with a multitude of signalling molecules. In normal cells, few HER2 molecules exist at the cell surface (Rubin and Yarden, 2001) but breast cancers cells can have up to 25–50 copies of the HER2 gene, and up to 40–100-fold increase in HER2 protein resulting in 2 million receptors expressed at the tumour cell surface (Kallioniemi et al., 1992). The DNA-based protein synthesised by the HER2 gene promotes cellular growth and proliferation. It and is over-expressed or amplified in approximately 20-30% of invasive breast cancers diagnosed (Freudenberg et al., 2009). Tumours that are established as being HER2 positive have an overall poorer prognosis, more aggressive disease (Cobleigh et al., 1999), and the risk of reoccurrence is also much greater than in patients diagnosed as being HER2 negative (Salomon et al., 1995). HER2 levels are usually analysed after a sample of the breast tissue is removed (biopsy) but also may be done once the surgery has been
completed. Knowing if a cancer is HER2 positive or negative assists in deciding which treatments are most suitable. Patients diagnosed with HER2 positive breast cancer have specific treatments available for them which block the HER2 receptors on the cell surface and thus slow or stop the growth of a tumour. Signalling pathways associated with the protein are also key targets and alert the body’s immune system to destroy cancer cells to which it is attached. It has been shown that when the protein is targeted and taken in combination with chemotherapy, that the patients' risk of recurrence of the disease is dramatically reduced (Orphanos and Kountourakis, 2012). The development of immune-targeted agents which originated from the identification of the HER2 protein as a potential molecular marker for the treatment of breast cancer promises prolonged survival and better quality of life for cancer patients.

2.1.6 Genetic signatures

The identification of a single genetic marker, as explained in above text, often leads to spotlight being turned on other genes in the same family or on genes upstream or downstream which may have a role in activation of a pathway influencing the disease. Furthermore, owing to new technologies and higher instrumentation throughputs, larger studies of genes in larger patient cohorts can be performed leading to the identification of a multi-gene “signature” assay panel that can guide clinicians in deciding a treatment that may best benefit a patient. Genetic signatures can complement classic prognostic factors to obtain more accurate patient-specific information by measuring a patients’ molecular dysregulation in a small subset of genes. The most commonly used multi-gene signature panel developed for colorectal cancer is the Oncotype DX Colon Cancer Assay® which has been developed by Genomic Health®. Consisting of a 12-multigene signature, the panel assesses a stage II colorectal cancer patient’s risk of the disease reoccurring post-surgical intervention to remove the tumour (Gray et al., 2011) and generates a recurrence score (RS). The panel consists of seven cancer related genes which have been selected from four major, independent, global clinical studies (O’Connell et al., 2010) and have been shown to be significantly associated with disease recurrence in patients. A total of 761 genes and 1,851 stage II and stage III tumour samples were assessed in total with independent validation allowed for the refinement of genes to be selected (Clark-Langone et al., 2007). Seven genes which include 3 cell
cycle genes (Ki-67, MYBL2, and c-MYC), 3 stromal genes (BGN, INHBA and FAP), and an early response gene (GADD45B), emerged as being statistically relevant for likelihood of reoccurrence after three years. Cell cycle and stromal activity regulation are key biological pathways and were identified in the developmental studies as being crucial in the risk of recurrence of the patients. The five other genes are classified as reference normalisation genes to which the expression levels are normalised. These genes are ATPSE, PGK1, GXP1, UBB and VDAC2. The result from the assay is a Recurrence Score value from 0 to 100 that corresponds to the likelihood that a patient’s cancer will reoccur three years post-surgery. The higher the RS which is obtained from the expression analysis will correspond to an increased risk of that patient’s cancer reoccurring. The signature panel provides additional supporting information for clinicians to support current risk assessment methodologies such as tumour stage, grade, lymph node involvement, and lymphovascular invasion.

Combining molecular signature panels such as mentioned, together with traditional clinicopathological grading methods that already exist to provide critical prognostic information, is conceived as the most forward-thinking method for the future of cancer development and outcome classification. The molecular basis of a tumour has evolved significantly over the last decade and has allowed for a greater understanding of carcinogenic disease development, progression and treatment. Publications with the joint subheadings of ‘neoplasm’ and ‘prognostic marker’ indexed in PubMed since 2006 have surpassed 26,000 and over 17,500 publications with ‘neoplasm’ and ‘predictive marker’. An expanded understanding of gene expression levels, DNA sequence alterations and protein function and structure has led to increased tumour biomarker discovery. This understanding, coupled with new technological advances such as DNA sequencing and reducing standard reaction volumes to the nanoscale, have greatly progressed the overall understanding of cancer mechanisms and allowed for new therapies to be developed. However, with over 14.1 million new cancer cases and 8.2 million cancer deaths worldwide in 2012 (Ferlay et al., 2014), the need for new molecular markers, highlighting novel biological interactions, is a primary objective for cancer research. A clearer understanding of how both cellular and noncellular components of the niche, in particular extracellular constituents, and how they influence cancer etiology and
progression is of growing importance and has become one of the key focus in cancer biology studies.

2.2 The Extracellular Matrix

2.2.1 Overview

As outlined at the beginning of this chapter, global cancer incidence rates are increasing significantly year on year. The global burden of cancer related illnesses has continued to increase largely in recent years particularly due to a growing and aging population (Thun et al., 2010) but also in part due to improved screening and detection rates. A key scientific focus to reverse this upward trend has been the identification of new prognostic markers and therapeutic targets which can be used to predict the incidence or outcome of cancer. These markers may be further utilised to screen the general population for differential diagnosis in symptomatic patients and for the clinical staging of the disease. Much effort has been devoted to determining how cellular components of the microenvironment initiate and promote cancer development (Bhowmick et al., 2004). Published expression profiling studies have uncovered a wide range of genes which have a range of roles in functions such as differentiation, cellular signalling and metastasis to name but a few. From these studies, greater understandings of the molecular mechanisms which direct how a cell undergoes transformations from a normal cell to a cancerous cell have been identified. Much of this data has become available freely online and allows researchers to gain an insight into how intricate these mechanisms can become. However, much of the published research fails to link results to one of the most important regulatory bodies with the human body, the cellular microenvironment and more specifically, the ECM.
The modulation of cellular function and tissue architecture in the body is controlled by guidance cues from the microenvironment composing both molecular and cellular elements (Harisi and Jeney, 2015) and specific sequences of dysregulation of gene-environmental interactions influences cancer development. The normal local microenvironment, or niche, of cells is a key component which supports and aids development of tissue and structures in the body. It constitutes diverse cell types (stem, progenitor, immune, fibroblasts, inflammatory cells etc.), blood vessels, signalling molecules and ECM which direct many essential functions and actions from embryonic development to organ growth. Cells in the body require many dynamic interactions with each of these components to modulate normal tissue development and homeostasis. The ECM is one of the most important regulators of cellular and tissue function in the body. It is a meshwork of both structural and
functional proteins assembled in unique tissue-specific architectures acting both as the mechanical framework for each tissue and organ and as a substrate for cell signalling (Brown and Badylak, 2014). The proteins which make up the ECM are large and complex macromolecules that initially were understood to constitute an inert, support structure which primarily only provided a physical scaffold to which the cells were attached. However, the biological dynamicity of the ECM which has emerged highlights its critical role in both normal homeostasis and more importantly its role in tumourgenesis. It has important roles in regulating normal cell and tissue development and function either through direct or indirect means (Bissell et al., 1982, Hay, 1993, Mecham, 2001, Moub et al., 2014) and plays crucial roles in the development of fibrotic disease and cancer (Cox and Erler, 2011, Bonnans et al., 2014, Lu et al., 2012). Cell growth, function, proliferation, migration and apoptosis are all influenced by cell-ECM interactions which make it indispensable for major developmental processes (Lu et al., 2011, Stickens, 2004). The “core” ECM matrisome comprises almost 300 proteins, including 43 collagen subunits, approximately 36 proteoglycans, and around 200 glycoproteins (Hynes and Naba, 2012). Important biological information is fed to the cells to control cell differentiation and proliferation as well as influencing cell polarity, regulating cell migration and cell turnover processes. Cells sense and respond to the mechanical environmental changes through transducers and convert mechanical input changes imposed by the ECM into complex intracellular signalling cascades that ultimately regulate cellular responses including adhesion, spreading, migration, and proliferation (Jaalouk and Lammerding, 2009). There is constant interaction between the ECM and surrounding epithelial cells by serving as ligands for cell receptors such as integrins, releasing growth factors such as epidermal growth factor (EGF) (Bonnans et al., 2014) and constantly rebuilding and restructuring tissue architecture (Lu et al., 2011). Local matrix composition and stiffness modulates cellular function through remodelling processes directed by the ECM where the mechanical environment undergoes both endogenous and exogenous forces (Bhatia, 2011). In addition, enzymes such as matrix metalloproteinases (MMPs) can act within the matrix to cleave and remodel the ECM proteins and proteolytic products thereby further remodelling the surrounding structures and influencing the overall microenvironment modification and development (Streuli, 1999). The wide spectrum
of dynamic-reciprocal ECM interactions in various cellular transformations and matrix remodelling highlights an essential role in cell and disease phenotype. When a cancer forms, these ECM interactions become amplified or reduced, dramatically increase malignant cell behaviour and allow for increased proliferation and metastatic progression (Joyce and Pollard, 2008). This multi-step process involves remodelling, basement membrane degradation and increased tumour cell motility which allows cancers to spread (Lu et al., 2012).

The ECM activity which is central to most of these interactive modifications is the primary focus of the research in this thesis and a key part is the selection of a definite panel of genes whose expression patterns may play a role in cancer progression. To explore the potential of the ECM genes as potential molecular markers for the research, a genetic panel subset was selected that hypothetically has an etiological role in cancer, particularly colorectal and breast carcinomas. Through direct or indirect means, the ECM regulates almost all cellular behaviour and is indispensable not only for major developmental processes (Lu et al., 2011, Stickens, 2004) but for tumour formation and progression also (Quail and Joyce, 2013). To date, only limited progress has taken place in the direct targeting of ECM components in cancer drug research, with integrins being the most targeted focus. However, many other drugs target ECM metabolism and subsequently modulate the composition and organisational structure of the ECM (Jarvelainen et al., 2009). Cell shape, function, proliferation, migration and apoptosis are all influenced by cell-ECM interactions which may, during pathological conditions such as cancer may become dysregulated leading to increased/decreased ECM gene activities. Therefore, the identification of prominent ECM tumour markers that derive biological insight into tumour development and progression would be of clinical value.

2.2.2 Fibrous proteins

The mammalian ECM is composed of a variety of multifunctional components including fibrous proteins, proteoglycans and glycoproteins and forms a complex network of three-dimensional structures among the cells of different tissues (Järveläinen et al., 2009). The primary fibrous proteins which constitute the ECM are collagen, elastin and fibronectin.
One of the most dominant of proteins within the ECM is collagen and it is the main structural component providing tensile strength, regulation of adhesion and direct tissue development (Rozario and DeSimone, 2010). Twenty eight types of collagen have been identified and described (Gordon and Hahn, 2010) and up to 30% of the mass of vertebrates is composed of collagen laying the structural fabric of the tissue architecture (Sherman et al., 2015). Primarily, it is secreted by fibroblasts within the stromal tissue structure (De Wever et al., 2008) which can influence its organisation and alignment into sheet or mesh-like structures with anchoring fibrils (Alberts B, 2002c). A primary characteristic of the main collagen type in the body (Type I) is enormous tensile strength which allow it to be stretched without being broken, making it ideal for withstanding muscle and tendon forces (Lodish H, 2000b). Collagen interacts with other fibrous proteins such as elastin to organise the interstitial ECM.

As an essential and unique part of various tissues that depend on elasticity, elastin is essential to fulfil the various mechanical requirements of basement membranes and connective tissues. Elastin supports structures by providing recoil to tissues that undergo repeated stretch (Frantz et al., 2010) and allows tissues to contract and

**Figure 16** - Schematic of connective tissue underlying an epithelium with primary ECM components such as collagen, proteoglycans and glycoproteins (Mahmud, 2013)
resume their shape. Again, it is primarily produced by fibroblasts during embryonic development and early childhood stages with almost no elastin produced by adult tissues. Approximately 10% of elastin fibre content is lost over a lifetime resulting in decreased flexibility, repair and strength of tissue structures. Abnormalities in the elastin gene are known to be responsible for several genetic disorders, disrupting the architecture and function of elastin-rich tissues (Wise and Weiss, 2009).

A third fibrous protein, fibronectin (FN) is intimately involved in directing the organisation of the interstitial ECM and, additionally, has a crucial role in mediating cell attachment and function (Frantz et al., 2010). It consists of two subunits and is assembled into an insoluble fibrillar matrix in a complex cell-mediated process (Wierzbicka-Patynowski and Schwarzbauer, 2003). Fibronectin fibrils are not static but are rearranged and recycled by cell movements, cell density and degradative processes (Hynes, 1999) which further influences interactions with previously mentioned collagen and elastin creating a dynamic environment. It is expressed by fibroblasts along with other cell types and can act as a ligand for a dozen members of the membrane-spanning integrin receptor family (Plow et al., 2000). One of the key roles of fibronectin fibrils is to assist in wound-healing processes where it assists in the formation of the membrane for the growth and adhesion of cells by binding to Type III collagen fibrils in particular (Engvall et al., 1978).

### 2.2.3 Proteoglycans and Glycoproteins

Proteoglycans can be secreted or cell-surface bound and serve diverse functions including ECM assembly and mediating cell adhesion and motility (Tsang et al., 2009). They are comprised of one or more glycosaminoglycans (GAG) chains attached to a core protein (Hynes and Naba, 2012) and have a wide variety of functions that reflect their unique buffering, hydration, binding and force-resistance properties (Frantz et al., 2010). Rather than providing structural strength like fibrous proteins, they confer additional properties on the ECM composition (Järveläinen et al., 2009) and are considered to be interfibrillary ECM molecules (Figure 17).
The primary function of proteoglycans derives from the hydrodynamic characteristics of the GAG components of the molecules which bind water to provide hydration and compressive resistance in the tissue structure (Mouw et al., 2014) which influences morphogenetic cell behaviours and regulates the diffusion of many secreted growth factors and morphogens (Rozario and DeSimone, 2010). They are upstream of many signalling cascades and are capable of affecting intracellular events and modulating distinct pathways (Schaefer and Schaefer, 2010b) in both normal and abnormal tissue structures including binding to and activation of epidermal growth factor receptor (EGFR), insulin-like growth factor 1 receptor (IGFIR) (Goldoni and Iozzo, 2008). Abnormal expression or deregulated function of these proteoglycans affect cancer and angiogenesis, and are critical for the evolution of the tumour microenvironment (Iozzo et al., 2009). Therefore, changes in the expression and degradation of the proteoglycan family members contribute to altered ECM and therefore could be potential molecular markers for cancer diagnosis and therapy. In addition to fibrous proteins and proteoglycans, a third main ECM
element is glycoproteins which, along with proteoglycans and hyaluronan, normally form a hydrated gel filling the interstices of the fibrous network (Alberts B, 2002c). Similar to proteoglycans, glycoproteins are interfibrillar ECM proteins with an attached sugar/carbohydrate unit, often called glycan, which gather into short oligosaccharide chains. The carbohydrate portions of many glycoproteins are carriers of information crucial for the functioning of many biological recognition systems (Cole and Smith, 1989). They are frequently present at the surface of cells and are actively secreted from cells where they have structural and crosslinking roles. They also function as integral membrane proteins, where they play a role in cell-cell interactions. The diverse range of glycoproteins and their functional roles in the human body are yet to be fully explored and their interactions allow for ECM assembly, domains and motifs promoting cell adhesion and also signalling into cells and other domains that bind growth factors (Hynes and Naba, 2012). The role of glycoproteins in tissue organisation and structure is crucial and dysregulation in the synthesis process for the proteins can influence interstitial matrix remodelling especially for fibrotic diseases and cancer (Bonnans et al., 2014). Glycoprotein targets highlighted for this study are Secreted protein acidic and rich in cysteine (SPARC), CD9 Molecule (CD9), N-Caderin (CDH2) and Carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1). SPARC is an acidic extracellular matrix glycoprotein that has shown to be significantly overexpressed in many human cancers such as colon and breast. Its overexpression in malignant tumours has been observed and correlates with disease progression and poor prognosis (Guweidhi et al., 2005) and plays a vital role in collagen binding and cell-matrix interactions. CD9 is a cell surface glycoprotein which is a member of the tetraspanin family. It has been identified as playing a role in the metastatic potential of colon carcinomas where its expression results in enhanced integrin dependent adhesion and inhibition of cell growth (Ovalle et al., 2007).CDH2 is a transmembrane glycoprotein which plays a role in cell-cell adhesion. Although originally identified in neural tissue, CDH2 is commonly found in cancer cells where it associates with proto-oncogene kinase pathways to break down intracellular connections and provide a mechanism for transendothelial migration. CEACAM1 is a cell adhesion glycoprotein and is primarily thought to be a tumour suppressor in humans. Multiple cellular processes are attributed to CEACAM1 such as metastasis
and apoptosis as well as being highlighted as a prognostic tool for breast cancer in preliminary studies (Wang et al., 2011)

2.2.4 Integrins

Integrins are a large class of transmembrane glycoprotein receptors on the cell surface involved in direct communication with the ECM (Hynes, 1987) and act as cell-cell and cell-matrix mediators (Koistinen P, 2000-2013). In the integrin family of receptors, 18 alpha (α) subunits and 8 beta (β) subunits have been identified which associate with each other to generate 24 different receptors with distinct ligand specificities (Hynes, 2002).

Integrins link the ECM to the actin cytoskeleton in response to signals from the ECM and function to stabilise the microstructure of the tissue, regulate mechanical state of tissue and participate in key morphological processes such as cellular adhesion and migration (Sastry and Burridge, 2000, Petit and Thiery, 2000, Lauffenburger and Horwitz, 1996). Integrins are a primary mechanical connection between the ECM and the cell connected through intracellular and extracellular structures (Schwartz, 2010) (Figure 18). The extracellular structures of integrins

![Integrins](image_url)

**Figure 18** - Integrins are transmembrane receptors that bridge between the intracellular and extracellular space. They are heterodimeric meaning that they have two different chains; an alpha and beta subunit.
determines binding specificity that recognise diverse matrix ligands such as collagen, fibronectin and laminin (Huttenlocher and Horwitz, 2011) while the intracellular components regulate ECM ligand affinity which leads to integrin activation (Shattil et al., 2010). Cells respond to force on integrin-mediated adhesions by remodelling the ECM and stabilising tissue structure and architecture. The forces which mechanically act upon cells can be transmitted through the membrane which in turn activates cellular responses in the form of multicomponent signalling complexes and genetic modifications which can control cell behaviours and fate. Therefore, by binding to the ECM at the cell surface, integrins provide mechanical links for cells along with transducing biomechanical cues from the microenvironment to the cell nucleus for downstream signalling changes. On ligation to the ECM, integrins cluster in the plane of the membrane and recruit various signalling and adaptor proteins to form structures known as focal adhesions (Desgroisellier and Cheresh, 2010) (Figure 19).

![Diagram](image)

**Figure 19** – Integrin mediated cell adhesion to the ECM is enhanced with the formation of large macromolecules on the inner membrane of the cell known as focal adhesions. Adapter and scaffold proteins provide strong mechanical link between the cytoskeleton and ECM.
Focal adhesions are large dynamic macromolecule protein complexes commonly with both mechanical components and cell signalling components found at the leading edge of migrating cells. The integrin clustering which leads to focal adhesions is sufficient to promote cellular changes such as focal adhesion kinase (FAK) phosphorylation. This allows FAK to cooperate with multiple signalling pathways through direct interaction with other cell surface receptors and adaptor proteins. Further specific integrin-mediated cellular changes interact with dynamic variations in their extracellular environments in order to regulate cellular adhesion, inside-out signalling, and polarised cell migration. Polarised cell migration is characterised by axisymmetric adhesion dynamics where adhesions occur at the leading edge and disassembly at the rear of the cell act to pull the cell forward to attractive forces such as growth factors or chemo-attractants (Huttenlocher et al., 1995). The formation and turnover of focal adhesion structures are vital processes during cell migration and are essential for the transmission of signals from the ECM through to the cell. The important biochemical pathways that are mediated by integrins are therefore not only important in normal tissue homeostasis but also in the pathology of cancer. A wide variety of integrins are highly implicated in tumourgenesis and disease progression (Baker and Zaman, 2010). An altered integrin pattern allows the cancer cells to recognize variable matrices, but it may also lead to altered ECM signalling and changes in gene expression (Koistinen P, 2000-2013). As many tumours originate from epithelial cells, the integrins expressed by these cells are predominantly retained in the tumour though the expression levels may be altered (Desgroisellier and Cheresh, 2010). The integrin receptors at the membrane of the cell coupled with chemical-led changes such as integrin-mediated cell attachment and signalling are critical in cancer initiation, development and metastasis. Integrin-clustering promoted focal adhesions can become the mechanism which a cancer cell uses to transition through the tissue to locations which are more favourable for it to thrive (Levental et al., 2009). The integrin clustering is particularly related to ECM stiffness in and around tumours (Paszek et al., 2005) and results in enhanced mechanotransduction that consequently promotes collagen crosslinking, enhanced kinase activity and further tumour cell migration and invasion (Cox and Erler, 2011, Lo et al., 2000).
2.2.5 Growth factors

Growth factors are naturally occurring substances in the body and are important for stimulating a range of cellular processes. Growth factors can be classified into different families on the basis of their target cells, role in cellular processes and overall structure with insulin-like growth factor (IGF), vascular endothelial growth factor (VEGF) and epidermal growth factor (EGF) families being the most well-known. They act as signalling molecules between cells but their activity can influence the rate of cellular growth, proliferation, differentiation and migration by promoting the synthesis of proteins and inhibiting degradation (Alberts B, 2002a). Typically, they bind to their receptor on the surface of target cells which induces their signal transduction thus are essential to normal cell cycle (Figure 20).

![Diagram](image)

**Figure 20**- Growth factors bind to receptors on the surface of the cell and transduce signals through pathways to influence cellular control systems (Lee et al., 2010)

The receptors are transmembrane proteins which transmit the information to the intracellular space. This in turn can lead to a molecular response from the cell and extracellular changes and responses such as binding of growth factors to ECM. Primarily, the ECM acts as a repository for growth factors (Kim et al., 2011) where they are released as soluble mediators in response to environmental cues. The ECM can directly interact with growth factors by binding and releasing to influence growth factor receptor signalling and can induce growth factor expression (Schultz and Wysocki, 2009) which is achieved by white blood cell adherence to the ECM.
which stimulates synthesis of growth factors. The ECM can also offer growth factors protection from degradation and helps develop concentration gradients by controlling the growth factor bioavailability which is important for further cellular processes such as migration (Wilgus, 2012). Inversely, growth factors are important facilitators for ECM regulation, helping regulate its production and degradation along with the physical organisation of tissue which highlights a bidirectional, dynamic relationship. Furthermore, growth factors stimulate cells to increase the production of ECM components or enhance synthesis of MMPs that break down ECM (Schultz and Wysocki, 2009). Cell-cell and cell-matrix interactions can be altered by the activity of proteases and their presence can limit the bio-availability of many growth factors, growth factor receptors and other significant molecules necessary for normal homeostasis of the ECM.

2.2.6 Matrix Metalloproteinases

Figure 21- MMPs function to alter and shape the basement membrane by a) remodelling, b) digestive, c) growth factor cleavage and d) inhibitor degradation processes. This allows MMPs to become a key player in the metastatic potential of cancer cells (Nagase et al. 2006)
Matrix-metalloproteinases are members of the protease family and play a key role in ECM protein degradation and tissue remodelling. Proteases are defined as a group of enzymes, whose catalytic function is to hydrolyse, or breakdown, proteins (Schumann, 2006) (Figure 21). The MMP family consists of approximately 25 secreted and cell surface enzymes that process or degrade ECM proteins, biological molecules and functional receptors as well as other substrates (Sternlicht and Werb, 2001) via proteolytic cleavage. Significant crosstalk occurs between members of the MMP family and the ECM to achieve successful remodelling of the environment. The interactions between the ECM and MMPs represent a prominent process of matrix turnover and tissue remodelling. In normal tissue homeostasis the expression of ECM remodelling enzymes are controlled at multiple levels from transcriptional to posttranslational regulation to ensure they are expressed at specific times in specific cells (Lu et al., 2011). However, it is during the tumourgenesis process that MMP activity can play important roles in cell migration and drive the dissemination potential of cancer cells into the normal adjacent tissue (Friedl and Wolf, 2008). MMPs can also be involved in disrupting the balance between growth and antigrowth signals in the microenvironment, as they potently influence the bioavailability or functionality of multiple important factors that regulate growth (Kessenbrock et al., 2010). This highlights the ECM-degrading proteases ability to play a critical role in both normal tissue functionality and also tumourgenesis, where they can act as positive, as well as negative, regulators of endothelial cell proliferation and vascular morphogenesis (DeClerck et al., 2004).

2.2.7 Other ECM Molecules

**Periostin (POSTN)** – Periostin is a secreted ECM protein which plays a wide variety of roles in tissue development. In many cancers, periostin binds to integrins on cancer cells, activating the Akt/PKB- and FAK-mediated signalling pathways (Morra and Moch, 2011). It is predominantly over-expressed in several types of cancer which can lead to increased cell survival, invasion, angiogenesis, metastasis, and the epithelial-mesenchymal transition.

**Interleukin 8 (IL8)** - Interleukin 8 is a chemokine associated with inflammation where it plays a role in neutrophil recruitment. It has been reported to promote tumour cell growth in colon cancer cell lines post binding to its receptors (Itoh et al.,
2005). It has also been implied to act as an autocrine growth factor for the possible increased migration colon carcinoma cells by cleaving metalloproteinase molecules (Li et al., 2005).

**C-X-C motif chemokine receptor 2 (CXCR2)** - CXCR2 is functionally important for multiple processes and its increased expression on tumours, along with its ligand IL8, has been associated with CRC growth, progression and recurrence in patients (Lee et al., 2012).

**Receptor of activated protein C kinase 1 (RACK1)** – RACK1 is a cytosolic protein involved in signalling pathways and protein trafficking within cells. Its diverse mechanisms permit it to being involved in normal cell growth, adhesion and movement as well as in multiple tumour types. Additionally it has been shown to interact with IGF-1R, ITGB1 and PKC in response to IGF1 stimulation influencing tumour cell proliferation and migration (Hermanto et al., 2002, Kiely et al., 2008).

**Focal Adhesion Kinase (FAK)** – FAK is a cytoplasmic protein that is primarily involved in the process of cell adhesion and migration. It is over expressed and activated in several advanced stage malignancies (Sulzmaier, 2014) and has been shown to decrease the motility of breast cancer cells when its expression is blocked (Leslie, 2009).

### 2.2.8 Gene selection procedure

A detailed analysis of extracellular matrisome literature reviews was undertaken to identify key ECM genes involved in cancer formation, development and progression. Prominent genes from a variety of genetic families were selected which code for cell surface proteins, growth factors and tissue remodelling proteins. Other key ECM proteins which could play a role in the development and progression of various carcinomas were also highlighted. Extensive datasets from published gene expression profiling studies of colorectal and breast cancer tissue, as well as cell lines, which focused on ECM genes were examined to aid in the selection process of genes which could be involved in various cancerous developmental processes. In support of this, gene omnibus repositories were studied for additional ECM genes which showed evidence of being more often involved in carcinomas of the colon, rectum or breast as it was the focus for this research. Furthermore, a detailed analysis of commercial ECM expression arrays, including Taqman® and Qiagen® ECM and
Adhesion Molecule arrays (Catalogue number: 4414133 and PAHS-013ZA)(Thermo Fisher Scientific, Qiagen;) was performed to support hypotheses developed for which genes could be most influential in the metastatic spread of the disease. Once completed, a panel of thirty-four ECM related genes was identified. The genes are part of the larger ECM gene family which code for approximately 300 proteins that make up the core matrisome. It is anticipated that differentially expressed genes in matched cancer samples will be identified highlighting variations wither between normal and malignant samples or variations withing individual samples. It is also anticipated that the expression patterns will correlate to an increased metastatic potential, lymph node involvement or another clinical parameter. A list of the selected genes for analysis in this study can be seen in Table 2 with references for studies which show specific dysregulation of these genes.
<table>
<thead>
<tr>
<th>Fibrous Proteins</th>
<th>Gene Name</th>
<th>Gene Symbol</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collagen, type I, alpha 1</td>
<td>COL1A1</td>
<td>(M. K. Bode, 2000, Thermo Fisher Scientific)</td>
<td></td>
</tr>
<tr>
<td>Collagen, type III, alpha 1</td>
<td>COL3A1</td>
<td>(Wu et al., 2012, Kahlert et al., 2011)</td>
<td></td>
</tr>
<tr>
<td>Laminin, alpha 1</td>
<td>LAMA1</td>
<td>(Thermo Fisher Scientific, Moser et al., 1994)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Glycoproteins</th>
<th>Gene Name</th>
<th>Gene Symbol</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secreted protein, acidic, cysteine-rich</td>
<td>SPARC</td>
<td>(Viana Lde et al., 2013, Liang et al., 2010)</td>
<td></td>
</tr>
<tr>
<td>Cadherin-2</td>
<td>CDH2</td>
<td>(Ding et al., 2013, Kroepil et al., 2012)</td>
<td></td>
</tr>
<tr>
<td>Carcinoembryonic antigen-related cell adhesion molecule 1</td>
<td>CEACAM1</td>
<td>(Arabzadeh and Beauchemin, 2012, Fiori et al., 2012)</td>
<td></td>
</tr>
<tr>
<td>CD9 Molecule</td>
<td>CD9</td>
<td>(Mori et al., 1998, Hashida et al., 2003)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transmembrane molecules</th>
<th>Gene Name</th>
<th>Gene Symbol</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrin, alpha 2</td>
<td>ITGA2</td>
<td>(Bertucci et al., 2004, Bianchini et al., 2006)</td>
<td></td>
</tr>
<tr>
<td>Integrin, alpha 5</td>
<td>ITGA5</td>
<td>(Viana Lde et al., 2013, Wu et al., 2012)</td>
<td></td>
</tr>
<tr>
<td>Integrin, alpha 8</td>
<td>ITGA8</td>
<td>(Thermo Fisher Scientific)</td>
<td></td>
</tr>
<tr>
<td>Integrin, alpha V</td>
<td>ITGAV</td>
<td>(Viana Lde et al., 2013, Bertucci et al., 2004)</td>
<td></td>
</tr>
<tr>
<td>Integrin, beta 1</td>
<td>ITGB1</td>
<td>(Thermo Fisher Scientific, Song et al., 2014)</td>
<td></td>
</tr>
<tr>
<td>Integrin, beta 4</td>
<td>ITGB4</td>
<td>(Thermo Fisher Scientific, Masugi et al., 2015)</td>
<td></td>
</tr>
<tr>
<td>Integrin, beta 5</td>
<td>ITGB5</td>
<td>(Thermo Fisher Scientific) (Wu et al., 2012)</td>
<td></td>
</tr>
<tr>
<td>Integrin, beta-like 1</td>
<td>ITGBL1</td>
<td>(Qiagen;)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Growth Factors</th>
<th>Gene Name</th>
<th>Gene Symbol</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epidermal growth factor</td>
<td>EGF</td>
<td>(Salomon et al., 1995, Feng et al., 2012)</td>
<td></td>
</tr>
<tr>
<td>Epidermal growth factor receptor</td>
<td>EGFR</td>
<td>(Spano et al., 2005b, Bertucci et al., 2004)</td>
<td></td>
</tr>
<tr>
<td>Insulin-like growth factor 1</td>
<td>IGF1</td>
<td>(Shiratsuchi et al., 2011, Giovannucci, 2001)</td>
<td></td>
</tr>
<tr>
<td>Insulin-like growth factor 1 receptor</td>
<td>IGF1R</td>
<td>(Freier et al., 1999, Baserga, 2000, Wang and Sun, 2002)</td>
<td></td>
</tr>
<tr>
<td>Insulin-like growth factor 2</td>
<td>IGF2</td>
<td>(Li et al., 2004, Oshima et al., 2008)</td>
<td></td>
</tr>
<tr>
<td>Insulin-like growth factor 2 receptor</td>
<td>IGF2R</td>
<td>(LeRoith and Roberts, 2003, Velázquez-Fernández et al., 2005)</td>
<td></td>
</tr>
<tr>
<td>Insulin-like growth factor binding protein 2</td>
<td>IGFBP2</td>
<td>(Rohan;, 2000, Walker et al., 2007)</td>
<td></td>
</tr>
</tbody>
</table>
**Table 2**- Candidate cancer-related genes that may play roles in regulating the composition of the ECM and be involved in the progression and development of cancer.
2.3 Chapter Closure

This chapter presents a review of the fundamental dynamics of cancer formation with a specific focus on colorectal and breast carcinoma development and progression. Incidence rates and risk factors which can propagate cancer formation such as age, diet and lifestyle factors are outlined along with familial and genetic changes that can influence tumourgenesis. It is highlighted that mutations in the DNA transcripts within cells in the body can inherently lead to the development of specific pathogens and be further exasperated by the involvement of the microenvironment which surrounds the cell. Biomarker discovery is then delineated and the process of identifying and developing potential new biomarkers for the personalised treatment of patients is explained. With the emergence of powerful genomic and proteomic technologies, new individual biomarkers and cancer signatures could be identified for early detection disease diagnosis and new therapies for patients.

Pathological alterations which occur during cancer may promote dysregulation of extracellular dynamics leading to increased or decreased ECM gene activities which further influences abnormal cellular behaviour. Therefore, an expanded review of the extracellular matrix and some of the major constituents is outlined. Included in the discussion are cell surface receptors, growth factors and ECM proteases which have engrained roles in both normal tissue homeostasis and carcinogenesis. It is further highlighted that dysregulation of the communicative pathways and remodelling networks during tumourgenesis can directly influence disease progression. Therefore, as the ECM is not just a passive support structure for cells, but also a bio-diverse and functional component of the tissue, a panel thirty-four ECM genes has been established to study for this research. Fundamental literature reviews of ECM publications and gene repositories, along with analysis of current established ECM and adhesion molecule arrays, has allowed for the concordance of these genes which will be examined for genetic dysregulation in the following chapters of this thesis.
CHAPTER 3
MATERIALS AND METHODS
3 Materials and Methods

3.1 Introduction

The primary objective of this thesis is to perform quantitative biological reactions on a microfluidic, droplet-based, PCR instrument using novel methods. This chapter details the constituents of the PCR mixes employed in each section of experimentation. The materials used for biological set-up are listed firstly. The clinical sample acquisition, sample preparation techniques and sample loading protocol are then examined. Following this, detail is given on specific PCR reaction components that make up the reagents along with an overview of the real-time PCR process are delineated. Cycle threshold levels are then outlined and the method for determining them is explained. The calculation of fold change is then discussed along with determination of the statistical analysis steps in the order that they are performed.

3.2 Materials

Plates

- Greiner CELLSTAR® 384 flat bottom well plates were purchased from Sigma-Aldrich Ltd. (Wicklow, IE)

Carrier fluid

- PD-5 silicone oil (density 0.92 g/cm³) was purchased from Brenntag (United Kingdom)

Cell lines

- HCT116 colon cancer cell line was obtained from ATCC (CCL-247™). They are from an adult male with colorectal carcinoma. They were maintained in Dulbecco Modified Eagle Medium (DMEM) supplemented with L-glutamine, penicillin/streptomycin and 10% FBS.

Cell Culture Reagents

- DMEM, foetal bovine serum (FBS), Penicillin/Streptomycin antibiotic mix and Trypsin-EDTA were purchased from Sigma-Aldrich Ltd (Wicklow IE).
- Optimem, Lipofectamine 2000 transfection reagents were purchased from Life Technologies (Dublin, IE).
General Chemical and Reagents

- The RNeasy® Plus Mini Kit was purchased from Qiagen (Manchester, UK).
- The Superscript® VILO™ cDNA synthesis kit was purchased from Invitrogen (Dublin, IE). siRNA were purchased from Sigma-Aldrich Ltd. (Wicklow IE).
- The Hyclone Hypure Molecular Biology Grade Water was purchased from Thermo Fisher Scientific (Waltam, MA, USA).
- Taqman® Gene Expression assays and Taqman® Gene Expression Master Mixes were purchased from Thermo Fisher Scientific (BioSciences, Dublin, IE).

3.3 Methods

3.3.1 Clinical Sample Retrieval

Colon and breast bio-banks were established in the University Hospital Limerick (UHL) which allowed for the collection of samples from patients who underwent surgical procedures at UHL. Informed consent was obtained from patients (ethical approval number 73/11).

Colorectal: Colorectal cancer tissues and paired adjacent physiologically normal mucosa tissue samples were collected from a heterogeneous group of 24 patients at different stages of malignancy, undergoing colorectal resection for cancer at University Hospital Limerick. Primarily the tumours are adenocarcinomas. After removal of the colon and associated mesentery the specimen was opened. Mucosal samples were taken from the tumour and normal mucosa at a distance of 10cm from the tumour. Samples were immediately placed in Allprotect tissue reagent (Qiagen) and stored in a -80-degree freezer.

Breast: Breast tumour, benign and adjacent pathologically normal tissue samples were harvested from a Caucasian female patient with a metaplastic spindle cell carcinoma. The samples were cut directly from the tumour post-surgery and a sample from the normal mucosa was taken approximately 10cm from the tumour in a section of uninvolved breast tissue. The samples were then subject to receptor status staining and tested negative for ER, PR receptors and also negative for HER2 overexpression (TNBC). Post-sectioning and staining, the tumour was stored in
Allprotect® Tissue Reagent (Quigen) for RNA stabilisation. The samples were then snap frozen immediately in liquid nitrogen and stored in a -80°C freezer until required for gene expression profiling using qPCR.

### 3.3.2 Workspace Preparation

Sample preparation was carried out using standard biological protocols in a contaminant free UV sterilisation cabinet. PCR is a highly sensitive process and minute amounts of DNA can compromise experimentation. The cabinet is fitted with an air filtration system which removes airborne particles from the working environment. Standard protocol advises turning on this system for 10 minutes before use to remove possible airborne particles. The UV light was also switched on for 10 minutes to denature any nucleic acids in the hood which further eliminated possible contamination sources. Surfaces, pipettes and associated apparatuses were sprayed with 70% ethanol before commencing hood operations.

### 3.3.3 RNA Extraction and cDNA Synthesis

*RNA was extracted from the tissue samples using the below protocols by a colleague, Dr. Catriona Dowling, and made available in the Kiely Laboratory.*

**Colorectal:** Frozen tissue was immersed in liquid nitrogen and ground into powder. Lysis buffer was added to tissue and the sample transferred to tubes. Total ribonucleic acid (RNA) was extracted using RNeasy Plus mini kit® (Qiagen) according to manufacturer’s instructions. RNA was evaluated for purity on the NanoDrop™ 1000 Spectrophotometer (Thermo Fisher). RNA has its absorption maximum at 260nM and the ratio of the absorbance at 260nM and 280nM is used to assess the RNA purity of an RNA preparation. RNA quality was evaluated through visualization of the 28S:18S ribosomal RNA ratio on a 1% agarose gel.

**Breast:** Breast tissue samples were also immersed in liquid nitrogen and homogenised into a powder. One millilitre of QUIzol® reagent (Quigen) was added. Subsequently the mixture was transferred to a Maxtract tube (Quigen) with 200ul of chloroform added. RNA was then extracted as per the RNeasy Lipid Tissue Mini KIT® (Quigen) to manufacturer’s instructions. RNA was eluted in 30μl of RNase free water and stored at -80°C. The mRNA is unstable and quickly degraded which makes it unsuited to direct quantification or frequent analysis. Therefore, once the RNA is extracted from the tissue it is then converted, or transcribed, to
complimentary DNA or cDNA which is a more stable genetic format using a
technique called reverse transcription. cDNA synthesis (*performed by a colleague, Dr. Susan Dwane*) was completed using Superscript® VILO cDNA Synthesis Kit (Invitrogen/Thermo Fisher, Carlsbad, USA) according to the manufacturer’s protocol. Reactions were incubated for 25 degrees for 10 minutes followed by 60°C for 45 minutes and 72°C for final denaturation for 15 minutes. Samples were stored at -20°C until ready for PCR analysis.

3.3.4 Stromal and Epithelial Content

The epithelial-stromal ratio can also vary considerably depending on factors such as the density of the tissue, tumour size and stiffness. These factors may influence the cellular content and ultimately the gene expression quantification when a particular gene is being assayed, highlighting the necessity of quantifying the cellular content ratios in each sample before performing gene expression analysis. Tumours principally consist of a wide range of cell types and this can have great influence over the progression of a malignancy. The tumour connective framework comprises of the malignant cell types and also includes fibroblasts, inflammatory cells, endothelial cells, adipocytes, and extracellular matrix (Mueller and Fusenig, 2004) which can all contribute to the development and progression of a carcinoma. It is also suggested that there is an interdependence of growth between the epithelial and stromal components in these tumours that explains their complex morphology (Sawhney et al., 1992). To determine if an epithelial and stromal component variation existed in the samples, two marker probes, *Cytokeratin-18* (KRT18) and *Vimentin* (VIM) were used to quantify the epithelial and stromal composition of each sample (Dowling et al., 2016a, Latil et al., 2001). Each of the twenty-four matched samples were evaluated by qPCR to approximate for tissue components with average composition shown to be consistent and can be seen in *Figure 22.*
Materials and Methods

For the breast tissue similar quantification is necessary to ensure the individual tumour sections have a similar stromal/epithelial ratio to that of the normal sample. No significant variation for each cell type was observed in samples TH2- TH8 which were shown to have 55% epithelial composition and 45% stromal composition, in comparison to the corresponding normal sample TH1 with approximately 52.5% epithelial composition and 48.5% stromal composition. The results of the stromal-epithelial assay can be seen in Figure 23.

![Figure 23](image)

**Figure 22** - The figure shows the determination of epithelial and stromal content of the colorectal samples.

For the breast tissue similar quantification is necessary to ensure the individual tumour sections have a similar stromal/epithelial ratio to that of the normal sample. No significant variation for each cell type was observed in samples TH2- TH8 which were shown to have 55% epithelial composition and 45% stromal composition, in comparison to the corresponding normal sample TH1 with approximately 52.5% epithelial composition and 48.5% stromal composition. The results of the stromal-epithelial assay can be seen in Figure 23.

![Figure 23](image)

**Figure 23** - The graph shows that there is minimal variation in cellular content between the cancerous breast tissue and the matched normal tissue sample.
3.3.5 Sample Preparation, Aspiration and Mixing.

DNA and assay were removed from the -20°C freezer. Mastermix, which is stored in a 3°C fridge, was also removed. The gene assay and DNA were thawed and vortexed to ensure homogeneity. Mastermix was gently mixed. The aliquots were then centrifuged for 5 seconds to ensure the reaction volume was spun down to the bottom of the tube. Two separate hoods were used to prepare DNA and reagents to avoid cross contamination. DNA (1:10) volumes of 6µl were diluted with 14µl of nuclease free water to create a total DNA well volume of 20µl. Mastermix, assay for the gene of interest (GOI) and water (22.5nl, 2.25nl, 5.25nl) were then prepared in an aliquot and pipetted into a second micro plate to create a total well volume on the reagent side of 30µl. The dead volume or volume that cannot be aspirated from the bottom of the well, for each experiment is 10µl. These volumes result in a final usable volume of 10µl (10000nl) available on the cDNA sample side and 20µl (20000nl) available on the assay-mastermix side for experimentation.

![Figure 24](image)

Figure 24 –The sample is aspirated from the well plate by the dipping tip which moves through the oil overlay into the sample premix before retracting and moving to the next well. On the right, a 384 well plate used for experimentation is shown which illustrates the modified wells to allow for an oil overlay to be added to the plate.

Using these volumes allows for 100 replicate experiments to be carried out on a single gene-assay expression analysis. This is a significant reduction in comparison to a commercial instrument's (AB 7900 HT) volume where 50µl cDNA sample, 25µl of assay, 250µl of mastermix would be required for approximately 100 reactions. The instrument volumes can be modified to suit individual experiments depending on genetic focus and volumes available. The cDNA is prepared into micro plates.
which have volumetric capacity of 112µl, with further plates prepared with the assays for the gene of interest and the mastermix. A total sample volume of 40µl can be pipetted into individual wells of the plate as the plates are modified to allow the instrument dipping head access multiple wells. The modification to the plate is a 12mm deep milled out section designed to hold a silicone-based oil overlay and allows continuous dipping while preventing air ingress (Figure 24). The plates were subjected to UV light sterilisation pre-sample addition. The oil overlay in the remaining well volume prevents the sample from being environmentally contaminated and also acts as a reservoir for fluid circulation during experimentation. The dipping tip can withdraw from a sample well and remain in the oil and access another well without interruption to the fluidic balance in the system. Once the pumping system reached the specific set point the dipping protocol commences. The dipping protocol was defined and controlled via integrated system software. The tip is preprogramed to dwell in the sample for a period of time which corresponds to a desired sample volume. Nanoliter quantities of the sample and reagents for the reaction are aspirated from the well and the tip moved back to the oil overlay. Additional wells can then be accessed to allow for combination dipping or no temple controls droplets to be aspirated. The components are combined in a stable liquid micro environment, the liquid bridge, which is a key design component of the platform. It allows for the cDNA and reaction chemistries to be mixed precisely without mechanical or physical exchanges. The final reaction volumes generated are of nanolitre scale and continuously flowing through the system carried by a carrier fluid without contacting surfaces such as tubing walls or junctions.

### 3.3.6 Sample Amplification

Based on the primary goal of this thesis, high throughput PCR was the method chosen to incorporate into microfluidic technologies developed to successfully determine genetic target quantitates. Quantification of these DNA targets allows for the determination of the exact genetic amounts expressed by cells at a given time, in a specific sample. The process involves the amplification of minut fragments of DNA to levels at which biological questions can be asked of it. It relies on repeated temperature cycling of the sample between two thermal zones of ~60°C and ~95°C.
for the DNA replication process to occur. The PCR requires several components and reagents for successful amplification of the DNA.

The components include:

- **DNA template**- the template contains the DNA fragment that needs to be amplified. It can either be cDNA or genomic DNA (gDNA) depending on whether one or two step PCR is performed.

- **Primers**- these are complementary to the DNA target’s 3 prime ends of the single stranded DNA (ssDNA). The primers are both forward and reverse primers and are specifically designed for the target sequence. The optimal primer length is 15-20 base pairs (bp) which is short enough to bind to the correct sequence at the annealing temperature and long enough to ensure adequate specificity.

- **DNA polymerase**- which is an enzyme isolated from a thermophilic bacterium that is able to withstand the denaturisation temperature.

- **Deoxynucleotide triphosphates (dNTPs)**- mix which contains the four deoxyribonucleotides- deoxyadenosine triphosphate, deoxycytidine triphosphate, deoxythymidine triphosphate and deoxyguanosine triphosphate.

- **Buffer solution**- provides a suitable chemical environment which gives maximum activity to the DNA polymerase

- **Magnesium chloride**- acts as source of magnesium ion for PCR, influencing the primer template annealing temperature, fidelity, specificity, and yield.

**Figure 25**- ds-DNA is heated to 90-95°C to denature the hydrogen bonds that bind the double helical structure and allow for the primers to attach at the specific sequence.
Heating the sample to \(~95^\circ C\) for 10 minutes (Figure 25) is the first step in the PCR process and allows for the denaturisation of the hydrogen bonds which are functioning to bind the DNA structures together to form double stranded DNA (dsDNA) helix. The helix is unwound to form two single strands of complimentary DNA bases. Once the double stranded DNA is denatured into two ssDNA sequences, the temperature is reduced to 60°C which allows the specific oligonucleotide sequences (primers) to anneal to the region of interest. The DNA polymerase then extends the primers along the ssDNA, forming a new dsDNA. A temperature of 72°C was originally required for extension of the primers but a redesign of reaction chemistries has allowed for annealing and extension to be carried out at the same temperature of 60°C which permits for simpler PCR devices and overall faster PCR (Figure 26).

![Figure 26](image)

**Figure 26-** A) Single stranded primers anneal to the ss-DNA template when the annealing process is being performed at a temperature of 60°C. B) The polymerase also attaches at this temperature and extends along the primers incorporating complimentary nucleotides to the template until a new ds-DNA strand is formed.

The PCR cycling process is repeated up to 40 times, potentially increasing the quantity of the original template section of DNA by \(2^{40}\) or over 1 billion times. The amount of DNA should, in theory, double at each cycle but depends on the efficiency of the reagents and the polymerase. This gives an exponential amplification of the
specific DNA fragment. These fragments are amplified to levels which become detectable within the PCR system.

### 3.3.7 Sample Detection and Threshold Cycle

For real-time PCR, data is collected throughout the PCR process rather than at the end of the PCR process (end-point PCR) and reactions are characterised by the point in time during cycling when target amplification is first detected. Fluorogenic probes or DNA binding dyes are used for fluorescence monitoring of the amplification. Binding dyes, such as SYBR Green I, are non-specific dyes which bind to all ds-DNA and are used in wide range of applications. However, as specific fragments are being targeted in this study, fluorogenic Taqman® probes which provide increased specificity and sensitivity are used. A reference probe is also incorporated to account for dye bleeding or other inconsistencies across the reaction samples. During a PCR run, the intensity of fluorescence generated by the probe is measured and used to quantitate the amount of newly generated double-stranded DNA strands. The rate of fluorescence accumulation which binds to the specific DNA fragment is dependent on the starting quantity of the template. The fluorescence accumulates as the reaction is thermocycled and is measured at the end of each cycle during which background noise levels within the system are accounted for and are normally determined from any static noise in detection plus the average of the first three no-template-control (NTC) droplets in a run.

The cycle number at which the intensity of fluorescence generated rises above background noise is known as the threshold cycle (Ct) or as described in the MIQE guidelines (Bustin et al., 2009), the Cq (quantification cycle). Standard threshold values are set automatically by the system at 0.2 (~10-standard deviations above background noise) or if background noise is minimal, the threshold can be reduced to 0.1 for more sensitivity. The Cq value is inversely proportional to the amount of nucleic acid in the starting sample and is typically read at the beginning of the cycle’s exponential phase on the s-curve graph ([Figure 27](#)). At this point, in theory, the template doubles at each cycle but is based on PCR chemistries and instrumentation efficiencies. A greater amount of template of that gene available at the start of the PCR results in an earlier detection rate of the signal or a lower Cq. Higher Cq levels indicate lower amounts of target nucleic acid and may also indicate
no detectable target in the sample. Typically, Cq values between 15 and 35 indicate an abundance of nucleic acids while values above 35 are excluded as they can be misinterpreted as amplification but can actually be the source of a possible contamination or noise within the latter phase of the reaction. The remaining reagents, polymerase and accumulated inhibitors can be the source of deviations during the final cycles.

![qPCR Curves](image)

**Figure 27**- PCR s-curve showing the accumulation of fluorescent dye. The earlier the signal is detected signifies a greater amount of template in the starting quantity for the sample. The sample indicated by the green line in the above graph is in greater abundance than the purple sample as the purple sample is only detected at a much later cycle. The cycle number is based on the threshold which is applied above background noise levels.

### 3.3.8 Data Normalisation

Relative quantification is based on the expression levels of a target gene versus one or more reference gene(s). The use of reference genes accounts for technical variations between sample replicates such as RNA input concentrations, reverse transcription yield, uneven loading, amplification efficiency, enzymatic activity and variation within experimental conditions (Rubie et al., 2005, Pfaffl, 2001). Normalisation of gene expression data is used to correct sample to sample variation and reference genes are used to normalise expression levels throughout the experiment as their expression is consistent between normal and cancer samples. Reference genes provide the core stability when analysing variations in gene
expression levels between samples. However, in different conditions, for different organisms and even tissues the constant level of the expression is not maintained for many reference genes (Kozera and Rapacz, 2013). It is important to select genes with suitable characteristics to be built into studies and to assess cross-condition variability before the true experiment commences. Ideal features of reference genes should include stable mRNA expression across a range of tissue/cellular samples and experimental conditions. Experimental procedures should not lead to varied expression of reference genes and it is imperative that any variations be identified to achieve correct quantification of gene of interest among different samples.

Originally, over 90% of RNA transcription studies in high impact journals used only one reference gene (Suzuki et al., 2000) which lead to inconsistent results and amplified fold changes which were unreplicable in follow-up studies. It is documented that normalisation of data with a single reference gene can lead to inaccurate interruption of results (Tricarico et al., 2002). More recently, the necessity of using multiple stable reference genes has become evident (Vandesompele et al., 2002, Dheda et al., 2004) to reduce technical variation between experiments throughout the qPCR workflow and to show relatively small fold-changes in RNA levels. Reference gene selection software, such as NormFinder and geNorm (MOMA, Biogazelle®) (Hellemans et al., 2007) is now highly regarded when choosing reference genes and assist with results generated to be MIQE guideline compliant (Bustin et al., 2009). Selecting between 2-5 reference genes for normalisation can achieve optimal normalisation for studies can reduce variability and reduce erroneous quantification.

**Colorectal:** Recent publications demonstrated that 97% of RT-qPCR studies on colorectal cancer contained information that was unreliable (Dijkstra et al., 2014) due to common errors in qPCR result reporting methods. When examining expression patterns on clinical samples, unstable reference genes could lead to the incorrect reporting of a dysregulated genes in cancerous tissue and invalidate data retrieved from experimentation. Ideal qPCR experimental design merits the testing of a panel of candidate reference genes and selecting the most stable for the type of sample being analysed and experimental conditions. Research to identify suitable reference genes was performed by a colleague (Dowling et al., 2016b). In the study,
nine reference genes were examined across 2D and 3D cell cultures and normal and malignant colon tissue. From the results of this research, Phosphoglycerate Kinase 1 (PGK1), Glucuronidase Beta (GUSB) were ranked as the most stable genes between normal and cancerous cell lines and colorectal tissue. Therefore, PGK1 and GUSB were used as primary reference genes. Peptidylprolyl Isomerase A (PPIA) and Hypoxanthine guanine phosphoribosyl transferase (HPRT) have been showed previously to be stably expressed in a large cohort of colorectal tissues (Kheirelseid et al., 2010a, Sorby et al., 2010) and so were selected to be included also. Utilisation of these four validated, stable expressed reference genes will add increased robustness to forthcoming profiling data in comparison to current published research in the field. The stability values as calculated by Normfinder are presented in Table 3.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Stability Value</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGK1</td>
<td>0.152</td>
<td>0.14</td>
</tr>
<tr>
<td>GUSB</td>
<td>0.307</td>
<td>0.148</td>
</tr>
<tr>
<td>PPIA</td>
<td>0.412</td>
<td>0.203</td>
</tr>
<tr>
<td>HPRT1</td>
<td>0.547</td>
<td>0.104</td>
</tr>
<tr>
<td>RPLP0</td>
<td>0.614</td>
<td>0.109</td>
</tr>
<tr>
<td>B2M</td>
<td>0.929</td>
<td>0.099</td>
</tr>
<tr>
<td>PMM1</td>
<td>1.004</td>
<td>0.138</td>
</tr>
<tr>
<td>TBP</td>
<td>1.472</td>
<td>0.268</td>
</tr>
<tr>
<td>ACTB</td>
<td>2.002</td>
<td>0.097</td>
</tr>
</tbody>
</table>

Table 3 - Reference gene stability in colon tissue samples (Dowling et al., 2016)

Breast: PPIA and ACTB have been showed to be two stable reference genes in basal breast cancer cells and were therefore chosen to normalise expression values in the breast tissue study (McNeill et al., 2007). Again, as a primary aim in this study is to show relatively small fold-changes in RNA levels additional reference genes were analysed to be included. Further to this, RPLP0 and PMM1 were also chosen to add improved robustness to the results (Lyng et al., 2008). A list of selected reference genes for breast tissue experimentation can be seen in Table 4.
<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Gene Symbol</th>
<th>Location</th>
<th>Assay ID</th>
<th>bp</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phosphoglycerate Kinase 1</td>
<td>PGK1</td>
<td>Xq13.3</td>
<td>Hs00943178_g1</td>
<td>73</td>
<td>(Dowling et al., 2016b), (Krzystek-Korpacka et al., 2014), (Sørby et al., 2010), (McNeill et al., 2007),</td>
</tr>
<tr>
<td>Peptidylprolyl Isomerase A</td>
<td>PPIA</td>
<td>7p13</td>
<td>Hs04194521_s1</td>
<td>97</td>
<td></td>
</tr>
<tr>
<td>Glucuronidase Beta</td>
<td>GUSB</td>
<td>7q21.11</td>
<td>Hs00939627_m1</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td>Hypoxanthine guanine phosphoribosyl transferase</td>
<td>HPRT</td>
<td>Xq26.1</td>
<td>Hs02800695_m1</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td>β-Actin</td>
<td>ACTB</td>
<td>7p22-p12</td>
<td>Hs01060665_g1</td>
<td>63</td>
<td>(Gur-Dedeoglu et al., 2009, Liu et al., 2015)</td>
</tr>
<tr>
<td>Phosphomannomutase 1</td>
<td>PMM1</td>
<td>22q13.2</td>
<td>Hs00160195_m1</td>
<td>111</td>
<td>(Lyng et al., 2008, Sharungbam et al., 2012).</td>
</tr>
<tr>
<td>Ribosomal Protein, Large, P0</td>
<td>RPLP0</td>
<td>12q24.2</td>
<td>Hs02992885_s1</td>
<td>98</td>
<td></td>
</tr>
</tbody>
</table>

Table 4- Reference genes used for normalisation

### 3.3.9 Expression Fold Change

The comparative Ct method is a mathematical model that calculates gene expression changes as a relative fold difference between an experiment and a calibrator sample (Livak and Schmittgen, 2001). When comparing the expression of the normal tissue to the malignant tissue, the amount of target is calculated relative to the calibrator control sample and also normalised to an endogenous reference control. The difference between reference gene expression when analysed in the normal (RG\text{normal}) and cancer (RG\text{cancer}) samples should, in theory, always be equal to zero as the reference gene is selected based on its invariability across multiple samples. Therefore;

\[ \Delta RG = RG_{normal} - RG_{cancer} = 0 \]

The relative change in expression for each sample, \( \Delta Ct \), is identified relative to the
Materials and Methods

calibrator and normalised to the geometric mean of four internal experimental reference genes, and is given by:

\[ \Delta C_t = C_{t, \text{sample}} - C_{t, \text{reference}} \]

where \( C_{t, \text{sample}} \) and \( C_{t, \text{reference}} \) are the Ct values for the sample and reference gene respectively. Therefore, two \( \Delta C_t \) values, one for the normal tissue sample (\( \Delta C_{t, \text{normal}} \)) and one for the cancer sample (\( \Delta C_{t, \text{cancer}} \)), are calculated. The \( \Delta \Delta C_t \) is then the difference between these:

\[ \Delta \Delta C_q = \left[ (C_q \text{ target gene }_{\text{normal sample}}) - (C_q \text{ reference gene }_{\text{normal sample}}) \right] - \left[ (C_q \text{ target gene }_{\text{cancer sample}}) - (C_q \text{ reference gene }_{\text{cancer sample}}) \right] \]

The fold change can then be calculated by:

\[ \text{Fold change} = 2^{-\Delta \Delta C_t} \]

3.3.10 Statistical Analysis

Statistical analysis is an essential step in the quest for confidence in data. This essential step in experiments has accelerated to keep up with the demand for more accurate outcomes in the fight against a range of diseases, especially cancer. For this study, data was analysed using a series of statistical methods to determine confidence levels in results. The primary package used was IBM SPSS Statistics 21 (SPSS Inc.) which uses empirical statistical methods to analyse the data to determine statistically relevant results in relation to the diagnostic study undertaken.

For the colorectal experimentation study, the mean of triplicates was used for statistical analysis and the distribution of the data was tested for normality using histograms for both normal and tumour samples and also using the Shapiro-Wilk test as the sample size is less than 50 samples. This test tests the null hypothesis that a sample \( x_1, \ldots, x_n \) came from a normally distributed population. Box and whisker diagrams were plotted as they are a convenient way of graphically depicting groups of numerical data through their quartiles. From these plots, outliers were determined and inspected in the corresponding data sets. Parametric tests (paired-sample student t-tests) and non-parametric tests (Wilcoxon signed-rank log test) were used to assess related samples based on whether the distribution of the data was normal or not normal. If homogeneity of variances was violated, then the both parametric and non-
parametric tests were performed to assess effect variations. A two-tailed p-value of <0.05 or less than 5% was considered statistically significant. Pearson's correlation coefficient and Spearman's rank correlation coefficient are two tests which are used to identify correlations. Pearson’s product-moment-coefficient correlation (or Pearson’s r) is a statistical metric that examines the strength of the linear relationship between two variables where as Spearmans rho (s) assumes a relationship exists between the two variables that does not fit a linear or non-linear pattern and is therefore less restrictive. In principle, it is a relationship that has the following characteristics; as the value of one variable increases, so does the value of the other variable; or as the value of one variable increases, the other variable value decreases. A monotonic relationship can also be defined as a linear one so it is important to investigate all possible relationships to choose the most appropriate one to potential identify co-expressed genes which may be biologically associated. For the breast tissue profiling, a commercial software tool (Relative expression software tool (REST®)) (Pfaffl et al., 2002) was used to perform analysis of fold changes in this research. The programs mathematical algorithm been developed to compute the expression ratio based on PCR efficiency and Cq deviations. The expression ratio results of the investigated transcripts are tested for significance by a Pair Wise Fixed Reallocation Randomisation Test © and plotted using standard error (SE) estimation via a complex Taylor algorithm negating the need for additional calculations.

3.4 Chapter Closure

The PCR sample constituents are first outlined in this chapter under the Materials section. Following this, the acquisition of clinical samples, extraction process and sample preparation techniques have been outlined. The PCR components are explained and the PCR process itself is outlined in detail along with the sample loading protocol and normalisation using internal control genes. Data analysis is then focused on with reference to generating the Ct value from the raw fluorescent signal. To conclude, an overview of the statistical significance determination required when examining the data produced is presented. This microfluidic platform incorporates all of the required processing steps for the complete analysis of samples. The reaction is created, thermocycled, amplified and analysed with minimal biological variants, handling or user interaction leading to a complete sample-to-Ct instrument.
CHAPTER 4

INSTRUMENTATION DEVELOPMENT
4 Instrument Development

4.1 Introduction

Improvements in developing microfluidic PCR devices in the last two decades have been very significant and therefore the utilisation of the technology for large-scale biological experimentation has been very much in-demand. In this chapter, current commercial microfluidic technologies that are used for genetic analysis are firstly discussed to allow for a full comprehension of such systems. The development of a quantitative-PCR continuous flow instrument is then discussed, highlighting the assembly of the platform in detail conducted as part of this research.

4.2 Microfluidic Technologies

Advancements in genomic profiling technologies, the targeting of specific molecules and improvements in validation techniques have all played a role in creating an improved outcome and survival for people who are diagnosed with cancers and other life threatening diseases. Microfluidics has some key advantages such as high throughput experimental flexibility and low consumption of reagents (Zhang and Ozdemir, 2009). The market for microfluidics technologies, which are defined, in general terms, as devices that handle very small (e.g., micro- or nanoliter) volumes of fluids, was valued at nearly €4.5 billion in 2011 and €4.9 billion in 2012 and total market value is expected to reach over €9 billion in 2017 (McWilliams, 2013). Utilising microfluidic approaches provide numerous advantages for DNA amplification (Zhang and Xing, 2010, Chang et al., 2013, McCalla and Tripathi, 2011) and other diagnostic applications (Eicher and Merten, 2011) such as economies of scale, parallelisation and automation, and increased sensitivity and precision that comes from small volume reactions. In addition to this, microfluidics is not only the ability to perform high-throughput sample analysis but also the ability to perform experiments out of the range of existing technologies (Kelly et al., 2007, Taly et al., 2007). PCR is the dominant method of choice for quickly generating a sufficient amount of identical genetic material for analysis in biological investigations. Other existing methodologies such as RNA microarrays (Chang, 1983, 2006), digital PCR (Sykes et al., 1992, White et al., 2013) and next generation DNA sequencing are extremely powerful tools for studies in that they allow one to
probe virtually the entire transcriptome to give an overall picture of gene expression behaviour (Spurgeon et al., 2008, Heredia et al., 2013) and genetic mutations (Vogelstein and Kinzler, 1999).

4.2.1 Chip Based Platforms

Devices which utilise manufactured chips from various substrates for biological reactions attract great interest for many different reasons. Small-volume, compact and fast integrated fluidic circuits (IFCs), or as they are more commonly known ‘Lab-on-a-chip’ systems, have made significant advances in recent years (Devonshire et al., 2013, Day et al., 2013, Zhang et al., 2006). Chip substrates can be manufactured from a range of materials but key properties such as thermal conductivity and biomolecule absorption rates are primary factors. Silicon and glass substrates were originally chosen due to good rigidity and strength (Zhang and Xing, 2007) but today polymer substrates such as polydimethylsiloxane (PDMS) allow for the manufacture of precise and complex chip structures.

![QuantStudio 3D digital PCR System and BioMark HD® System (Fluidigm Corporation)](image)

**Figure 28** - The QuantStudio 3D digital PCR ® System and BioMark HD® System (Fluidigm Corporation)

Functional components and product detection methods are integrated into the chips to allow for complete biological analysis. Commercial dPCR microfluidic platforms such the BioMark HD® System (Fluidigm) and QuantStudio 3D digital PCR ® System (Thermo Fisher) (Figure 28) use microfluidic technology combined with IFCs to generate and analyse partitioned samples and enable high throughput gene expression measurement (Morrison et al., 2006, Dixon et al., 2009). This facilitates thousands of assays to be performed in parallel. These technologies allow for the screening for multiple potential biomarkers and drug target but are sometimes
limited by the necessity to pre amplify targets which has been suggested to introduce amplification bias (Ginsberg, 2005). Also, limitations in dynamic range make the detection of transcripts in low abundance problematic (Marioni et al., 2008) and their results for any given gene are often ambiguous due to the system noise interference (Liang, 2007).

More recently the SmartChip system (Wafergen) (Figure 29) targets the gap between hybridization-based microarray technology and PCR. The platform combines the high throughput nature of microarrays with the sensitivity, precision and dynamic range of quantitative real-time PCR by combining a nano-dispenser module and a PCR unit which performs thermal cycling and data collection in a complete stand-alone unit. The chip can also be used as a next-generation sequencing target enrichment system which further increases its versatility in the laboratory. Integrated systems like this that enable a flexible workflow for high density gene expression profiling allow for more genetic analysis to be performed with more selective test variation and more complex interactions to be probed.

Despite many benefits such as economies of scale and functionality, there are many limitations to microchip technology: Some materials have been shown to inhibit the PCR, absorb PCR products and require thermal insulation which increases overall complexity(Zhang and Xing, 2007). Many systems do not employ real-time

Figure 29- Four-quadrant chip. Each quadrant contains 841 nanowells. One chip can process 1, 2 or 4 samples for a maximum number of 3364, 1682, and 841 amplicons, respectively (Wafergen SmartChip®)
detection, and thus require additional methods or devices such as gel electrophoresis for the quantification of amplicons (Koo et al., 2013). Furthermore, the cost of microchip manufacturing can be high with silicon and glass being the most expensive to fabricate.

Advantages:
- Simultaneous analysis of multiple genes for novel biomarker discovery
- Small chip size with large number of wells allows for high-throughput screening.
- Chip fabrication techniques constantly improving.

Disadvantages
- Chip design and manufacture can be expensive.
- Relatively short shelf life
- Limited dynamic range remains a challenge.

4.2.2 Droplet-based platforms

![Droplet-based platforms](image)

**Figure 30**- This figure reflects publications indexed within all databases for the search term “droplet microfluidics” (Chou et al., 2015). Publications have increased substantially since 2009 showing the increasing popularity of microfluidic applications for a range of industries.

Droplet-based microfluidic technologies have maintained a strong influential presence in the biological analysis market and this is boosted by current demands for
higher throughput processes while using reduced sample quantities. Publications in the field of droplet microfluidics for a range of industrial applications is increasing year on year (Figure 30), with over 500 “droplet microfluidics” articles published in 2014 (according to ISI Web of Knowledge), highlighting the current upward trend in the utilisation of microfluidic technology in industry. Microfluidic technology provides some advantages over current technologies in the field of genetic analysis such as the implementation of nanoscale reactors, minimisation of sample absorption and enhanced mixing and mass transfer inside droplets (Zhu and Fang, 2013). In addition, microfluidic systems can have high flow rates and which can maximise the potential for increased throughputs. Furthermore small volumes allow several thousand PCR reactions to be performed in parallel (Nakano et al., 2003, Curcio and Roeraade, 2003) or a continuous-flow based approach in which the temperature is kept constant over time at specific locations in the system, with the sample moved between the individual temperature zones for cycling (Liu et al., 2002, Nakayama et al., 2006). This approach has a smaller thermal inertia because only the PCR mixture needs to be heated rather than an entire chip (Zhang and Ozdemir, 2009) or well-plate in different systems. Additionally, the challenge associated with realising the desired economies of scale in microfluidic devices is to simultaneously reduce the number of pipetting steps while amortising the sample volume from each pipetting step over a large number of independent assays (Liu et al., 2003). The type of instrument that the end user chooses ultimately depends on the specific application that will be carried out and the complexity of the experimental design which is being undertaken. Many microfluidic systems for PCR which have been developed (Kiss et al., 2008, Chabert et al., 2006, Mazutis et al., 2009) offer many advantages such as expedited biological/chemical analysis, higher throughput capabilities and reduced consumption of samples and reagents (Zhang and Jiang, 2016).

Current standard PCR systems are well-based systems where the samples and reagents are mixed offline before being loaded into the PCR system in the well plate. The PCR system then cycles the temperature of the samples in the wells. Droplet-based microfluidic systems are created by two immiscible phases, typically aqueous droplets held within a non-aqueous, immiscible, carrier fluid such as silicone or fluorocarbon oil (Markey et al., 2010). This creates a wrapped droplet which acts as a bioreactor in this system.
One reactor set consisting of five aqueous microfluidic plugs is illustrated in Figure 31. The set comprises of four reactor droplets with the sample of interest to be assayed and one ‘no template control’ (NTC), a PCR reaction without the presence of a template. The droplets produced are usually in the microliter range and can be produced at rates of tens of thousands per hour. Wrapped droplets in small channels also allow fluid flows with no dispersion; therefore, contamination will not be left in the tube for the next droplet to pick up. In addition, when such droplets are surrounded by an immiscible fluid this prevents contact between the surfaces of the channels and the sample within the droplet, eliminating adverse effects due to the large surface to volume ratios (Markey et al., 2010). The combination of droplets with further reagents and chemistries can also be exploited by matching density and surface tension properties of fluids in a controlled environment. This adds further reasoning to utilisation of this type of technology to perform complex biological experimentation.

Advantages:

- Amplifying a specific DNA sequence using droplet PCR is faster and less technically difficult than previous methods.
- Increased dynamic range of detection and has been shown to detect as low as two fold change.
- Multiplex analysis of different targets possible.
Disadvantages

- PCR product increases exponentially which can lead to variation when transformed to linear range.
- Can require expensive equipment initially and reagents.

In conclusion, technological advances in the field of genetic research have advanced significantly in recent years. Vast quantities of data and information regarding a person’s genetic profile, SNPs and germline mutations can be achieved from relatively small quantities of sample and also smaller volumes of reagents. In fact, some technology’s rapid progress is so advanced that it has now out-paced computational processing as predicted by Moore’s law (Shokralla et al., 2012) which in itself, becomes an additional challenge for significant clinical outcomes to be achieved from the research. Whole genome sequencing is an example of this and needs further review in the area of data analysis and processing to fully utilise its benefits. However, to combat this, it has become common practice to check the results of a genome-wide study with RT-qPCR which has excellent sensitivity, dynamic range, and reproducibility and is widely regarded as a “gold standard” measurement (Spurgeon et al., 2008). Considering the current use of qPCR for molecular microbiological testing in the clinical laboratory, high-throughput RT-qPCR devices are also likely to be at the forefront of transcript-based diagnostics in the near-future (Devonshire et al., 2013).

4.3 Instrumentation Characterisation

4.3.1 Introduction

- This section presents research to advance the development of the core microfluidic and amplification devices for the micro total analysis system (µTAS). Once complete, a genetic diagnostic platform capable of high throughput DNA amplification and detection will be assembled for the identification of genetic biomarkers in a range of oncogenic diseases and other chronic illnesses.

The diagnostic instrumentation developed for this thesis expands on the approach of using microfluidic droplets, acting as distinct miniature reactors continuously flowing through the system, from which it will be possible to quantify gene
expression levels in DNA samples. Key engineering and biological methods are implemented to create a stand-alone genetic diagnostic system capable of combining precise fluidic control with specific molecular analysis technologies. Early prototypes were designed to be reusable without carryover contamination (the unwanted transfer of molecular material from one reaction to another) that may ultimately lead to a false positive by encapsulating the droplets by immiscible oil. This platform is also based on a continuous flow concept, where a large number of reactions will be generated with small reaction volumes leading to a fundamental requirement for a system that is capable of manipulating such small samples.

A number of engineering and biological challenges need to be addressed;

- The entire process from reactor droplet generation to data acquisition needs to be scaled and automated to ensure accuracy and consistent measurements.
- High throughput capability needs to be an important consideration as a large number of samples are expected to be assayed. Each reaction will be performed in quadruplicate. Therefore, assaying the expression of e.g. 50 genetic markers in 30 samples necessitates 6,000 reactions. If the samples are matched patient samples, this figure becomes 12,000. The addition of 1,000 negative control reactions brings the minimum total number of reactions to 13,000.
- As the number of reactions required will be large, sample volumes need to be small to maximise system capabilities and efficacy. Final reaction droplet size is expected to range in the 250-350nl range.
- The reactor droplets need to be reliably encapsulated in the carrier fluid and no carry over contamination should occur between the droplet and surfaces such as mixing capillaries or conduit boundaries.
- Cost effective compared to current qPCR commercial platforms.

The design processes of some of the primary instrumentation modules were completed prior to the commencement of this research. These components were used for the development of a prototype genotyping platform driven towards the detection and characterisation of Single Nucleotide Polymorphisms (SNPs) in agricultural crops. The instrumentation was also a droplet based platform with a high throughput
capability of approximately 18,000 data points per hour processed over 96 instrument lines. A unique feature of this micro-total analysis system was droplet manipulation, stable thermal amplification of plant DNA fragments and optical detection of the minute fluorescence of the binding dyes which are necessary for the PCR was performed on one instrument.

These characteristics were desired to be included in a new platform which could potentially be used for RNA quantification studies. RNA quantification measures the amount of RNA present for specific genetic targets which may influence the progression of a disease. Would it be possible to analyse RNA, a molecule which carries genetic information from the nucleus to the sites of protein synthesis within the cytoplasm of the cell with this technology? Could the technology be used to discover biomarkers in mammalian specimens? To fully understand if a platform based on the technology could perform the biological experiments, the design, development and operation of some of the previously designed key components are

**Figure 32-** Schematic of the GEI (Gene Expression Instrument) design showing primary components and overall process flow. The main components of the instrument comprise of liquid bridge technology, fluidic handling systems, an integrated pumping system, thermal cycling module and optical detection platform.
outlined in this chapter. Using some of the components, the new platform is assembled as a four-line system which allows for a range of complex experiments to be carried out in parallel. It again uses a flowing reactor droplet concept where an organic carrier fluid will convey aqueous sample mixtures, which are mixed with a suitable inter-calculating dye, through a micro-analysis system (Figure 32). The entire process, from sample preparation to data acquisition, is performed on one platform and is automated to ensure speed and consistency in measurement. The instrument consists of three key functional modules which have been adapted from previous instrument iterations. A key adopted module is the droplet dispensing module or liquid bridge which allows for the creation of the bioreactor and combination of a number of micro droplets of varying constituents required for successful nucleic acid amplification. The second adopted module is the thermal cycling module which is required to heat the sample for the PCR. The development of the dual temperature segments and thermocycling parameters are outlined. Thirdly, the optical detection system is described and the key features used to achieve bio-fluorescence are outlined. In addition to these key modules, optimisation of the pumping system design, fluidic balance and fluorescence detection improvements were enhanced as part of this thesis and discussed throughout the chapter. In conclusion, a discussion of the assembly of a novel platform and hypothesised approach to perform continuous high-throughput RNA biological screens is presented.

4.3.2 Droplet Generation

- The method for the generation of the micro-reactor droplets for this research is described for final reactions of approximately 300nl in size.
- Fluidic manipulation designs previously developed are presented explaining mixing of individual droplets of DNA, reagents and TAQ polymerase occurring in a controlled micro environment.

The first step in the microfluidic life cycle of a droplet is its production. Intricate methods have been developed to accurately generate low volume, precise droplets which can be reproduced quickly for a wide variety of experiments. Droplet size is determined by the application requirement and is predominantly in the nanolitre
scale but pico- and femtolitre droplets have been demonstrated to perform single-cell and single-molecule analyses (Leman et al., 2015).

The development of micro-droplet uses either active or passive methods (Zeng et al., 2013). In passive methods, the droplet generation depends on the geometry and dimensions of the device. Microfluidic t-junctions and flow focusing devices are common passive methods which can achieve a uniform and evenly spaced continuous stream of droplets. The drop size is usually controlled by tuning the flow rates of the continuous and the dispersed phase (Gu et al., 2008). However, dynamic equilibrium can vary and change the size distribution which can hinder the prediction of the flow behaviour. There are also many limitations with these methods such as droplet polydispersity in droplet streams and flexibility in manipulation of droplets.

The active method for generation of droplets involves the use of an external factor such as an electric field for droplet generation. Two techniques that fall in this category are di-electrophoresis (DEP) (Fan et al., 2009), which applies a non-uniform electrical field to produce an electrical force upon a droplet, and electro-wetting on dielectric (EWOD) (Pollack et al., 2011) which relies on the change in contact angle of a droplet on an electrically charged substrate. Using these microfluidic techniques, droplets can be generated and transported from one charged region to another easily and make the technology very appealing in the field of digital microfluidics. However, as these two techniques exploit electrical fields to generate droplets, limitations can occur in the control of interfacial instabilities where the surface wettability (contact angle) can be difficult to manipulate and control.

In the instrumentation, a capillary-based system that precisely generates specific, stable droplets without contamination between neighbouring droplets is delineated. Droplets of aqueous solution surrounded by a fluorinated carrier fluid provide a simple platform for manipulating samples with no dispersion or losses to interface Micro-scale polytetrafluoroethylene (PTFE) capillary tubing is used as the flow conduit in the qPCR instrument. PTFE is a chemically inert thermoplastic polymer that is commonly used in biological applications. It has a low coefficient of friction which allows the oil carrier fluid and aqueous droplets to pass smoothly through the tubing with minimal resistance or backpressure (Markey et al., 2010). Also, as PTFE is a hydrophobic material, strong wetting of the surface walls by the carrier fluid is
produced and the internal tolerances on the capillaries prevent the droplets from sticking to the walls on imperfections as seen similar PCR designs (Dorfman et al., 2005). The tubing (ID ~150um, OD ~712um) is integrated into dipping heads that are attached to modular 3-axis robotic stages (CMMS-AS Motor Controllers, FESTO, Germany) which move in pre-programmed steps aspirating different samples and reagents from different wells. The dipping head flexibility allows for access to each individual well on a variety of plates including 384 well, flat-bottom, micro plates (Figure 33).

![Diagram of microfluidic dipping process using PTFE tubing.](image)

**Figure 33** - Microfluidic dipping process using PTFE tubing. The tip is initially submerged in an oil overlay continuously drawing oil into the system at a predefined flow rate. The tip is moved into the sample (b) the pump draws sample into the tube and towards the liquid bridge. Tip returns to the oil (c) creating the wrapped droplet. (Dalton, 2011)

The robotic stages programmed sequence allows dipping into the solutions and aspirate predefined quantities of DNA and reagents. This allows versatility in the instruments robotic movements and allows the sample side to remain in the well while the multilumen head moves from well to well aspirating the genetic marker targets. This process can also be reversed where the multilumen head remains in the well while the single tip side moves from well to well aspirating a range of samples. On the single tip side, four individual dipping heads aspirate identical volumes of DNA sample, while a second, separate bundle of four dipping heads aspirate the reagent and Mastermix volume (Figure 34).
Droplets are created by moving the dipping heads from an oil overlay in the microplate to the sample for a specific time. The sample is withdrawn and the dipping head moves back to the oil overlay, creating the wrapped droplet. The droplet volume is determined by the dwell time of the dipping head and the volumetric flow rate. Accurate and precise incremental dipping accounts for the reduced time the aspiration tip is in the sample due to the reduced height level of sample. Initially the system is primed with density-matched oil. To combine and mix the sample with the reagents within one droplet, which may then be used as a microfluidic biochemical reactor, liquid bridge technology that has been developed (Curran, 2005) is exploited.

Figure 34- a) Single tip (individual sample dipping) and b) Multilumen (bundled reagent dipping) heads
A liquid bridge is formed when a volume of fluid bridges transiently between two opposing capillaries (Forget, 2009) and a concept of the design can be seen in Figure 35. The bridge comprises a first inlet port of approximately 300µm in diameter which is at the end of a capillary, two narrower inlet ports each perpendicular to the first port, an outlet port (also of 300µm) which is at the end of a capillary and a chamber for silicone oil. The chamber where the droplets combine is ~5mm in diameter and ~3mm in depth. The two phase nature of microfluidic flow necessitates a means of combining fluids where one phase (the aqueous) forms a ‘bridge’ between the capillary tips while the other phase (carrier oil) inflicts instability in the bridge. The oil is density-matched with the reactor droplets such that a neutrally buoyant environment is created within the chamber. The separation of the droplets by the oil phase, which is directly proportional to the ratio of oil to aqueous flow rates, guarantees the droplets act as individual micro-reactors (Dalton, 2011) and also continuously replenishes the oil in the chamber. This causes the droplets to assume a stable capillary suspended spherical form upon entering the chamber. The first and second droplets remain suspended at the end of the capillary until the third droplet enters the system causing the formation of an unstable funicular liquid bridge allowing the spherical shape to quickly rupture and span the gap between the ports, forming an axisymmetric liquid bridge. This method of
periodic, stable mixing allows for the creation of the micro-reactor droplets with DNA sample and reagents for functional experimentation (Figure 36).

After droplets are produced they are transported in the micro-capillary tubing by the carrier fluid. The droplets can be moved, mixed and monitored in an isolated environment with minimum risk of carry-over contamination. In previous work on this type of PCR, cross contamination between droplets was attributed to droplet instability and the formation of small satellite droplets (Curcio and Roeraade, 2003). In single-phase microfluidic PCR devices, PCR mixture can be easily absorbed at flow channel surfaces, which causes PCR inhibition, carryover contamination (Zhang and Jiang, 2016) and chemical absorption of reaction component (Skeggs, 1966). Cross over contamination between successive samples can be augmented further by diffusion of samples through the flow phases. To inhibit this, when droplets are produced they are transported in the micro-capillary tubing by the carrier fluid where the interfacial properties of the carrier fluid coupled with the constant flow of fluid into the system separating the droplet trains means carry-over contamination risks are reduced. This is due to the biphasic flow in the capillary in flow regime where a liquid film exists between the droplets and capillary wall. The oil film prevents the aqueous droplets from wetting the wall of the tube, preventing contamination of the flow conduit (Morris, 2008). Droplet contamination prevention is discussed further in Chapter 5.
4.3.3 Fluidic Control and Conduits

- The alteration of the pumping system for this research from a 96-line platform to a more stable 4-line system is outlined. Further modifications include the addition of manifolds and a priming system.

Precise fluidic control is essential to achieve successful continuous-flow droplet PCR. Droplets are in constant motion through the system so fluid flow rates are a primary focus. Dwell time of a droplet in a thermal zone to have efficient amplification is highly dependent on the flow regime throughout the system and is also the dominant influence in determining droplet size. Therefore, the tubing geometry and internal surfaces are critical to meeting the design requirements for microfluidic PCR. In a micro-channel, the reduction of sample size accelerates the completion of chemical reactions and improves heat transfer due to shorter diffusion length and an increased surface to volume ratio (Markey et al., 2010). The reduction of sample size also offers safety benefits, including containment where some potentially hazardous materials are being used. For biphasic capillary flow, the minimum droplet volume is the volume of a sphere with radius equal to the internal tubing radius (Dalton, 2011). Droplets which are smaller than this are more difficult to manipulate and can be more prone to contamination between surfaces and the conduit wall. In the system outlined in this thesis, fluidic pumping is delivered using micro-annular gear pumps (HNP Mikrosysteme, Germany) (Figure 37). This active, integrated method provides a flow rate by displacing fluid between and internal and external rotor. High precision, low volume dosage, compact design and an integrated microcontroller are key advantages of using this device. With the ability to operate at a flow rate in the range of 0.003-18ml/min coupled with a low pulsation of <6%, the micro-annular pumps are ideal for the continuous, stable convection of droplets through the system. This system enables independent control of fluidic modulation to achieve active control of droplet size and generation frequency by adjusting the pumping displacement which varies flow conditions. Each of the DNA sample and reagent droplets are processed in a strict order to ensure reproducible biological balance is adhered to each time a fluid is aspirated from a well plate.
There are four micro-pumps in the system, each one individually programmed to specific micro-litre flow rates and volumetric displacements. The samples and reagents are aspirated from the micro well plates using one of the pumps (Pump A) which then transports the droplets to the mixing chamber in the liquid bridge. This pump operates at approximately 450rpm, generating a flow rate of 300µl/min for each line. The carrier fluid flow is drawn from the well plate through the bridge, continuously priming the system. A bypass manifold was installed on this line to allow for an improved flowrate and ensures the fluidic imbalance or ‘lag’ is at a minimum in the system. When the micro-droplets are at the capillary ends within the liquid bridge, they are combined into one reaction and conveyed through the system using a second pump (Pump B) operating at approximately 50rpm at a flow rate of 12µl/min for each line. This rate is determined from tubing geometry and thermal cycler dimensions to ensure correct dwell time over thermal zones. Two further pumps (Pump C and Pump D) function as ‘sheathing lines’ which continuously replenish the oil supply that flows over the aspiration tips in the dipping heads. This prevents air from entering the system and allows for the movement of the dipping heads in air between the park position trays and the well plates. This sheathing system also replenishes the oil in the well plates and overlay of the sample. Flow rate monitoring is achieved using a LabVIEW control algorithm which is designed to monitor flowrates on a continuous, real-time basis and has implemented warning systems to notify the user of any deviations outside the predetermined limits. This software was modified from previous instrument iterations for a 4-line system.
Fluidic balance was enhanced in the pumping system by incorporating a range of manifolds through the flow process to compensate for imbalances which may occur within a reduced system (Figure 38). The manifolds also prevent air ingress events from creating flow rate fluctuations and droplet size inconsistencies by trapping air bubbles within predesigned ‘air traps’ which are later removed from the system. Furthermore, the manifolds protect the pumps from air ingestion which has the potential to damage internal gear components and cause power surges leading to experiment complications and/or pump failure.

The pumping system also comprises of key flow intersections regulated by two-way and three-way valves to allow for precise control and high speed direction control of the fluid. The valves are simple in design yet allow for the rapid control of the oil through individual ports and pathways. Volumetric flow rate feedback is provided using a calibrated liquid flow sensor (Sensiron ASL1600-1/ASL1600-20) (Figure 39) yielding accurate, pulseless flow rates even at low operating pressures. Real-time flow rates are relayed to the central control system which in turn adjusts the pump speeds to counteract any possible peaks in flow rate and keep the system overall flow steady and controlled. Post experimentation, the amplified micro reactor droplets flow into a collection manifold which is diverted to a waste storage tank.
To remove any remaining inconsistent droplets and other possibly micro impurities from the system, a priming system was also incorporated into the platform. This allows the system to be purged at the beginning and/or end of each experiment removing any air bubbles or possible contaminants. Priming of the system is dependent on the reversal of one of the system pumps (Pump A) and diversion of either oil or a sterilising fluid into the highest point of the liquid bridge. The liquid then passes down through the bridge and exits at the base of the bridge to be diverted to waste. Liquid is also forced through the experimentation conduits at an increased velocity to perform the same function. A total system prime of the liquid bridge, conduits and associated components can be completed in 6- 8 minutes. The carrier fluid, which is key to droplet development, mixing and convection of reactors through the system, is filtered using a 0.11µm particle filter and located within a basement compartment of the instrument pumping system alongside the waste oil tank. Each reservoir is 10 litres in capacity which allows for approximately 140 hours of continuous instrument operation, highlighting an important walk-away automation aspect of the platform.

4.3.4 Thermal Amplification Process

- The thermal cycler design previously developed is presented. The cycler was re-tubed using PTFE tubing and additional feedback thermocouples were installed for optimised thermal control. Calibration of the system was also performed periodically.
The instrument that employs accurate temperature control and rapid temperature changes to conduct PCR reactions is called a thermal cycler. A large body of research (Zhang and Xing, 2010, Schoder et al., 2003, Neuzil et al., 2006, Belgrader et al., 2001) has focused on developing faster thermal cycling technologies for PCR. A mentioned previously, to carry out the PCR, thermal control of the reaction droplet is a crucial step to maximise the full efficiency of the amplification process. A complete PCR amplification in a conventional PCR device based on a metal block with 20–40 cycles will usually take 1–3hrs because thermal ramp is at the range 1-2°C/s (Zhang et al., 2006, Zhang et al., 2007). Numerous methodologies have been utilised in the field of thermal control for PCR including internally embedded heaters (Morganti et al., 2011, Woolley et al., 1996) thermoelectric devices (Khandurina et al., 2000, Yang et al., 2002) and infrared (Oda et al., 1998, Shaw et al., 2010) with the ultimate goal of providing speed, uniform thermal control and cost effective amplification of DNA. To cool the reaction process, air gaps between the device coupled with fans or Peltier devices have been implemented. These technologies provide several benefits for many applications including increased throughput and improved specificity but engineering trade-offs mean that reaction times are typically more than 25 minutes (Logan, 2009). The majority of microchip systems utilise fabricated micro-, nano- or picoliter reservoirs for conventional thermocycling where the entire chip is heated and cooled. (Marcus et al., 2006, Matsubara et al., 2005, Morrison et al., 2006, Nagai et al., 2001) Some of the reported drawbacks of on-chip amplification are difficulty in creating parallel reactions, difficulty in adjusting the number of cycles once design is complete, and the high cost of such devices. Furthermore, the number of micro-reactors is fixed, reducing experimental flexibility.

An interesting principle which has been exploited is the dynamic “flow-in-capillary PCR reactor”, where a continuous flow of samples and reagents is passed through three zones, which are kept at constant temperatures optimized for denaturation, annealing, and elongation of the DNA fragments. This combats non-uniform temperature fields which may lead to low amplification efficiency of nucleic acids and even non-specific PCR products due to insufficient annealing temperature of the PCR process (Wittwer et al., 1994). The principle was first described where a Teflon capillary was routed through three constant temperature oil baths to provide
30-cycle amplification more rapid than thermal-cyclers with conventional metal blocks (Nakano et al., 1994). The four-line continuous-flow thermal-cycler design utilised in this thesis takes from this principle of a continuous flow microreactor. As these reactors are of the nanolitre range, they undergo equal repeatable conditions for PCR which ensures high reproducibility, and an overall homogeneous process leading to more reproducible results. Thermal cycling of the reaction for this research is performed in a flowing serpentine cycler which gives superior performance because of the relative size of the droplets therefore resulting in more sensitive experiments.

**Figure 40-** Phases of amplification of the DNA target in the thermally controlled environment

Depending on enzymatic kinetics of the chosen chemistries, the PCR amplification (Figure 40) process can either be a two/three step process consisting of consecutive temperature zones of denaturisation, annealing and, if required, additional extension phase. Double-stranded DNA segments are disassembled during the denaturation phase at ~95° into two single stranded DNA molecules. During the annealing phase at ~60°C the sequence specific primer is attached to the single stranded DNA and the polymerization enzyme extends the targeted DNA using deoxynucleotide triphosphates present in the reaction. If the denaturation temperature is too high, the DNA is likely to degrade and conversely if it is too low then the DNA double helical
structure will not fully unwind resulting in poor amplification efficiency. Similarly, if annealing temperatures are too high, the primers will be unable to bind and the target sequence will not amplify. The optimal annealing temperature (Ta Opt) is approximately 5°C below the melting temperature of the primer (Chester and Marshak, 1993), however; high resolution melt-point analysis should be performed to determine optimum annealing temperature if manufacturer recommendations are not provided. Ta Opt for any given primer pair on a particular target can be calculated as: Ta Opt = 0.3 x(Tm of primer) + 0.7 x(Tm of product) - 25; where Tm of primer is the melting temperature of the less stable primer-template pair, and Tm of product is the melting temperature of the PCR product (Rychlik et al., 1990).

Annealing temperatures were recommended by the reagent manufacturer.

![Figure 41- 3D model of thermal module used for activation of enzymes and to thermal cycle the PCR reagents](image)

Accurate thermal control is a key aspect for successful PCR. The adopted design for this research (Figure 41) (Barrett, 2008) is characterised by two independent temperature zones with no cyclic hold times inspired in part by the observation that much of the canonical 20–40 cycle denature (90–95°C), anneal (50–65°C) and elongation (70–80°C) temperature sequence can be abbreviated without sacrificing performance (Bartsch et al., 2015). The extension phase is an additional step to enhance the activity of the polymerase to extend the primers to form ds-DNA and is required if the optimum temperature for polymerase activity is higher that the temperature required for annealing. Extension temperatures normally range between 70°C-74°C. However, by using specific Taqman® chemistries for experimentation (which have been designed to have higher enzyme activity at lower temperatures and
primers with shorter amplicons) can reduce the necessity for this temperature stage. The primers can hybridise to the template at the same time that the polymerase is amplifying the target DNA so both activities can occur at the annealing temperature of ~60°C. This allows for more efficient reactions and for the design of more compact and more effective thermal cyclers in comparison to conventional models. The thermal characteristics are based upon the AB7900HT Fast Real-Time PCR System which is a widely-used PCR platform manufactured by Applied Biosystems. The system specifies a temperature uniformity of ± 0.5°C and accuracy of ± 0.25°C so using it as a benchmark instrument allowed for the same thermal cycles to be replicated in the design of the continuous-flow platform. The AB7900HT system is further discussed in Chapter 5.

Consequently, two thermally-insulated, static blocks are arranged in parallel and constitute as two thermal zones, a 95 °C denaturisation region and a 60°C annealing region with incorporated ~1mm serpentine channels. Since the blocks are isothermal, there is no unnecessary ramping time as seen with most PCR platforms that operate by repeatedly heating and cooling blocks. The elimination of unnecessary ramping time has the potential of reducing the overall thermal cycling time compared to conventional thermal cyclers. The channel length of each zone is purposely designed to define the time duration for the micro-reactors to pass through based at a specific flowrate. Zones are separated by an air gap of approximately 1mm to reduce heat transfer convection rates between the two surfaces. Heating elements (Minco Heat, Minneapolis, USA) are applied to the blocks which allow for fast heating from steady state to elevated control temperature in order to maximize test speed and throughput. These thin, flexible components consist of an etched-foil resistive heating element laminated between layers of flexible insulation. The two temperature zones are thermally insulated to reduce temperature fluctuations along the length and because of this gives superior temperature control due to the reduced heat dissipation capacity. Furthermore, because the PCR mixtures are confined to small droplets in the nanolitre range any temperature change will be achieved uniformly.
The micro-droplet enters the thermal cycling system in a single capillary and travels through the 95°C section for initial denaturisation/dsDNA unwinding of 10 minutes prior to PCR cycling. Similar designs have been demonstrated in literature where sample is flowed through fused silica tubing which is in turn wound helically around a cylindrical device subdivided into discretely heated, constant temperature segments (Park et al., 2003) and which was later expanded upon (Zhang C-S, 2011) with the integration of analysis systems to detect amplification. The micro reactor droplet follows the serpentine shape passing over the two thermal zones and provides for one cycle of PCR for each length of travel along the serpentine design. The serpentine channel length was designed to yield a dwell time duration ratio of 1:3 for denaturation and annealing. The time for each cycle is approximately 60 seconds (15 seconds denaturisation/ 45 seconds annealing) and further dependent on the internal diameter of the tubing (ID ~584µm, OD ~1040µm) (Zeus Inc. Tubing, Orangeburg, U.S.A) and the flow rate (~12ul/min) of the PCR through the system. In total, the sample undergoes 40 thermal cycles before exiting the thermal cycler for sample collection.

Figure 42- Temperature control schematic
From entering the system to exit, the droplet spends a total of 55 minutes in the thermal cycler. During experimentation, temperature deviations were observed to vary a maximum of +/- 0.5°C along the surface of the heating blocks when monitored using four k-type thermocouples integrated along the heating blocks. The thermocouples provide a feedback loop to two proportional-integral-derivative (PID) controllers where the temperatures are pre-set and to a LabVIEW control program to maintain a steady temperature profile. A schematic of this design can be seen in Figure 42. This configuration employs high heat conducting material and efficient thermal insulation providing a homogenous temperature over the block allowing for a compact, efficient and reliable thermal-cycler design and operation.

4.3.5 Detection and Data Analysis System

Detection and analysis of amplified DNA is performed on-line by light-emitting diode fluorescence detection and two high speed spectrometers. The inclusion of fluorescent probes in the PCR reaction mix permits the amplification process to be monitored within individual droplets at specific locations. This previously designed configuration to detect real-time amplification was calibrated and validated post-assembly.

A prerequisite of qPCR experimentation is that the fluorescent signal needs to be monitored at each cycle. Unlike end point PCR, real time qPCR allows for quantification evaluation at any point in the amplification process to accurately predict the quantification cycle or point at which the baseline threshold is passed. A comprehensive data modelling system allows for the signals to be interpreted and expression ratios to be determined. The bio-fluorescence platform described in this thesis analyses the microfluidic reactor droplets with superior accuracy, sensitivity and efficiency for the qPCR by simultaneously detecting, tracking, exciting and detecting different fluorophores in each individual droplet. The optical platform is integrated with the thermal cycler so that as amplification of the template is occurring, the signal is being remotely monitored and recorded. Two dyes are incorporated into the mixed droplet. The first dye, FAM™ (6-carboxyfluorescein) is attached to the probe and releases a fluorophore once the nucleotides bind to the specific sequence. TaqMan® probes consist of an 18-22 bp oligonucleotide probe which is labelled with a reporter fluorophore at the 5’ end and a quencher.
fluorophore at the 3’ end. As the target DNA amplified, this signal increases and a quencher molecule stops the fluorescence from being emitted once the extension phase of the PCR process, in conjunction with the oligonucleotide primers, has completed (Figure 43).

**Figure 43-** Description of the Taqman Probe (a) sequence specific primers and Taqman probe anneal to complementary sequences in the DNA template. The quencher molecule restricts the fluorescent signal from the reporter dye. (b) Extension. Taqman polymerase extends the complimentary strand of DNA (c) While extending, the polymerase cleaves the probe and enables the release of a fluorescent signal by the reporter (McCarthy CE, 2015).

A second dye, ROX (*carboxy-X-rhodamine*) acts as a passive reference dye and is incorporated in the PCR mastermix. The fluorescence levels of ROX have no relationship to the quantity of DNA in the reaction. However, the inclusion of this dye in a reaction provides a constant fluorescent signal during the amplification cycles for normalisation of fluorescent reporter signal in real-time quantitative PCR. The ROX signal normalises for fluorescence variations and it is the increase in target concentration (due to the target DNA sequence been amplified) that is directly proportional to an increase in normalised fluorescence signal. Fluorescence wavelengths for the dyes which are used for this research can be seen in Figure 44.

Dye calibrations were performed to ensure that dye fluorescence was detected accurately by the platform. Concentrations of the dyes were mixed with nuclease free water and passed through the system with the thermocycler deactivated. Fluorescent
signals were then mapped as droplets passed through the system generating a dye calibration curve for each line and applied post-experimentation to normalise for fluorescence detection anomalies that may occur. The platform is designed for additional dyes for multiplex reactions where a number of assays could be targeted at once. However, as only single-plex assays are performed in this thesis, additional signals are not analysed.

Unlike end point PCR, real time PCR allows quantification evaluation at any point in the amplification process. To detect fluorescence from the reaction and subsequent expression levels for each target, the platform uses highly sensitive optical probes. Each probe consists of seven fibre cores each measuring 200µm in diameter. To excite the fluorochrome, LED light travels through six back-light illumination fibres which are integrated in a radial orientation bundle into thermal cycler portholes throughout one length of travel of the capillary through the unit. The LEDs fully disperse the light throughout the droplet ensuring maximum illumination of the fluorophores within the reaction mix.

Figure 44- Fluorescence excitation and emission spectrum for ROX and FAM. ROX dye has a peak excitation and emission of 575nm and 610nm respectively. FAM dye has a peak excitation and emission wavelength of 495nm and 519nm. Dashed lines depict excitation while solid lines depict emission.
Figure 45 Illustration of the optical detection platform showing the individual components designed for optimum fluorescence analysis of each micro droplet as it is conveyed through the system. The design and orientation of the fibre bundle can be seen where the illumination lights are radially positioned around a central detection core. The LED light source, fibre splitter and tiled array are shown. TC cassettes stacked vertically resulting in an array of portholes to facilitate the integration of fibre optics. High speed cameras (1 of 3 shown) are focused on the tiled array (40 x 4) to detect fluorescence from the dyes. The application of masks over the matrix of fibres determines the focus of the cameras ensuring precise fluorophore variation detection.
Once the dye in the reaction is excited, the central single detection core detects the fluorescence from the droplet as it passes the fibre. The optically clear capillaries allow the fluorescence to give a direct, uninhibited signal to be relayed. In total, forty excitation/detection fibre combinations are distribution per line and arrayed into a coherent bundle to a fibre splitter. Fluorescence emission is then transmitted back to a tiled array into which the fibres are positioned in a matrix orientation where it is analysed by two high-resolution digital cameras (*Hamamatsu CCD, ORCA-03G*) (*Figure 45*) which each have a resolution of 1.37 million pixels and capture at a frame rate of 43f/s. The cameras are calibrated to analyse individual fluorescence using a signal intensity mapping system enabling analysis of targets within the sample. Normalizing the fluorescence intensity is accomplished in real-time PCR software by dividing the emission intensity of the reporter dye by the emission intensity of the reference dye.

![Graph](image.png)

**Figure 46** - An example of an s-curve generated by the instrumentation highlighting the detection of Cq's for individual samples. Sample number, amplification cycle and test line tools can be selected from the user interface. Further calibration options are also available.

Modelling software analyses the raw fluorescence (*Rn*) from the *6-FAM* reporter and then normalised against the ROX reference dye. If the sequence is not present, no amplification will take place and therefore only weak background noise will be detected. The fluorescence produced by the sample which is detected by the system
is proportional to the amount of initial targets for the amplification reaction (Bustin, 2005). The resulting signal (ΔRn) is plotted against the number of cycles on a logarithmic scale (Figure 46). A threshold for fluorescent detection is set above the background levels as described in Chapter 3. Once the signal passes the Cq, then, in theory, the sequence of the DNA target doubles every cycle but is dependent on the efficiency of amplification.

4.4 Chapter Close

In this chapter, microfluidic applications in the field of genomic technology were discussed. This stimulated the development of the microfluidic qPCR instrumentation as part of this research (Figure 47) which provides sensitivity and quantitative precision for differentiating gene expression variability between biological samples. The microfluidic characteristics of the instrument are outlined along with how they benefit a qPCR gene expression study. Different modules for the generation and analysis of micro-reactor droplets containing DNA targets and reaction chemistries are described. Some of the components designed in iterations of a similar instrument were assembled into a PCR platform for RNA based studies. These modules include the liquid bridge, a two-step thermal cycler and optical detection unit. Further to this, components such as a redesigned pumping system, improved dipping head functionality and increased thermal feedback loops were integrated into the system. The developed technology serves as the basis for a next-generation gene expression instrument potentially capable of generating more data than traditional systems in the same amount of time – typically 1,000 data points per ~60 minutes, compared with traditional systems that can process 384 plated samples in ~30-90 minutes. While these traditional systems remain an attractive option for many applications, new technology enables breakthroughs that are impossible with current PCR instruments. The system is readily available for further optimisation and automation and serves as a first step toward a high-throughput, quantitative continuous-flow PCR apparatus.
Figure 47 - Photograph of complete instrument showing some individual modules

- PID thermal control module
- Optical Detection Unit
- Optical fibre unit
- DNA dipping head
- Reagent dipping head
- Robotic Control Modules
- Pumping system (at rear)
Chapter 5
INSTRUMENT VALIDATION
5 Instrument Validation

5.1 Introduction

As outlined, a key aim of this research was to develop and test a fully functional continuous flow quantitative real-time PCR platform. The amplification and accumulation of product is quantified as the reaction progresses, with detection of fluorescent signal after each cycle. Proof of concept experiments undertaken prior to gene expression profiling studies are described in this chapter. The chapter validates the instrument capabilities of amplifying a target DNA molecule and the feasibility of the droplet generation and processing techniques required for large scale biological processing. A commercial qPCR platform that was used to benchmark experiments is also presented in this chapter. In comparison to commercial instrumentation, the benefits of using droplet-based microfluidic methods for the qPCR process can include:

- Reproducible, high quality results from cell populations and tissue sample on a constant, systematic basis.
- Significant increases in screening throughput and reduction in reagent usage.
- Expanded possibilities by using multiple reagents and different sample and assay configurations. This leads to allow for more varied experimental design.
- The incidence of user-to-user variation, which is particularly critical when using sub microliter volumes, is challenged by using automated dispensing/aspirating technology and leads to improved data consistency.
- Time and financial savings where the technology produces more data at less cost and in a shorter time frame than standard protocol.

To verify the benefits of microfluidics, a number of experimental testing conditions are examined and assessed to ensure the technology provides stable, reliable data. Gene knockdown tests are carried out showing the reduced expression values of genes post biological treatment. Experimental studies are performed demonstrating successful amplification and fluorescent detection. These experiments are carried out in parallel with commercial based platform to detect concordance. Experiments to
identify the possibility of line to line variance within the four-line qPCR system are then performed. Following this, contamination tests are completed on the system to inspect for possible crossover contamination within the system that may occur. Finally, the stability of selected reference genes for quantitative PCR experimentation in Chapter 6 is discussed.

5.2 Applied Biosystems AB7900HT PCR Platform

A qPCR platform manufactured by Applied Biosystems, the AB7900HT Instrumentation (ABI) (Figure 48) was used to benchmark the quantitative results and to interrogate biological samples. Positive and negative controls were simultaneously performed on the system, replicating thermal zones, fluorophore detection levels and cycle numbers set in the continuous flow system.

Reactions on the commercial platform were pipetted into 384 flat-bottomed plates and sealed with an optical adhesive film (MicroAmp™). Care was taken to ensure bubble formation did not occur inside the well as this will affect fluorescence intensity during thermal cycling. Additionally, the plates were centrifuged at 5000rpm for thirty seconds to ensure sample is not attached to well wall. After this, they were placed on a robotic arm and loaded into the machine where settings for cycling and detection are set. Once the protocol is set, the entire plate is cycled through the temperature zones required for PCR. A 488nm argon laser excites each
of the samples on the plate and the resulting fluorescence signal is detected by a 
CCD camera mounted on a spectrograph after each cycle. Instrument software then 
normalises the fluorescence data, compiles and smoothens s-curves and returns Cq 
values for export and analysis.

5.3 Validation Tests

5.3.1 Assay Efficiency Validation

A major prerequisite of any real-time PCR–based assay is the amplification 
efficiency (E) of the PCR reaction. PCR reactions with low efficiency will have 
lower sensitivity. A sensitive method for assessing if two amplicons have the same 
efficiency is to assess how the Ct values vary with template dilution. Protocols 
suggest examining the efficiency using 10-fold dilutions over 5 or 6 logs; however, 
this is not possible for all targets. The efficiency of an assay was determined by 
means of a calibration curve where the logarithm of the initial template nucleic acid 
concentration is plotted on the x-axis and the Cq plotted on the y-axis. Assays are 
assessed for efficiency to determine if they are equal to 1.0 (100%) or within an 
acceptable range of +/-10% as stated by manufacturer specifications. Calibration 
curves showing a slope of 3.32 +/-10% reflects an efficiency of 100% +/-10%.

To accurately determine the amplification efficiency, a serial dilution series of 
cDNA from a HCT116 colon cancer cell population were prepared with reductions in 
starting template DNA from 1.25ng/µl to 0.00125ng/µl. A random sampling of 
assays from the panel was selected for efficiency experiments, namely CD9, HPRT1, 
and RACK1 (Figure 49). Efficiency is calculated from the slope of the standard 
curve using the following formula;

$$E = -1 + 10^{(-1/\text{slope})}$$

and converted to a percentage. Amplification efficiency ranges elucidated for a 
sampling of assays selected were between 93-105%. Another critical parameter that 
was assessed is the $R^2$ values generated for PCR efficiency. $R^2$ values indicate good 
PCR reproducibility in both experiments. When the $R^2$ is 1, the value of Y (Cq) can 
be used to accurately predict the value of X(concentration). An $R^2$ value = >0.99 
provides good confidence in correlating two values.
Figure 49—Linear regression curves for a number of expression assays. The slope produced by a PCR standard curve is used to calculate % efficiency. Efficiencies of 102.73% (HPRT1), 101.86 (CD9) and 99.34% (RACK1) were determined. The presence of inhibitors in the reaction could potentially lead to efficiencies over 100% as the difference between Cq values is smaller than predicted and therefore the amplification seems more efficient.
Parallel tests on the GEI instrumentation and commercial platform analysed the expression of fourth assay, *Beta-2-Microglobulin (B2M)* (Reference Sequence: NM_004048.2) (Figure 50). B2M is a reference gene which is commonly used in relative quantification studies either independently or in combination with other endogenous controls. Evaluation of data points and efficiencies for both instruments shows good concordance.

- **Figure 50** - Determination of efficiency of B2M assay using the Ct-slope method with 4 concentrations covering a 3-log range. The calculated efficiency for this assay is GEI 94.36% - ABI 93.76%.

### 5.3.2 Line-to-Line Variance Assessment

A four-line system allows for a variation of experiments to be carried out in parallel. The instrumentation can be up scaled or downscaled to modify the number of lines depending on individual experimental requirements. Further serial dilution tests assessing possible line to line variance of the developed instrumentation are shown. Firstly, a serial dilution experiment was repeated three times across one line (Figure 51-Graph A). The Cq values established show minimal variation across the line ensuring good reproducibility when only using individual lines. Secondly, the system was evaluated using a serial dilution across the four system lines, again showing minimal variation when carrying out parallel experiments and operating at maximum capacity (Figure 51-Graph B).
As an additional overall metric for concordance between the developed instrumentation and commercial platform, an experiment was prepared to perform repeat dilution series (Figure 52). The cumulative dataset for nine GEI Instrument tests (blue 1-9) can be compared to the three AB7900HT tests (orange 1-3) which used replicate cDNA, assays and mastermix. The results again show good concordance and minimal differences in Cq values. Reproducible amplification, no cross-contamination and detection of low concentrations were demonstrated on numerous consecutive sample drops.

**Figure 51**—HCT116 cDNA vs. β2M expression tests. (A) Graphical representation of data across one instrument test line for four dilutions repeated three times. (B) Repeated serial dilution for four lines operating on instrument showing minimal line to line variance when system is at maximum capacity. An incorrect optical detection setting was applied to Line 2 which accounts for slight variation.
5.3.3 Expression Knockdown Validation

Four repeat experiments were performed to investigate the instruments’ sensitivity and reproducibility capabilities by amplifying a gene target under normal PCR amplification conditions. The gene being analysed was Receptor for Activated C Kinase 1 (RACK1) along with one reference gene PPIA. RNA was extracted from a HCT116 colon cancer cell line and was separated into two volumes, one of which was transfected with small interfering RNA or siRNA oligonucleotide to suppress the expression of the RACK1 gene. This transient knockdown of the RNA is achieved by the binding of the oligonucleotide to the active gene transcript for RACK1 and causes decreased expression of the gene. cDNA was then synthesised. The standard cDNA and the transfected cDNA samples were prepared individually on a well plate. Gene assay and mastermix for experimentation were plated on a second plate as described in Chapter 3. The results from the transfection and inhibition of RACK1 expression are shown in Figure 53. Ct values of the reference gene PPIA are also shown. The expression of the reference gene does not change significantly between pre- and post-transfection but RACK1 expression is inhibited as expected. Pre-transfection
RACK1 exhibits a threshold cycle of 22 while post-transfection samples show the threshold cycle is 27 which, when calculated using $2^{(ΔΔCt)}$ formula, (Livak and Schmittgen., 2001) indicates a fourteen-fold knockdown of the RACK1 gene by the transfection process.

5.3.4 Multi-gene Expression Validation

Dual expression validation tests were performed to determine the specificity of the instrumentation when analysing the expression of gene assays with different properties and expected expression values. Experiments targeted the RACK1 gene and Phosphoglycerate Kinase 1 (PGK1) (Reference Sequence: NM_000291.3) reference gene. cDNA was synthesised from the same colon carcinoma cell line (HCT116) and Taqman® chemistries were used for detection. RACK1 was chosen for this experiment because it has been shown to be expressed intensely in colon cancer cells (Saito, 2002) and PGK1 was selected because of its stability showing no or only minimal variations in expression levels in similar experiments (Falkenberg et al., 2011). Results comparing the expression of these genes can be seen in Table 5, with concurrent tests on the AB7900HT using identical experimental preparation.
conditions also shown. The amplification plots for the GEI and ABI data can be seen in (Figure 54). The mean ($\mu$), standard deviation ($\sigma$) and coefficient of variation (CV) across the 3 wells are shown and also on the 4 instrument individual lines. The closeness of data points to the mean influence the standard deviation value and subsequently affects the coefficient of variance (CV) of the Ct.

**Figure 54** - Above: AB7900HT Real-Time PCR amplification plot showing the expression of RACK1 and expression of PPIA in HCT116 cDNA. Below: Corresponding data from the GEI instrumentation. The corresponding Cq values can be seen in Table 5.
Table 5- RACK1 and PGK1 expression (Ct values) in HCT116 colon cancer cell line cDNA. Averages of the replicates (AB7900HT three-well average and GEI triplicate average across the four instrument lines) were used when comparing instruments.

5.3.5 Contamination Testing Validation

An important issue that warranted consideration for performing biological reactions is the effect, if any, of carryover from each reaction to the next and how carryover components may affect the trailing reaction. In previous work on this type of PCR, cross-contamination between droplets was attributed to droplet instability and the formation of small satellite droplet (Curcio and Roeraade, 2003). In the instrumentation developed, droplets are transported in the micro-capillary tubing by the carrier fluid where the interfacial properties of the carrier fluid, coupled with the constant flow of fluid into the system separating the droplet trains, means carry-over contamination risks are reduced. The hydrophobicity of the PTFE tubing also reduces the risk of sample contamination. In addition, due to the biphasic flow in the capillary, a liquid film exists between the droplets and capillary wall. This oil film prevents the aqueous droplets from wetting the wall of the tube, preventing contamination of the flow conduit (Morris, 2008).

During testing, the risk of carryover contamination was measured by analysing the amplification signal obtained by substituting the DNA template in the reaction with water and processing it immediately after loading a positive sample containing DNA in an interspersed dipping sequence (Figure 55). Droplets were evaluated by alternating between NTC reactions \( n = 65 \) and template nanodroplet reactions \( n = 50 \) in single dipping mode. NTCs contain the primers and mastermix but do not
contain any template. There is no observable contamination between different droplets and furthermore amplification is consistent for each of the reaction droplets.

**Figure 55**-Contamination tests performed on instrument. The signal intensity heat map shows droplet number on the y-axis and cycle number on the x-axis with no NTC amplification detected. Also in the figure is a zoomed view of the reporter dye (FAM) fluorescence at cycle 36 showing clear distinctions between positive reactions and NTCs. The amplification plot with a Cq equal to 28.15 ± 0.7 at 0.2 threshold is also shown.

**5.4 Chapter Close**

This chapter outlines the validation of the developed qPCR instrumentation which will be used for more advanced gene expression studies in this thesis. The chapter interrogates parallel tests which are performed on both the GEI instrument and the AB7900HT platform. Efficiency tests on randomly selected assays are performed analysing the percentage efficiency of each reaction, CV values and R² values. Line to line variance is also assessed along with expression knockdown and multi-gene expression tests. Finally, contamination risks are examined and tests performed to confirm that the risk of droplet to droplet contamination between samples is controlled during experimentation. Results that have been generated validate the instrumentation for more detailed expression profiling studies which will be outlined in **Chapter 6**.
Chapter 6

EXPERIMENTAL RESULTS
6 Experimental Results

6.1 Abstract
The following chapter presents and discusses results from analytical work performed using the microfluidic qPCR platform. A brief outline of specific experimental design is discussed in relation to preparation of the instrument. The primary experiments focus on analysing the RNA expression of a panel of thirty-four ECM genes in a cohort of twenty-four pairs of colorectal cancer samples and matched associated normal tissue. Differential expression patterns emerged between the normal and malignant tissue and correlated with histopathological parameters and overall surgical staging. The differential expression study between the normal and cancerous matched patient samples is followed by analysis of expression patterns between different patient groupings. The findings demonstrate that a droplet-based microfluidic quantitative PCR system enables biomarker classification. Results of a verification study performed on a single, randomly-selected colorectal sample from the cohort are then shown. The sample is analysed on a commercial platform to determine if the expression of five of the most dysregulated genes corresponds to the expression profile identified on the microfluidic platform.

In the second part of this chapter, the ECM genetic panel is assayed in a malignant breast specimen which has been sectioned into seven core samples. A matched normal sample is attained from the same patient. Tumour heterogeneity is highlighted where differential expression for the corresponding panel of thirty-four ECM genes is observed in each individual sample. Again, a verification study of the most dysregulated genes is performed to verify genetic dysregulation and instrument reliability. By analysing a small panel of ECM genes in a tumour sample using microfluidic methods to determine changes in expression patterns it is anticipated to achieve a more informed prognosis for the patient to support existing cancer categorisation strategies.

6.2 Colorectal Profiling ECM study

6.2.1 Differential gene expression in matched samples
RNA expression levels of thirty-four ECM genes (Table 2) were quantitatively analysed by qPCR. Twenty-four malignant patient samples were included in this
Experimental Results

pilot study in addition to matched corresponding normal tissue. Eighteen of the malignant samples categorised into four UICC stages (stage I-IV) alongside six pre-cancerous samples, some with inflammatory conditions. Gene expression differences were observed in the ECM genes assayed in the tumour tissue in comparison to the individual matched normal samples. Tabulated results can be seen in Table 6.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Patient cohort (n=24)</th>
<th>Fold change</th>
<th>Stat sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEACAM1</td>
<td>-4.1</td>
<td>0.048*</td>
<td></td>
</tr>
<tr>
<td>COL1A1</td>
<td>+6.2</td>
<td>0.039*</td>
<td></td>
</tr>
<tr>
<td>CXCR2</td>
<td>+4.0</td>
<td>0.02*</td>
<td></td>
</tr>
<tr>
<td>EGF</td>
<td>-5.0</td>
<td>0.014*</td>
<td></td>
</tr>
<tr>
<td>EGFR</td>
<td>-3.0</td>
<td>0.05*</td>
<td></td>
</tr>
<tr>
<td>ITGA2</td>
<td>+4.0</td>
<td>0.007*</td>
<td></td>
</tr>
<tr>
<td>ITGA8</td>
<td>-3.6</td>
<td>0.003*</td>
<td></td>
</tr>
<tr>
<td>ITGBL1</td>
<td>+11.8</td>
<td>0.003*</td>
<td></td>
</tr>
<tr>
<td>IGF1</td>
<td>-7.7</td>
<td>0.007*</td>
<td></td>
</tr>
<tr>
<td>IL-8</td>
<td>+18.45</td>
<td>&lt;0.001*</td>
<td></td>
</tr>
<tr>
<td>MMP7</td>
<td>+31.1</td>
<td>&lt;0.001*</td>
<td></td>
</tr>
<tr>
<td>MMP9</td>
<td>+7.4</td>
<td>0.02*</td>
<td></td>
</tr>
<tr>
<td>RACK1</td>
<td>+2.4</td>
<td>0.019*</td>
<td></td>
</tr>
<tr>
<td>SPARC</td>
<td>+3.8</td>
<td>0.033*</td>
<td></td>
</tr>
</tbody>
</table>

Table 6 - The genetic panel was analysed as a combined cohort in the samples with no bias regarding staging or other factors. Shown are expression fold change between normal and cancerous samples and associated p-values (*statistically significant)

Differential gene expression greater than two-fold difference between normal and malignant tissue was observed in twenty-two genes. The differences identified in expression were statistically significant in the case of fourteen genes. These genes have varying biological functions and code for cell surface proteins, growth factors, proteins involved in tissue remodelling and cell adhesion. The greatest difference in gene expression was observed in MMP7 (+31 fold, p=<0.001) between the matched tumour and normal mucosa in this study. IL-8 expression was up-regulated (+18-fold, p=<0.001) and ITGBL1 expression was shown to be increased (+11-fold, p=0.003) in malignant versus matched normal tissue.
Figure 56 - Gene expression fold change levels for the patients included in the study with statistically significant differences between normal mucosa and matched tumour samples. Average fold change relative to normal tissue sample normalised to the geometric mean of the reference genes.
MMP9 expression was observed to be upregulated (+7-fold, p=0.02) in the malignant tissue as was COL1A1 expression (+6-fold, p=0.039) and ITGA2 (4-fold, p=0.007). CXCR2 (4-fold, p=0.022), SPARC (3.5-fold, p=0.033) and RACK1 (2.5 fold, p=0.019) were also identified as being upregulated in the matched comparison. Conversely, decreased expression levels were observed in six of the genes in the ECM panel. IGF-1 expression levels were lower (-8-fold, p=0.007) in cancer tissue than in the normal tissue. Decreases in expression levels for EGF (-5 fold, p=0.014) and EGFR (-3-fold, p=0.05) were also observed. CEACAM1, was also identified as being decreased (-4-fold, p=0.048). ITGA8 gene expression levels were shown to be down regulated (-3.6-fold, p=0.003) in the cohort. A graphical representation of this can be seen in Figure 56.

6.2.2 Dysregulation Discussion

The differential expression of fourteen genes which were statistically significant was identified. Identifying differentially expressed genes acts as an additional clinical tool that may be used to diagnose patients and may influence the potential treatment decisions. The global dysregulation observed in this gene expression study of the ECM indicates that substantial dysregulation and remodelling of the ECM occurs during colorectal pathogenesis and this can lead to increased or decreased metastatic potential of the tumour.

Upregulation.

Matrix metalloproteinases: The expression of MMP7 was observed to be the elevated in all of the samples profiled. MMP7 cleaves cell surface proteins, promotes adhesion of cancer cells, and increases the potential of tumour metastasis (Edman et al., 2011) by degrading key components of the basement tissue structures. When profiled, the expression of the MMP7 gene in the cancer samples was increased an average of 31 fold (p<=0.001) in comparison to the matched normal sample. In particular expression of MMP7 appears to be increased in samples diagnosed with later stage colorectal cancer. This upregulation is very significant in understanding how cancer cells, which already have increased motility and metastatic functionality, can break away from the primary tumour and move through the basement membrane with a higher degree of ease in comparison to other cell types. MMP9 is a second protease from the initial panel that emerged as having differential expression in
malignant tissue in comparison to the normal tissue. Its overexpression has been observed in colon cancer studies where it is again largely associated with metastatic progression of colorectal cancer and overall poor prognosis (Yang et al., 2014, Li et al., 2013).

**IL-8:** Coupled with the upregulation of *MMP7* and *MMP9*, the next most dysregulated gene which emerged from the ECM panel was *IL-8*. A paired sample student t-test revealed a significant overall differences in *IL8* group mean expression levels between tumours and normal colorectal tissue where the overall expression of *IL8* was significantly up-regulated (18-fold, *p*=<0.001). *IL-8* overexpression has been detected in many carcinomas, including colorectal cancer and is associated with poor prognosis (Kheirelseid et al., 2010b, Ning et al., 2010). It has been suggested that tumour cells produce IL-8 as an autocrine growth factor that promote tumour growth, tissue invasion and metastatic spread (Xie, 2001). The gene also functions to activate multiple intracellular signalling pathways and is also known to promote angiogenic responses in endothelial cells, increase proliferation and survival of endothelial and cancer cells, and potentiate the migration of cancer cells (Waugh and Wilson, 2008). Therefore, it was expected that the expression of *IL8* would be higher in the cancer samples in comparison to the normal.

**Integrins:** Two members of the integrin family were also observed to be statistically upregulated in the samples analysed. *ITGBL1* and *ITGA2* have varied roles in regular cellular behaviour and tissue dynamics. *ITGBL1* expression was shown to be increased in malignant versus normal tissue (+11-fold, *p*=0.003). Despite not being highlighted prominently in literature, *ITGBL1* may expedite any migratory functionality gained by cells as it has been shown to have a role in the breast-to-bone cancer metastases pathway by activating TGFβ (Li et al., 2015a). The most significant increase is seen in the later staged cancers in comparison to earlier staged and benign tissue samples. *ITGA2* expression in the cohort is shown to be upregulated four times more than that of the normal tissues (*p*=0.007). This increase was most evident in earlier stages of cancer development which may indicate *ITGA2* having a role in early epithelial-mesenchymal transition (EMT) processes. The overexpression may also associate *ITGA2* with a pre-cancerous functionality which may allow cells to gain early stage migratory and invasive characteristics.
**SPARC, CXCR2, RACK1**: Increased expression of *SPARC* was shown in the profiling which conforms with literature (Said N, 2013) while *CXCR2*, which is a receptor for *LAMA1* and also a critical mediator of cellular senescence was observed to be overexpressed. Levels of *RACK1* overexpression were also evident between samples and it has previously been shown to interact with *PTK2* and *IGF1R* to promote cellular migration (Kiely et al., 2005).

**Downregulation.**

Decreased expression levels in some prominent genes in the test panel were observed between the normal and cancerous samples. In total, six genes were downregulated when the means of each sample were assessed by paired sample t-test.

**IGF1**: *IGF1* was one of the genes in the panel which was dysregulated when assayed. *IGF1* is a growth hormone regulated that exerts powerful effects on the promotion of cellular behaviour and tissue development. In cancer, levels of *IGF1* gene expression has been reported to be higher in colorectal cancer samples than that of normal mucosa (Freier et al., 1999, Rong Zhang, 2013). It has been shown to become dysregulated due to binding activity to its receptor *IGF1-R* and through the activation of signalling pathways by key regulatory proteins such as Protein Kinase B (Nicholson and Anderson, 2002) and RACK1 (Kiely et al., 2006). *IGF1* secreted from intestinal fibroblasts has two important effects on intestinal epithelial cells: it stimulates their proliferation and promotes migration (Simmons et al., 1999). However, when profiled, the expression of the *IGF1* gene in the cancer samples was decreased an average of 8-fold (*p* =< 0.007) in comparison to the matched normal sample. In particular expression of IGF1 appears to be decreased in 7/11 stage II samples and downregulated in 6/7 stage III samples. This downregulation could play a role in changes to cellular motility levels and aid suppressing the metastatic potential of the tumour.

**EGF and EGFR**: *EGF* is a growth factor which becomes activated when it binds with its growth factor *EGFR*. Both ligands have both been shown to be involved in several stages of oncogenesis such as invasion and growth and are particularly relevant in the targeting of colorectal cancer (Shaheen et al., 2001, Mendelsohn, 2001) by blocking *EGFR* when it is overexpressed to prevent *EGF* from binding to it. Monoclonal antibodies or tyrosine kinase inhibitors which block specific proteins
Experimental Results

and signal transduction pathways are most effective for this. In this study, EGF was expressed 5 fold less (p=0.014) in the cancerous tissue in comparison to the corresponding normal sample while its receptor EGFR was also downregulated 3-fold (p=0.05). For both genes, the expression was most decreased in stage II samples. The expression in the stage III and benign samples was most similar to the normal sample indicating that the patients in the cohort would most likely not be suitable for anti-EGFR therapies.

**CEACAM1**: CEACAMs are cell surface glycoproteins primarily involved in intracellular binding. The expression of CEACAM1 was observed to be the most under expressed in all of the samples profiled. The role of CEACAM1, in some cases, is contradictory in whether the gene has a tumour suppressor or tumour promoter role in cancer (Fiori et al., 2012). Primarily, it is an adhesion molecule that, in most malignant transformations, is thought to be a tumour suppressing force where levels decrease in early stages of the oncogenic process and allow cellular transformations to become less anomalous. In some tumour cell lines, restoration of CEACAM1 expression reduces the level of modulation and dysregulation and, in some cases, abolishes their oncogenicity in vivo (Kammerer et al., 2004). In this study, levels of CEACAM1 were observed to be, on average, 4-fold (p=0.048) downregulated across the twenty-four tumour samples in comparison to the matched normal tissue.

**ITGA8**: The ITGA8 gene encodes the alpha 8 subunit of the heterodimeric integrin alpha8-beta1 protein which is a transmembrane receptor protein. Several integrin family members are well documented to be involved in carcinogenic processes, especially colorectal and breast, but the direct role of ITGA8 is largely unknown. Similar to other integrins, the primary role of ITGA8 is to mediate cellular processes such a cellular adhesion, communication with the ECM and activation of signals for downstream growth factors. Further to this, ITGA8 has been shown to be one of a panel of genes on an enriched molecular pathway in colorectal cancer (Kok-Sin et al., 2015). When assayed in the twenty-four colorectal cancer tissues, the expression of ITGA8 was principally downregulated an average of 3.6-fold (p=0.003) across the cohort. The downregulation was more definitive in stage III samples and some advanced stage II samples. Overall, if the stage III cohort is taken independently, expression of ITGA8 is downregulated 10-fold in comparison to the corresponding
normal sample. This may indicate that the adhesive properties for which integrins including \textit{ITG}A8 are responsible for, becoming ineffective and possibly leading to a less constrained cancer cell which could work to increase the metastasis rate of the tumour. After the initial gene expression levels were determined by experimental means, gene pairs which had similar expression patterns from the fourteen gene panel were sought. The associations between the significant gene expression arrays are shown in Table 7.

6.2.3 Expression correlation and gene networking

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|}
\hline
\textbf{Correlations} & \textbf{ITGA2} & \textbf{ITGA8} & \textbf{ITGBL1} & \textbf{IGF1} & \textbf{SPARC} & \textbf{EGF} & \textbf{EGFR} & \textbf{MMP7} & \textbf{MMP9} & \textbf{COL1A1} & \textbf{CEACA1} & \textbf{M1} & \textbf{RACK1} & \textbf{IL8} & \textbf{CXCR2} \\
\hline
\textbf{ITGA2} & & & & & & & & & & & & & & & \\
\textbf{ITGA8} & PC & 1 & \textbf{0.082} & 1 & & & & & & & & & & & \\
\textbf{ITGBL1} & PC & 0.157 & 0.339 & 1 & & & & & & & & & & & \\
\textbf{IGF1} & PC & 0.012 & 0.371 & 0.285 & 1 & & & & & & & & & & \\
\textbf{SPARC} & PC & 0.211 & 0.264 & \textbf{0.493} & 0.235 & 1 & & & & & & & & & \textbf{0.470} \\
\textbf{EGF} & PC & 0.086 & 0.077 & \textbf{0.487} & 0.072 & 0.047 & 1 & & & & & & & & \\
\textbf{EGFR} & PC & 0.227 & 0.161 & 0.172 & 0.32 & \textbf{-0.277} & \textbf{0.470} & 1 & & & & & & & \\
\textbf{MMP7} & PC & 0.522 & 0.081 & 0.003 & -0.102 & 0.197 & 0.051 & 0.049 & 1 & & & & & & \\
\textbf{MMP9} & PC & 0.146 & 0.053 & 0.053 & 0.041 & 0.298 & 0.017 & 0.169 & 0.051 & 1 & & & & & \\
\textbf{COL1A1} & PC & 0.079 & -0.323 & 0.174 & -0.292 & \textbf{0.763} & -0.13 & 0.283 & 0.131 & 0.277 & 1 & & & & \\
\textbf{CEACA1} & PC & 0.322 & 0.081 & 0.003 & -0.102 & 0.197 & 0.051 & 0.049 & 1 & & & & & & \\
\textbf{RACK1} & PC & 0.274 & 0.125 & 0.09 & -1.08 & 0.357 & 0.002 & 0.077 & 0.066 & 0.354 & \textbf{0.451} & 0.258 & 1 & & \\
\textbf{IL8} & PC & 0.141 & -0.313 & 0.21 & -0.311 & -0.099 & -0.35 & 0.073 & 0.035 & 0.058 & -0.073 & 0.145 & 0.094 & 1 & \\
\textbf{CXCR2} & PC & 0.392 & -0.1 & 0.403 & -0.124 & 0.141 & 0.054 & 0.369 & 0.105 & 0.017 & 0.289 & -0.11 & 0.25 & 0.118 & 1 \\
\hline
\end{tabular}
\caption{Correlations between gene expression levels. Correlation is significant at the 0.05 level (2-tailed). * Correlation is significant at the 0.01 level (2-tailed). ** Correlation is significant at the 0.01 level (2-tailed).}
\end{table}

Table 7 - Genes with dysregulated expression from Figure 1 are shown. Pearsons correlated (PC)/co-expressed genes are highlighted in bold.
A relationship between the genes can either be linear or monotonic and tests were performed to allow for the construction of co-expression networks which identify genes which show either a similar expression pattern across genes or genes which rise and fall inversely across samples. These networks are of biological interest as co-expressed genes are either controlled by the same transcriptional regulatory program, functionally related or members of the same pathway or protein complex (Weirauch, 2011). The strength of a relationship is determined by closeness to 1 (direct correlation) or -1 (inverse correlation). Once the most suitable relationship is identified, a co-expression network can be constructed by looking at which genes have a similar or converse expression pattern across a range of samples. Connectivity and gene co-expression relationships can be defined from these networks and have been associated with key disease-related pathways based on the detection of genes with critical functional roles. A basic gene co-expression network from which genes with multiple interactive roles can be identified is shown in Figure 57.

![Gene Co-Expression Network](image)

**Figure 57**- Gene Co-Expression Network **. Correlation is significant at the 0.01 level (2-tailed). Significance threshold $|r(Gi, Gj)| > 0.6$

The strongest positive relationships was observed between \textit{COL1A1} and \textit{SPARC} expression ($\rho = 0.763$, p-value $< 0.001$) and \textit{COL1A1} and \textit{RACK1} ($\rho = 0.451$, p-value $= 0.032$). \textit{COL1A1} and \textit{SPARC} genes participate in ECM and cytoskeletal control and are also involved in the carcinogenesis of many malignancies and both have been shown to be relevant for tumour prognosis (Helleman et al., 2008). The association between \textit{COLLAGEN} and \textit{SPARC} has been highlighted (Wei et al., 2012).
Experimental Results

where the silencing of the *SPARC* gene by transfection of cells with *SPARC* siRNA led to the expression of *COL1A1-COL3A1* mRNAs and proteins to be significantly enhanced. *MMP7* and *ITGA2* also showed a direct correlation (\(\rho = 0.552, p\text{-value} = 0.018\)) indicating a potential link between the two genes. Correlations between *ITGBL1* and both *SPARC* expression (\(\rho = -0.720, p\text{-value} = 0.001\)) and *EGF* (\(\rho = -0.699, p\text{-value} = <0.001\)) also emerged as being statistically significant. A strong positive correlation was observed between *EGF* and its receptor *EGFR* expression (\(\rho = -0.470, p\text{-value}=0.032\)). Using the approach of assessing the co-expression of genes can be very influential in determining relationships between genes which perform a wide variety of functions to modulate disease development and metastasis.

Subsequently, the patient cohort was stratified according to UICC staging to identify further differential expression patterns. The cohort was categorised based on the histopathological diagnoses as follows; Group 1: Benign samples which included diverticulosis, dysplasia, tubulovillus adenoma (TVA) polyps and an early Stage I sample \((n=6)\), Group 2; Stage II carcinomas \((n=11)\), Group 3; Stage III and Stage IV

Figure 58- The figure shows expression fold-difference for the gene panel between normal and cancer samples when samples are categorised according to UICC staging. Normal samples are normalised to one and highlighted in red, Group 1 is green, Group 2 is purple and Group 3 is blue. Error bars show standard error for each grouping. Statistical significance is also shown (p-value; *=<0.05; **=<0.01).

Subsequently, the patient cohort was stratified according to UICC staging to identify further differential expression patterns. The cohort was categorised based on the histopathological diagnoses as follows; Group 1: Benign samples which included diverticulosis, dysplasia, tubulovillus adenoma (TVA) polyps and an early Stage I sample \((n=6)\), Group 2; Stage II carcinomas \((n=11)\), Group 3; Stage III and Stage IV
advanced carcinomas \((n=7)\). Based on this categorisation, a number differentially expression patterns could be observed for the gene panel \((\text{Figure 58})\). For Group 1, ECM gene dysregulation between the normal and cancer samples for patients was noted only in an upregulation of \(IL-8\) (15-fold, \(p=0.05\)). In Group 2, statistically significant up-regulation was observed for \(MMP7\) (48 fold, \(p=0.004\)), \(IL-8\) (17-fold, \(p=0.015\)), \(COL1-A1\) (8.7 fold, \(p=0.049\)), \(LAMA1\) (11-fold, \(p=0.033\)), \(ITGBL1\) (8-fold, \(p=0.016\)), \(MMP9\) (5.2 fold, \(p=0.05\)), \(ITGA2\) (5.1, \(p=0.039\)), and \(COL3A1\) (4-fold, \(p=0.041\)). Genes which were observed to be statistically down regulated for Group 2 samples included \(EGF\) (-7.4 fold, \(p=0.004\)) , \(IGF1\) (-4.7 fold, \(p=0.034\)), and \(CEACAM1\) (-4.5-fold, \(p=0.044\)). Finally, in Group 3, up regulation of \(IL-8\) (25 fold, \(p=0.002\)), \(ITGBL1\) (22.5-fold, \(p=0.021\)), \(MMP7\) (17.8-fold, \(p=0.035\)), \(SPARC\) (5.6 fold, \(p=0.05\)), \(ITGA2\) (5.1 fold, \(p=0.029\)), \(RACK1\) (5 fold, \(p=0.006\)) and \(ITGAV\) (2 fold, \(p=0.012\)). Down regulation of \(IGF1\) (-11.7 fold, \(p=0.04\)), \(LAMA1\) (-9.4 fold, \(p=0.08\)), \(ITGA8\) (-7.8 fold, \(p=0.048\)) and \(EGF\) (-7 fold, \(p=0.03\)). A simplified diagram showing this dysregulation between groups can be seen in \(\text{Figure 59}\).

\begin{center}
\textbf{Figure 59}-Potential genetic biomarkers for CRC diagnosis based on UICC staging. Patients were grouped into categories depending on the staging of the carcinoma.
\end{center}
### 6.2.4 Histopathological parameters associations

Analysis was performed to identify possible associations between malignant sample gene expression and histopathological parameters. Parameters recorded during surgical resection can be observed in Table 8. A graphical representation of which parameters are most associated with specific genes can be seen in Figure 60. Individual associations can be seen in Appendix A.

**Tumour diameter:** The gene expression of three ECM genes in the panel, *IGF1* \((r=-0.596, p=0.009)\), *ITGBL1* \((r=0.669, p=0.005)\) and *VEGF* \((r=-0.487, p=0.04)\) correlated with the diameter of the samples which were recorded during surgical examination. The tumour size at the largest point was recorded in millimetres (mm) with three categories of <30mm, 30-40 mm, and >40mm. *IGF1* decreased expression (-13 fold), and *ITGBL1* increased expression (26-fold) was observed to be present in tumours with a diameter greater than 40mm.

**Stage:** The clinical staging of the patients’ tumour was determined using the American Joint Committee on Cancer (AJCC) Tumour-Node-Metastasis (TNM) system. *ITGBL1* expression \((r=0.588, p=0.013)\) was identified as showing a stepwise increase in expression which correlated with the staging of the patient cohort with an ~8-fold increase in expression for patients with Stage II and ~22-fold increase for Stage III in comparison to Stage I. *LAMA1* expression showed an increase in Stage II (+20-fold) but a significant decrease in expression in Stage III (-40-fold) \((r=0.509, p=0.037)\).

**Differentiation:** The expression levels of three genes showed a moderate correlation with the level of differentiation which was recorded during surgical resection. Upregulation of *COL3A1* (7-fold) was significantly associated with poorer differentiation \((r=0.469, p=0.024)\) along with upregulation of *MMP9* (6.1-fold) \((r=0.423, p=0.05)\). It has been shown (Zhang et al., 2012) that increased levels of *MMP9* correlate with poorer differentiation. A slight increase in *IGF1R* (2.4-fold) \((r=0.533, p=0.049)\) levels also correlated with poorer differentiation.

**Recurrence:** Three of the samples in the cohort were taken from patients who later presented with a recurrence of the carcinoma. *PTK2* \((r=0.435, p=0.043)\) levels in these patients were shown to be elevated in comparison with patients whose cancers had not recurred.
### Experimental Results

<table>
<thead>
<tr>
<th>Number of Patients</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>TVA/other pathology</td>
<td>6 (25%)</td>
</tr>
<tr>
<td>Carcinoma</td>
<td>18 (75%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
</tr>
<tr>
<td>Mean ±SD</td>
</tr>
<tr>
<td>Median</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>UICC Staging (n=24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N/A</td>
</tr>
<tr>
<td>Stage I</td>
</tr>
<tr>
<td>Stage II</td>
</tr>
<tr>
<td>Stage III</td>
</tr>
<tr>
<td>Stage IV</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Location (n=24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right colon (male/female)</td>
</tr>
<tr>
<td>Left colon (male/female)</td>
</tr>
<tr>
<td>Rectum/Sigmoid (male/female)</td>
</tr>
<tr>
<td>Total Colectomy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tumour diameter (n=18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;30mm</td>
</tr>
<tr>
<td>30–40mm</td>
</tr>
<tr>
<td>&gt;40mm</td>
</tr>
<tr>
<td>Unknown</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tumour Grade (n=18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
</tr>
<tr>
<td>T2</td>
</tr>
<tr>
<td>T3</td>
</tr>
<tr>
<td>T4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nodal Status (n=24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NX</td>
</tr>
<tr>
<td>N0</td>
</tr>
<tr>
<td>N1</td>
</tr>
<tr>
<td>N2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Differentiation (n=24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well-moderate</td>
</tr>
<tr>
<td>Moderate</td>
</tr>
<tr>
<td>Moderate-poor</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Invasive (n=24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mucin Secretion (n=24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mucinous</td>
</tr>
<tr>
<td>Not stated</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lymphovascular Invasion (n=18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
</tr>
<tr>
<td>Absent</td>
</tr>
<tr>
<td>Suspicious</td>
</tr>
<tr>
<td>Not stated</td>
</tr>
</tbody>
</table>

Table 8 - Clinicopathological data of patients selected for the analysis study.
**Experimental Results**

**Age:** Patients were categorised into two groups of being up to and including sixty-four years of age and sixty-five and older. Increased MMP7 expression correlated with increased age in the cohort \((r=0.488, p=0.018)\). Patients aged sixty-four and under had an average increase of 10-fold expression of MMP7 while patients aged sixty-five and over had an average of 40-fold increase. CDH2 levels were also correlated \((r=0.490, p=0.015)\) with age where patients under 64 had six fold decreased CDH2 expression than people aged over 65. Increased IGFBP3 expression \((+2\text{-fold})\) correlated \((r=0.473, p=0.023)\) with an increased age compared to the younger group.

**Invasiveness:** The invasiveness of the tumour is closely linked to the size and the grade of the tumour and is determined by whether or not the tumour penetrates the basement membrane and/or muscle layers in the lumen of the colon. In the twenty-four patient cohort, nineteen samples are documented to be invasive malignancies of varying degree. The depth of invasion was not recorded. SPARC and COL1A1 expression levels were shown to be increased significantly \((4.2\text{-fold}) (8.1\text{-fold})\) in samples which were invasive \((r=0.465, p=0.022) \,(r=0.494, p=0.017)\).

**Gender:** IGFBP3 expression levels were noted to correlate \((r=0.424, p=0.044)\) with patient gender with male patients having an increased expression \((+4\text{-fold})\) compared to females.

**Lymph Node Status:** In non-metastatic colorectal cancer, lymph node status is the strongest pathologic predictor of patient outcome and is used for determining the most appropriate adjuvant treatment for patients. In the cohort, patient's lymph nodes were examined for abnormalities and inflammation. The nodal status was determined by the TNM system mentioned previously where the "N" portion of this classification system refers to the nodal status indicating the presence, if any, of cancer spread to lymph nodes. The system is: NX: Regional lymph nodes cannot be assessed, N0: Cancer has not spread to regional lymph nodes, N1: Cancer has spread to 1 to 3 lymph nodes, and N2: Cancer has spread to 4 to 9 lymph nodes. Four patients were classified as NX, twelve patients were classified as N0, six were classified as N1, and 2 were classified as N2. The expression of PTK2, or Focal Adhesion Kinase (FAK) as it is more commonly known, showed a step-wise increase in expression from N0 \((+1.5\text{ fold})\) to N1 \((+4.2\text{-fold})\) to N2 \((+7.4\text{ fold})\)\((r=0.587, p=0.006)\). A stepwise decrease was also observed for ITGA8 for N0 to N2 \((-2.1\text{-fold})\).
Experimental Results

fold)(- 4-fold)(-15-fold)(r = -0.523, p = 0.018). EGF expression was upregulated in N1 and downregulated in N0 and N2.

**Perineural Invasion (PNI):** PNI is grossly underreported in CRC and could serve as an independent prognostic factor of outcomes in these patients (Liebig et al., 2009). PNI absence was reported in six of the 20 tumour samples. PNI was not stated for fourteen others. MMP9 levels were significantly decreased in patients with no PNI (r=0.481, p=0.037). Increases in MMP9 aid a cancer cell’s progression to metastasis and have been shown to correlate with increases in PNI in literature. (Duan et al., 2013)

**Mucinous Component:** From the cohort, four of the nineteen tumours were documented as having a mucinous component and all were Stage III carcinomas. Colorectal mucinous carcinomas present at a more advanced stage, predominantly in men, with a higher right colon location rate and a worse overall 5-year survival rate than the non-mucinous colorectal cancer (Papadopoulos et al., 2004). Three of the four patients with mucinous component were male. ITGBL1 and ITGA8 expression correlated with whether or not a mucinous component was present. Increased expression in ITGBL1 by 30 fold (r=0.669, p=0.005) was observed in mucin positive tumours and average expression of ITGA8 was decreased 10 fold (r=0.655, p=0.002). Three of the four patients received chemotherapy.

**Grade:** ITGA2, ITGB1, and CD9 gene expression levels were correlated with the grade of the malignant samples analysed albeit with a small sample size. ITGA2 and ITGB1 were observed be lower in T3 and T4 tumour samples in comparison to T1 and T2 tumour samples (r = -0.657, p = 0.004)(r = -0.501, p = 0.041). CD9 expression levels also correlated with the grade of the cancer samples with a stepwise decrease in expression (r = -0.613, p = 0.007) from T1 (+10 fold), T2 (+2.14 fold), T3 (+0.05 fold) to T4 (-11.5 fold). In addition to this analysis, the T1 sample and the two T4 samples were removed from the dataset and T2 and T3 correlations were evaluated independently. When this was done, the correlation between ITGA2 and ITGB1 expression with the grade of the tumour remained statistically significant. However, the correlation between CD9 expression and grade did not remain statistically significant.
6.2.5 Verification study using qPCR Instrument

Following determination of gene expression levels of the samples using the microfluidic gene expression instrumentation, conformation of the results was sought by performing a parallel real-time PCR experiment on the commercial instrument, the AB7900HT platform. Post verification, the process of identifying differentially expressed RNA levels using microfluidic methods would be supported and could potentially greatly expedite the process as well as the added benefit of using significantly reduced volumes of input materials. Additionally, increased throughput could identify potential biological markers quicker from extensive studies and future instrument revisions could act as an additional clinical tool to diagnose patients and influence the treatment decision. For the verification study, a parallel test on both instruments is designed which has the potential to identify discrepancies that could exist between instruments and investigate whether or not specific gene expression fold changes were unique to the microfluidic platform. The three genes which were identified as most upregulated and two genes most downregulated from the microfluidic instrumentation experimentation were selected for verification on the commercial platform along with three stable housekeepers.
### Table 9 – Sample and Reagent Volumes for ABI PCR verification studies

<table>
<thead>
<tr>
<th>Reagents</th>
<th>Volume (µl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template cDNA</td>
<td>0.5</td>
</tr>
<tr>
<td>Molecular Grade Water</td>
<td>1.75</td>
</tr>
<tr>
<td>Gene Assay</td>
<td>0.25</td>
</tr>
<tr>
<td>MasterMix</td>
<td>2.5</td>
</tr>
</tbody>
</table>

The largest gene expression fold change increase observed using the microfluidic instrument was in the expression of MMP7, IL8 and ITGBL1 and the largest decrease was observed in IGF1 and CEACAM1 and therefore these genes were chosen. Reference genes PPIA, PGK1 and HRPT1 were included for normalisation. Table 9 shows volumes of sample and reagent used for ABI instrument analysis. Reactions containing gene expression MasterMix and gene assay were prepared with 1:10 dilution cDNA which was synthesised from 1µg/µl of RNA. This is the same concentration which was used on the microfluidic instrument, so therefore Cq differences between GOIs and HKs for both platforms are expected to be zero (ΔCq=0).

**Figure 61** - Plate Layout for ABI Instrument. The samples, assay and mastermix are pipetted into individual wells. Green icon wells indicate normal samples while red icon wells indicate cancerous samples with blue indication NTC samples.
Experimental Results

Reactions were performed in triplicate for each gene of interest. NTCs were also prepared and run on the same plate to account for possible contamination. The plate layout for the ABI instrument can be seen in Figure 61. The PCR was then performed on the samples and the gene expression data was analysed for statistical significance by randomisation tests (Pair Wise Fixed Reallocation Randomisation Test©/REST Software). Statistical significance was set at p=<0.05 or less than 5% error probability. The trend of expression when analysed was replicated the commercial platform and the fold change intensity was also comparable (Table 10).

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Reference Genes</th>
<th>Genes of Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PPIA</td>
<td>PGK1</td>
</tr>
<tr>
<td>GEI</td>
<td>0.84</td>
<td>1.42</td>
</tr>
<tr>
<td>ABI</td>
<td>0.59</td>
<td>1.15</td>
</tr>
</tbody>
</table>

Table 10- Results from parallel verification experiments on the GEI and ABI platforms.

On the GEI instrument, MMP7 expression was again the most upregulated gene in the panel where its expression was increased significantly in comparison to the matched normal sample. On the ABI platform its expression was upregulated over 46-fold in comparison to the GEI which showed upregulation of 57-fold. Furthermore, IL8 and ITGBL1 were upregulated 33-fold and 3.5-fold respectively in comparison to the GEI which showed a 43 and 5-fold increase. Conversely IGF1 showed downregulation of 18-fold in comparison to 15-fold on the GEI as did CEACAM1 which showed 6-fold downregulation in comparison to the GEI which showed 2.7-fold. Despite slight variations in fold change amounts, these results signify substantial differential expression patterns in these genes between normal and malignant samples. Disparities in fold change intensity could be caused in instrument sensitivity, smoothening algorithms in the commercial platform or slight variations in reference gene expression. NTC wells showed no amplification. In conclusion, the verification of results, initially determined using microfluidic methods, on the commercial platform indicate the significant potential for this technology to be used for larger genetic and patient studies. A graphical representation of the fold changes for the 5 selected genes is shown in Figure 62.
The concept of tumour heterogeneity was first introduced in 1984 and up until this point cancers were mainly described as monoclonal (a group of cells produced from a single ancestral cell by repeated cellular replication) (Heppner, 1984). In this pilot study, the aim was to identify specific patterns of gene expression from the panel of ECM genes within core biopsies of a given breast tumour sample and compare the expression with the expression of a matched normal sample. The breast cancer (BrCa) tumour was pathologically examined and identified as estrogen receptor (ER) negative, progesterone receptor (PR) negative and not overexpressing the HER2/neu gene, otherwise known as triple negative breast cancer (TNBC). To examine distinct morphologically different entities within the tumour, ECM genes which are key regulators of cellular transformations and tissue developmental behaviours were identified from literature for gene expression analysis. The identification of differential mRNA expressions among the genes in each of the cores aims to highlight the heterogeneity within the single tumour and potentially indicate the need.

Figure 62-The above graph depicts the fold change for 5 genes (3 upregulated and 2 downregulated) as observed in a parallel experiment performed on the microfluidic (GEI) and commercial well-based (ABI) platforms. Results depicted show replicable and concurrent results. Error bars show standard error for each grouping.

6.3 Breast Profiling ECM study

6.3.1 Abstract

The concept of tumour heterogeneity was first introduced in 1984 and up until this point cancers were mainly described as monoclonal (a group of cells produced from a single ancestral cell by repeated cellular replication) (Heppner, 1984). In this pilot study, the aim was to identify specific patterns of gene expression from the panel of ECM genes within core biopsies of a given breast tumour sample and compare the expression with the expression of a matched normal sample. The breast cancer (BrCa) tumour was pathologically examined and identified as estrogen receptor (ER) negative, progesterone receptor (PR) negative and not overexpressing the HER2/neu gene, otherwise known as triple negative breast cancer (TNBC). To examine distinct morphologically different entities within the tumour, ECM genes which are key regulators of cellular transformations and tissue developmental behaviours were identified from literature for gene expression analysis. The identification of differential mRNA expressions among the genes in each of the cores aims to highlight the heterogeneity within the single tumour and potentially indicate the need.
for individual types of targeted gene therapy. Furthermore, there is a need to elucidate targets specific to TNBC based on the understanding of tumours with heterogeneous cell populations and identification of pathways that may be potential targets for patients.

### 6.3.2 Breast Profiling Findings

For this case study, thirty-four ECM genes were assayed by qPCR in seven tumour samples and also in a matched non-cancerous tissue sample to identify differential gene expression patterns. The selected genes are not clustered on a cancer related pathway or collection of cancer networks, but do constitute the ECM which has important roles in regulating normal cell and tissue development and function (Bissell et al., 1982, Pickup et al., 2014).

![Figure 63](image.png)

**Figure 63** - This image shows the location of the carcinoma within the upper region of the left breast of the patient. The tumour was identified as a metaplastic spindle cell carcinoma (7cm in diameter) negative for ER and PR receptors and had non-elevated levels of ERBB2. Core biopsy locations taken from the tumour are shown also (1-6) along with a normal (7) and random (8) sample.

The seven samples of the tumour were identified as middle-posterior (TH1), middle-superior (TH2), middle-inferior (TH3), middle-anterior (TH4), mid-medial (TH5), mid-lateral (TH6) a random sample (TH8) along with the normal sample (TH7) (Figure 63).
<table>
<thead>
<tr>
<th>Expression</th>
<th>TH1</th>
<th>TH2</th>
<th>TH3</th>
<th>TH4</th>
<th>TH5</th>
<th>TH6</th>
<th>TH7</th>
<th>TH8</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITGA2</td>
<td>1.125</td>
<td>1.558</td>
<td>1.162</td>
<td>-1.013</td>
<td>1.216</td>
<td>1.882</td>
<td>-2.293</td>
<td></td>
</tr>
<tr>
<td>ITGA5</td>
<td>4.013</td>
<td>5.848</td>
<td>5.008</td>
<td>3.773</td>
<td>4.526</td>
<td>6.76</td>
<td>6.901</td>
<td></td>
</tr>
<tr>
<td>ITGA8</td>
<td>1.91</td>
<td>2.595</td>
<td>5.049</td>
<td>2.46</td>
<td>3.046</td>
<td>4.652</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITGA9</td>
<td>8.55</td>
<td>10.449</td>
<td>11.818</td>
<td>10.64</td>
<td>5.428</td>
<td>8.513</td>
<td>3.239</td>
<td></td>
</tr>
<tr>
<td>ITGB4</td>
<td>-2.3923</td>
<td>1.106</td>
<td>-2.222</td>
<td>-3.460</td>
<td>-4.38</td>
<td>-52.632</td>
<td>-4.587</td>
<td></td>
</tr>
<tr>
<td>ITGB5</td>
<td>3.154</td>
<td>3.399</td>
<td>3.803</td>
<td>9.7</td>
<td>4.175</td>
<td>4.517</td>
<td>4.574</td>
<td></td>
</tr>
<tr>
<td>ITGBL1</td>
<td>5.613</td>
<td>5.451</td>
<td>5.464</td>
<td>4.093</td>
<td>1.629</td>
<td>2.193</td>
<td>7.239</td>
<td></td>
</tr>
<tr>
<td>IGF1R</td>
<td>-3.846</td>
<td>-1.287</td>
<td>1.973</td>
<td>1.7</td>
<td>-2.247</td>
<td>-1.9417</td>
<td>-1.666</td>
<td></td>
</tr>
<tr>
<td>IGF2</td>
<td>-2.136</td>
<td>-1.824</td>
<td>1.517</td>
<td>1.08</td>
<td>-2.538</td>
<td>-2.392</td>
<td>-1.683</td>
<td></td>
</tr>
<tr>
<td>IGF2R</td>
<td>5.701</td>
<td>7.392</td>
<td>8.539</td>
<td>12.887</td>
<td>2.488</td>
<td>8.484</td>
<td>2.423</td>
<td></td>
</tr>
<tr>
<td>IGFBP4</td>
<td>2.59</td>
<td>2.356</td>
<td>1.642</td>
<td>1.558</td>
<td>1.454</td>
<td>1.17</td>
<td>1.777</td>
<td></td>
</tr>
<tr>
<td>EGF</td>
<td>-1.1287</td>
<td>-1.594</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2.685</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>EGFR</td>
<td>17.889</td>
<td>11.663</td>
<td>11.227</td>
<td>-1.308</td>
<td>1.837</td>
<td>2.299</td>
<td>1.267</td>
<td></td>
</tr>
<tr>
<td>RACK1</td>
<td>9.521</td>
<td>12.209</td>
<td>5.979</td>
<td>5.657</td>
<td>1.781</td>
<td>5.322</td>
<td>3.518</td>
<td></td>
</tr>
<tr>
<td>PTK2</td>
<td>-1.118</td>
<td>-1.262</td>
<td>1.464</td>
<td>1.077</td>
<td>1.019</td>
<td>1.507</td>
<td>-1.028</td>
<td></td>
</tr>
<tr>
<td>IL8</td>
<td>10.562</td>
<td>10.305</td>
<td>7.342</td>
<td>5.21</td>
<td>16.077</td>
<td>18.884</td>
<td>35.762</td>
<td></td>
</tr>
<tr>
<td>MMP7</td>
<td>-1.136</td>
<td>-2.645</td>
<td>1.072</td>
<td>-8.771</td>
<td>6.543</td>
<td>4.424</td>
<td>-5.076</td>
<td></td>
</tr>
<tr>
<td>MMP9</td>
<td>103.622</td>
<td>168.39</td>
<td>132.807</td>
<td>70.104</td>
<td>16.235</td>
<td>103.071</td>
<td>33.784</td>
<td></td>
</tr>
<tr>
<td>MMP2</td>
<td>7.077</td>
<td>6.422</td>
<td>1.243</td>
<td>-1.112</td>
<td>9.789</td>
<td>5.684</td>
<td>1.183</td>
<td></td>
</tr>
<tr>
<td>MMP3</td>
<td>-7.6923</td>
<td>-9.888</td>
<td>-2.3923</td>
<td>-</td>
<td>6.723</td>
<td>3.093</td>
<td>1.428</td>
<td></td>
</tr>
<tr>
<td>CXCR2</td>
<td>2.807</td>
<td>1.986</td>
<td>2.891</td>
<td>-3.164</td>
<td>2.553</td>
<td>3.343</td>
<td>-3.174</td>
<td></td>
</tr>
<tr>
<td>COL1A1</td>
<td>9.473</td>
<td>5.528</td>
<td>5.315</td>
<td>3.967</td>
<td>14.117</td>
<td>7.35</td>
<td>7.584</td>
<td></td>
</tr>
<tr>
<td>LAMA1</td>
<td>3.552</td>
<td>5.46</td>
<td>10.817</td>
<td>11.372</td>
<td>3.264</td>
<td>1.788</td>
<td>3.524</td>
<td></td>
</tr>
<tr>
<td>CD9</td>
<td>-3.891</td>
<td>1.018</td>
<td>-1.607</td>
<td>-</td>
<td>-3.355</td>
<td>-3.1646</td>
<td>-5.405</td>
<td></td>
</tr>
<tr>
<td>CDH2</td>
<td>-5.154</td>
<td>-2.375</td>
<td>-2.611</td>
<td>-8.620</td>
<td>1.042</td>
<td>-1.087</td>
<td>-1.153</td>
<td></td>
</tr>
</tbody>
</table>

**Table 11** – Differential expression fold change patterns determined by REST software. *Statistically significant is (p<0.05). Statistically significant differences are highlighted using bold font. Dashed lines indicate undetermined levels of expression. The normal sample is normalised to 1 in each case.*
Differences in gene expression were determined by REST© which also identified statistically significant fold changes within the dataset. Identifying differential expression patterns in different subsections of the tumour could potentially highlight the heterogeneous nature of the tumour and identify certain genes which may have the potential to increase the metastatic activity of the malignant cells. When profiled, a clear pattern of differential expression emerged for genes not only between malignant and normal tissue samples but across individual biopsy samples also. The extent of dysregulation varied and not all differentially expressed genes were statistically significant. The entire expression variation can be observed in Table 11.

6.3.3 Dysregulation Discussion

**Matrix Metalloproteinases:** It has been shown that matrix metalloproteinase dysregulation can lead to increased tissue dysregulation (Roy et al., 2009) and therefore enhances malignant cells potential to spread from the primary tumour location to secondary sites making differential MMP expression a potential gauge in TNBC studies. In this study, MMP9 levels across the 7 samples were observed to be the most differential expressed between malignant and normal samples (Figure 64). MMP9 production by tumour and stromal cells is one of the most important factors for metastatic behaviour of tumour cells (Noel et al., 2008). It is one member of the matrix metalloproteinases family which have been shown to be ‘functional digesters’ of tissue and also play key roles in tissue architecture remodelling. Furthermore it has been investigated as a potential biomarker for cancer invasion and metastasis (Merdad et al., 2014). The largest MMP9 fold difference was a 168-fold increase in TH2 followed by a 132-fold and 103-fold increase in TH3 and TH1.

**Matrix Metalloproteinases:** It has been shown that matrix metalloproteinase dysregulation can lead to increased tissue dysregulation (Roy et al., 2009) and therefore enhances malignant cells potential to spread from the primary tumour location to secondary sites making differential MMP expression a potential gauge in TNBC studies. In this study, MMP9 levels across the 7 samples were observed to be the most differential expressed between malignant and normal samples (Figure 64). MMP9 production by tumour and stromal cells is one of the most important factors for metastatic behaviour of tumour cells (Noel et al., 2008).
Experimental Results

It is one member of the matrix metalloproteinases family which have been shown to be 'functional digesters' of tissue and also play key roles in tissue architecture remodelling. Furthermore it has been investigated as a potential biomarker for cancer invasion and metastasis (Merdad et al., 2014). The largest $MMP9$ fold difference was a 168-fold increase in TH2 followed by a 132-fold and 103-fold increase in TH3 and TH1. Increases in the expression of the $MMP9$ gene in carcinomas of the breast have been known to increase the invasive properties of the carcinoma by degradation of the organisational proteins of the ECM and surrounding tissue structures such as collagen and elastin (Wu et al., 2014). The smallest observed fold change of $MMP9$ in this study was in TH6 in which a 16-fold increase in expression. $MMP7$ expression was increased in the mid medial TH6 (+6.5 fold) and mid lateral TH7 (+4.4 fold) samples and decreased in the random sample taken (-5 fold). $MMP2$ expression was increased in four samples with the highest increase in TH6 of 9-fold. $MMP3$ expression varied ranging from -7 to +6 fold changes across the tumour.

**Structural constituents:** Increased levels of collagen have been shown to significantly increase the risk of developing breast cancer due to density changes in the breast tissue (Huo et al., 2015). Further research has also shown that increased

---

**Figure 64** - Matrix Metalloproteinase differential expression in the seven malignant samples in comparison to the matched normal sample. MMP9 expression is highly over expressed in comparison to the other genes assayed.
collagen in breast tissue (Zhang et al., 2013a) and increased linearisation of collagen (Acerbi et al., 2015) could increase the potential of breast tumours spreading and becoming more invasive. This increase in collagen expression could also be correlated to this tumour’s pathological classification of being a metaplastic (spindle cell) carcinoma where the metaplastic cells can be an instigator for increased collagen production (Battifora, 1976).

COL1A1 and COL3A1 expression levels were assessed in the tissue samples and increases in expression of both genes was observed with values ranging between 2 and 14-fold increases. TH6 showed the largest increase in COL1A1 expression (+14-fold), and TH5 had the lowest increase (+3.9 fold). Concurrently, TH6 expression for COL3A1 had also the largest change, increasing approximately 6.8-fold, while TH4 had the smallest increase (+2.4 fold). The largest increases observed are in samples which were taken at the boundaries of the tumour which may indicate increases in collagen deposition at the outer limits as opposed to the central core of the tumour. Further to this, the expression pattern of LAMA1, which is another key structural constituent of the extracellular matrix and tissue microenvironment, is statistically significantly in different sections of the tumour. Laminins are a family of additional ECM components which function to bridge between primary structural constituents to reinforce the overall network as well as connecting the ECM to cells.
and to soluble molecules within the extracellular space (Mouw et al., 2014). Significant differential expression was observed in four of the seven tumour samples in comparison to the adjacent normal tissue. The largest increase in expression for *LAMA1* was observed in the superior mid-section of the tumour (TH5) where expression of the gene was 5 times more than that of the corresponding normal sample. *LAMA1* expression has been implicated in a wide variety of biological processes including cellular differentiation, adhesion and metastasis due to it being a key component of the basement membrane. The dysregulation of structural components in this study can be seen in Figure 65.

**Integrins**: The role of integrins in the initiation and progression of a range of carcinomas, and particularly carcinomas of the breast, from benign hyperplasia to carcinoma in-situ is well documented (Berry et al., 2004, Lambert et al., 2012, Glukhova and Streuli, 2013, White and Muller, 2007). A selection of both the alpha and beta subunits were analysed which are key bridges for cell-cell and cell-ECM interactions. From the selected integrin panel, expression patterns of the targets in the individual samples are well defined and primarily in the same direction, albeit in varying amounts (Figure 66).

![Integrin differential expression in the seven malignant samples in comparison to the matched normal sample](image-url)

**Figure 66** - Integrin differential expression in the seven malignant samples in comparison to the matched normal sample
The largest gene upregulation is observed in *ITGB1* with an average increase of 18-fold in six of the seven samples with the largest being in TH3 (28-fold) followed by TH2 (27-fold) and TH4 (25-fold). Other integrin expression was significantly different although not across all samples which highlights further heterogeneity in this tumour. *ITGA5*, *ITGB5* and *ITGAV* also showed significant increases in expression across six of the samples in comparison to the normal tissue. Expression of *ITGA5* was upregulated by 4-7 fold across the samples while *ITGAV* was more differential but still upregulated by 3-12 fold. Enhancement of *ITGA5* expression can promote *ITGA5* mediated cell adhesion and metastasis in breast cancer (Qin et al., 2011). It is further postulated that *ITGA5* and *ITGAV* play key roles at the boundary between the cell membrane and the ECM (Viana Lde et al., 2013). However, the function and role that these two proteins play in the breast cancer has not been fully elucidated in literature. The relationship between the two genes has also previously been examined in colorectal cancer to identify molecular interactions and the *ITGAV* gene emerged as being associated with higher progression and spread of CRC via perineural invasion (Viana Lde et al., 2013, Waisberg et al., 2014). Perineural invasion was not documented for this tumour so it is unclear if there is evidence to support this hypothesis for this particular carcinoma.

Significant *ITGB5* over expression of 3 to 9.7-fold was observed in six of seven samples. TH7 was the only sample outside statistical remits but it still upregulated nonetheless. Elevated expression of 2 to 5-fold was observed for *ITGA8* in six of eight samples with the exception of TH5 and TH7 where the fold change was not significant. *ITGBL1* was upregulated in 6/7 samples with the highest overexpression observed in TH8 of 7.2 fold. In contrast to the predominantly overexpression of the integrin genes in this study, *ITGB4* was the sole integrin whose expression was for the most part downregulated across the samples. Minor downregulation of approximately 2-4-fold was observed in TH2 and TH6 while the expression of *ITGB4* was upregulated in only the superior mid-section (TH3) where expression was observed to be +2-fold. However this was not deemed statistically significant. Literature observations indicate that *ITGB4* expression in human breast cancer is restricted and associated with basal-like cancers (Lu et al., 2008). The disparity in expression between the samples highlights significant variations of *ITGB4* in the
tumour and may suggest that decreases in the gene could function in concert with other integrin proteins to facilitate the aggressive behaviour of this type of tumour.

**Growth factors:** Growth factors (GFs) have been shown to have significant roles in the development and progression in a number of cancers including breast (Sachdev and Yee, 2001, Kiely et al., 2006), colorectal (Shiratsuchi et al., 2011, Giovannucci, 2001) and prostate (Roberts, 2004, Monti et al., 2007). However, despite promising experimental evidence, the specific role of GFs in tumourgenesis and metastasis is still unclear. At a fundamental level, it is hypothesised that increased *IGF1* and *IGF2* expression levels along with higher levels of circulating *IGFs* lead to increased cellular proliferation, migration and cell survival rates in triple negative cell lines (Davison et al., 2011).

![Growth Factors](image)

**Figure 67** – Growth factor differential expression in the seven malignant samples in comparison to the matched normal sample

Evidence also shows that increased levels of *IGF1* also inhibits apoptosis which greatly influences tumour growth and progression rates (Cory et al., 1999). It is this proliferative and anti-apoptotic effect that can propagate the spread of carcinomas. *IGFs* exert their actions by binding with a receptor at the cell surface. These receptors play a prominent role in IGF activity by either promoting or supressing signalling pathway activities and have been shown to contribute to tumour mass
Experimental Results

Increases or tumour reoccurrence particularly in breast carcinomas. The pattern of differential growth factor expression in this study can be seen in Figure 67. On an overall level, IGF1 gene expression has been reported to be higher in breast cancer than that of normal mucosa with findings indicating a tumour-promoting role (Christopoulos et al., 2015). Decreased levels of IGF are not commonly associated with increased breast cancer risk or progression. However, in this study, expression levels for IGF1 were, on average, statistically significantly lower (~17.7 fold) in the cancer tissue than in the normal mucosa with the largest decrease observed in the superior mid-section TH3 of the tumour (~24.4 fold). This decrease in expression corresponded with decreased levels of IGF1R, the receptor for IGF1, which showed average down regulation of the gene in three of the seven samples (~2.6 fold). IGF protein and receptor cross-linking interactions have a direct effect on cellular activities and can stimulate the activity of a wide range of other molecules such as tumour suppressor genes, other growth factors (Yu and Rohan, 2000) and binding proteins. Binding proteins belong to a family of six proteins which bind to IGF1 and IGF2 with high affinity and modulate their biological actions. One such IGF binding protein which can modulate IGF activity, IGFBP3, was observed to be dysregulated across 5/7 tumour biopsy samples. The dysregulation was increased in all samples by an average statistically fold difference of 6.7-fold with the largest increase observed in the random sample TH8 of 11-fold. Increased expression was also observed in the IGFBP2 gene (+16.4 fold) in TH8 Dysregulation of the gene was not statistically significant elsewhere. Statistically significant IGFBP4 dysregulation was confined to TH2 (+2.6 fold) and TH8 (+1.7 fold). Statistically significant changes in expression were also observed for EGF in TH6 (+2.7 fold) and its receptor EGFR in three of the seven samples, TH3 +11 fold, TH4 +11 fold and TH6 +1.8 fold). Expression levels for EGF were consistently low across the tumour leading to small or undetectable changes in expression. In 4/7 tumour samples analysed, VEGF was over expressed at significantly higher levels as opposed to the corresponding normal tissue. The largest increases of VEGF was observed in TH3 (54 fold), TH6 (+23.9 fold) and TH8 (+16.4 fold). The VEGF signalling system may be a crucial step in increased angiogenic activity in a tumour (Niu and Chen, 2010) and the pattern of dysregulation highlighted here merit the an increased focus on the growth factor family. In conclusion, IGF, IGF receptors, IGF binding protein, EFG and VEGF
expression patterns appear to be consistently differentially expressed across the tumour. *IGF1* is downregulated along with *IGF1R* and each binding protein is upregulated which may be important in influencing cellular behavioural changes within the tumour and signalling mechanisms. Furthermore, consist large variations in gene expression in different tumour sections highlights the importance of these genes in the dynamic microenvironment surrounding the tumour.

**Other ECM Constituents:** Differential expression for these genes can be seen in Figure 68.

![Other ECM Constituents](image)

**Figure 68** – Other ECM constituents differential expression in the seven malignant samples in comparison to the matched normal sample

**IL8:** *IL8* expression was also significantly upregulated in five of the seven malignant samples in comparison to the normal mucosa. *IL8* has been identified in literature as a potential marker for tumour aggressiveness (Freund et al., 2003). Samples TH1 and TH2 displayed an overall increase in expression of 10 fold while samples TH5 and TH6 were upregulated 5 and 16 fold approximately. Interestingly, *IL8* expression in TH8 was significantly increased more than the other samples where the fold increase was approximately 35 fold. As stated previously TH8 is a random core biopsy taken from the tumour with no information of its particular location recorded. The differential expression of *IL8* in this tumour could be due to many factors. Overexpression of *IL8* has been detected in many carcinomas and it has been
suggested that tumour cells produce *IL8* as an autocrine growth factor that function to promote tumour growth, tissue invasion and metastatic spread (Brew et al., 2000). It has been shown that tumour associated fibroblasts and tumour associated macrophages along with mesenchymal stem cells within the tumour microenvironment secrete leukins such as *IL8* that can lead to replenishment of breast cancer stem cells (Korkaya et al., 2011). It has also been shown that *IL8* expression can correlate with ER status where data shows that IL-8 expression is negatively linked to ER status of breast and ovarian cancer cells (Freund et al., 2002). We know from patient histological data that the sample used for this study was identified as being ER negative thus correlating with the large increase in *IL8* observed in this sample.

**POSTN:** Coupled with increased *IL8* production, *POSTN* can also be a key contributor to increased invasiveness of breast cancer cells and is shown to be increased in each of the samples in this study. *POSTN* was statistically differentially expressed in all of the malignant samples. The largest increase in *POSTN* was observed in TH6 (+23 fold), TH3 (+18 fold) and TH2 (+17 fold). Significant increases were also noted in TH7 (+14 fold) and TH4 (+11 fold) while the smallest increases were observed in TH8 (+8.7 fold) and TH5 (+3.4 fold) further indicating *POSTN* expression variations throughout the tumour. This also indicates an influential role for *POSTN* in TNBC development and supports studies where tumours with higher levels of *POSTN* were shown to attain significantly more distant bone metastasis and worse disease-specific survival than those with none or low-expressed *POSTN* protein (Xu et al., 2012).

**SPARC:** The expression of *SPARC* in the TNBC samples overall indicated an upward trend. The dynamics of the *SPARC* gene and its role in various cancers and specifically breast cancer is not completely understood. However, it is predominantly hypothesised that patients expressing high *SPARC* levels have worse disease-free survival and overall survival (Zhu et al., 2016). In the samples analysed, *SPARC* expression was increased in 5/6 samples which were statistically significant. The largest increase was observed in TH3 where the expression of this gene was increased 20-fold.
OUTSTANDING GENES: The pattern of expression was also dysregulated for RACK1, CDH2, CXCR2, CEACAM1, CD9 and PTK2 although the standard error of the expression fell outside the limits of 5% for many samples. Therefore, the dysregulation was deemed statistically insignificant for some of the samples. RACK1 expression was upregulated in all of the malignant samples, with four being statistically significant. CDH2 expression was downregulated in 6 of the 7 malignant samples, with four being statistically significant. CXCR2 expression was upregulated in 5 of the 7 malignant samples and downregulated in 2, with four being statistically significant. CEACAM1 and CD9 expression was downregulated in all of the malignant samples, with two samples being statistically significant. PTK2 expression was only slightly dysregulated in the malignant samples with none of the dysregulation being statistically significant.
6.3.4 Verification study using commercial qPCR instrument

Having determined the gene expression levels of the samples using the microfluidic instrumentation, conformation of the results was sought by performing a parallel real-time PCR experiment on the commercial instrument, the AB7900HT platform. This parallel test has the potential to identify discrepancies that may occur between instruments and investigate whether any large gene expression fold changes were unique to the microfluidic instrument. The largest gene expression fold changes observed when using the microfluidic platform were in the expression of IL-8, IGF1, ITGB1 and MMP9 and therefore these genes were selected for further verification analysis on the commercial platform. Reference genes PMM1, RPLP0 and PGK1 were also included. The layout of the ABI well plate is shown in Figure 69.

Reactions containing gene expression MasterMix and gene assay were prepared with 1:10 dilution cDNA which was synthesised from 500ng/µl of RNA. Since the same concentrations were used on the microfluidic instrument, Cq values for GOIs and HKs are expected to be the same as the microfluidic instrument (ΔCq= 0). The reactions for each gene of interest were performed in triplicate. NTCs were also prepared and run on the same plate to ensure no contamination was present. NTC wells showed no amplification.
Experimental Results

To verify that the results are concurrent between the two platforms, Table 12 shows Cq values for the four GOIs and the three HK genes in the seven cancer samples and the one corresponding normal sample. The threshold was set at 0.1 on both instruments. The Cq values were analysed to determine significant differential gene expression fold changes between the seven malignant samples and corresponding normal sample. In Figure 70, the ABI s-curve shows the difference in Cq values between the normal sample TH1 and the posterior mid section TH2 of the breast tumour for MMP9. From this we can see that, in particular, MMP9 expression was substantially increased in the cancer samples in comparison to the normal sample on both platforms.

<table>
<thead>
<tr>
<th></th>
<th>TH1</th>
<th>TH2</th>
<th>TH3</th>
<th>TH4</th>
<th>TH5</th>
<th>TH6</th>
<th>TH7</th>
<th>TH8</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEI</td>
<td>IL8</td>
<td>31.117</td>
<td>27.681</td>
<td>27.606</td>
<td>28.088</td>
<td>28.129</td>
<td>27.272</td>
<td>27.597</td>
</tr>
<tr>
<td>GEI</td>
<td>IGF1</td>
<td>26.61</td>
<td>30.952</td>
<td>30.969</td>
<td>30.096</td>
<td>30.339</td>
<td>30.687</td>
<td>30.63</td>
</tr>
<tr>
<td>ABI</td>
<td>IGF1</td>
<td>28.968</td>
<td>31.907</td>
<td>32.685</td>
<td>29.322</td>
<td>31.275</td>
<td>31.638</td>
<td>29.69</td>
</tr>
<tr>
<td>ABI</td>
<td>ITGB1</td>
<td>25.304</td>
<td>23.015</td>
<td>23.068</td>
<td>22.85</td>
<td>24.478</td>
<td>23.57</td>
<td>22.595</td>
</tr>
<tr>
<td>ABI</td>
<td>RPLP0</td>
<td>22.103</td>
<td>22.033</td>
<td>23.86</td>
<td>24.95</td>
<td>22.705</td>
<td>23.188</td>
<td>23.265</td>
</tr>
<tr>
<td>GEI</td>
<td>PGK1</td>
<td>23.382</td>
<td>23.185</td>
<td>23.325</td>
<td>23.227</td>
<td>23.081</td>
<td>23.561</td>
<td>23.579</td>
</tr>
</tbody>
</table>

**Table 12- Cq Values for GEI versus ABI verification study**
Experimental Results

Overall, differences in expression between control sample and cancer samples were assessed in group means for statistical significance by randomisation tests using REST Software. The pattern of fold change in the GOIs was similar on both instruments (Table 13) with the expression intensities also replicated in most samples, only limited by the sensitivity of the platforms, smoothening software in the commercial platform and stability of reference genes. Reference gene variation was between the acceptable range of 0.8 and 1.2 cycles showing that the reference genes were suitably stable for the study.

Table 13 – Expression fold changes between normal sample and corresponding malignant samples. Data which is shaded is deemed statistically significant at p<0.05 level. Green indicates upregulation while red indicated downregulation.
Experimental Results

Figure 71 – Expression fold changes between normal sample and corresponding malignant samples as determined by the two platforms.

When plotted, it is clear that the dysregulation pattern for the four selected genes does vary significantly across the tumour depending on which malignant sample is analysed (Figure 71). These results confirm the hypothesis that a high degree of tumour heterogeneity exists within individual tumours. More explicitly, in this tumour, a very large increase in \textit{MMP9} expression, a substantial increase in \textit{IL8} expression (in particular in the random sample TH8), a varying decrease in \textit{IGF1} expression and an increase of varying magnitude in \textit{ITGB1} expression was observed when using the microfluidic platform and verified using a commercial platform. This verification study therefore highlights a high degree of dysregulation in the ECM genes assayed indicating individual gene expression variations within a TNBC tumour along with a high degree of tumour heterogeneity within the tumour itself. Also the study indicates reproducibility on the microfluidic platform when compared with a commercial platform that indicates that the platform would be suitable for further high throughput studies.
6.4 Chapter Closure

This chapter has demonstrated successful genetic profiling of twenty-four matched colorectal tissue samples and seven tissue biopsies from a single breast carcinoma and matched normal using both microfluidic and well based techniques. The identification of genes which are dysregulated can allow researchers to build on past physiological understanding of carcinomas and expand this to include the molecular mechanisms which influence cancers.

The primary study highlights ECM genes which show differential expression between colorectal tumour tissue and matched normal colon mucosa. Expression patterns observed in the colorectal study highlighted statistically significant differential expression in 14 of the 24 genes assayed. The genes which emerged as being differentially expressed have been previously shown in literature to play key roles in the carcinoma process for a range of individual cancers and in particular colorectal carcinoma. A pairwise correlation pattern emerged also which showed that certain genes were co-expressed. Further to this, correlations emerged between the expression pattern and clinicopathological parameters such as patient gender, tumour diameter, lymphovascular invasion and nodal status which were recorded during surgical resection. The genes identified in this pilot study provides insight into differential expression patterns during the development and progression of colorectal cancer. The identified genes may also represent important targets that could be integrated into clinicopathological tools in the management and treatment of the disease. The results determined by microfluidic means were verified on a commercial platform.

In the second part of the study, tumour heterogeneity in a TNBC sample was demonstrated by assessing seven individual core sections of a tumour and comparing them to a corresponding normal sample from the same patient. Primarily, equivalent patterns of expression were observed but the magnitude of expression varied significantly within individual sections. These genes could play a role in the progression of the tumour or lead to increase metaplastic behaviour which could hinder effective therapeutic strategies. The pattern of dysregulation observed within this tumour signifies a possible ECM signature profile which may help refine the selection process of TNBC patients for alternative treatment therapies. The results
determined were again verified on a commercial platform which produced concurrent results.

Overall, the global dysregulation observed in these genetic studies of ECM expression in different malignant samples indicate that as a cancer develops and progresses it can lead to substantial ECM remodelling along with increased vasculature development leading to heightened metastatic potential. Expanding the patient cohort and genetic panel further would yield additional potential biomarkers for metastasis and other targets. Based on the results, we hypothesis that sets of ECM genes which are essential components for tissue and cellular homeostasis in the normal network are becoming altered, leading to the formation of a tumour cell-hosting environment or cancer network and so may serve as an indicator for tumourgenesis and progression. This goes to highlight that the optimal treatment of patients with cancer depends on establishing accurate diagnoses by using a complex combination of clinical, histopathological and genetic data.
CHAPTER 7

CONCLUSIONS AND RECOMMENDATIONS
7 Conclusions and Recommendations

7.1 Conclusions

By examination of the literature, it was observed that some current instrumentation for PCR experimentation were limited in their capabilities by utilising large reaction volumes and low-medium throughput methods. Additionally, it was noted that large amounts of fresh patient tissue samples were often difficult to attain which highlighted a need for an alternative, novel method for carrying out genetic profiling in a more innovative way. Therefore, the primary objective of this research was to develop a continuous flowing microfluidic platform to identify genetic dysregulation in patient samples using quantitative PCR methods. In this chapter, the conclusions from this research study are presented and recommendations for future work are proposed.

- In order to achieve the objective, a continuous-flow quantitative platform was assembled and characterised.

- Three key instrumentation modules, previously designed, were adopted and assembled with additional components into a stand-alone platform to create a high-throughput microfluidic PCR system.

- Droplet generation was achieved using liquid bridge technology and produced consistent low volume droplets in the order of 300nl containing DNA and reagents for experimentation. Optical detection systems were calibrated and dye calibration experiments completed to achieve optimum fluorescent detection from micro droplets.

- The instrumentation was further validated by an array of genetic tests to confirm that reproducible data could be obtained from future studies.

- By the demonstration of this technology, it is evident that it could be used for other studies examining tumour heterogeneity, single cells etc. achieving high throughput screening of approximately 1,000 reactions per hour. Furthermore, the use of small sample volumes of 6ul of DNA, 45nl of polymerase and 4.5nl of gene assay per 100 reactions would be very advantageous.
Conclusions and Recommendations

- Research was conducted using microfluidic droplet qPCR for analysis of 34 ECM genes in 24 matched colorectal tumour samples yielded the following results:

- Statistically significant differential gene expression between normal and malignant colorectal tissue was observed in fourteen genes from the panel selected for analysis.

- Of these 14 genes, 9 genes were upregulated while 5 were downregulated when assessed in the cohort overall.

- \textit{MMP7} showed the greatest upregulation with increased expression of over 31-fold. Following this \textit{IL8} expression was up 18-fold and \textit{ITGBL1} expression was up 11-fold.

- The greatest decrease in expression was \textit{IGF1} expression which was down regulated 8-fold

- Correlations between co-expressed genes were identified using statistical methods.

- Certain gene expression patterns in the malignant samples correlated with histological parameters indicating a possible link between gene expression and factors such as invasiveness, diameter and mucinous components.

- Dysregulation was observed in certain genes with analysed with UICC staging identifying potential early and late stage tumour markers.

- The expression patterns of five of the most dysregulated genes on the microfluidic platform were verified on a commercial platform.

- For the analysis of the 34 ECM genes in seven breast tumour samples and one corresponding normal sample, the following results were determined:

- Gene expression patterns for the selected genes were diverse firstly between the normal and each of the cancer samples and secondly across individual samples highlighting a large element of tumour heterogeneity existing within the tumour.
Conclusions and Recommendations

- Expression in four genes *MMP9*, *COL1A1* and *POSTN*, *ITGA5* were statistically upregulated in all seven malignant samples. The largest expression increase was observed in *MMP9* expression (average +89-fold increase).

- All of the structural constituent genes analysed (*COL1A1*, *COL3A1* and *LAMA1*) showed increased expression across the samples although some were not statistically significant. The magnitude of upregulation was varied across the samples highlighting a degree of stromal gene expression.

- The expression patterns of four of the most dysregulated genes identified on the microfluidic platform were verified on the commercial platform.

- The successful utilisation of novel microfluidic technology to perform biological experiments potentially allows us to maximise biomarker discovery rates and improve cancer patient prognosis.

The conclusions presented here are a demonstration of the microfluidic platform and the potential for future revisions of this instrument to be used for conducting high-throughput differential gene expression studies, determining heterogeneity within individual tumours as demonstrated here, as well as further studies of single cell analysis, epigenetic modifications etc. Utilising the next generation platforms of similar design and throughput for biological assays could allow for a more revised consideration when selecting therapeutic strategies for patients.

### 7.2 Recommendations for future work

On completion of experimental and theoretical efforts for this research, some recommendations for future work from both an engineering and biological perspective emerged which would have a positive impact on the advancement of this research.

#### 7.2.1 Instrumentation Recommendations

- To design a simplified pumping system for the transport of fluid within the system. The current system is slightly complicated and tubing lengths are
Conclusions and Recommendations

elongated which could lead to minor pressure changes within the system depending on the position and quantity of droplets. This differential pressure gradient could lead to increased droplet stresses and increase the risk of droplet rupture. Further to this, reducing the distance from the sample pick-up point to the mixing station in the liquid bridge would further reduce pressure drops in conduits and allow for more accurate volumes to be aspirated and faster dipping sequences to be implemented. These changes would also allow for the reduction of the number of air traps and bypass systems which are installed as a corrective measure.

- Introducing an incremental dipping setting into the robotic stages would be beneficial. Modifying the dipping depth could prevent reduced quantities of sample or reagent from being aspirated due to volumes in well plates decreasing. Currently the system decreases to the maximum safe distance for approximately 10mm from dip 1 and continues for all subsequent dips. This could be achieved by programming the stages to decrease approximately 200 microns extra into the well after each experimental triplicate train of reactions from a start height of ~4mm.

- Incorporate a fully automated priming system to allow the instrumentation to be quickly refilled with carrier fluid or purged with a cleaning agent. Possible contamination was constantly monitored and controlled in the current system by individual repeated tests. However, an automated system would cleanse the instrument lines at certain time points and prevent the risk of elements remaining in the system and allow these contamination tests to be reduced or discontinued.

- Increase the overall number of thermal cycles to 50 cycles. This would allow more thermal cycling time for amplification and possibly increase detection of low-abundant targets in samples. Alternatively, the inclusion of a pre-amplification step to enrich targets of interest in the original sample would aid with detecting low levels of expression.
7.2.2 Biological Recommendations

- Further studies involving larger patient population size (>100 pax) are merited to increase the potential for a more informed diagnosis. Furthermore, increasing the cohort would allow for more statistical tests to be performed on different grouping of patients along with increasing the robustness of current analysis tests.

- Further studies involving larger gene panels from this matrisome would allow for more comprehensive gene networks to be established. The mammalian core matrisome or ECM is composed of approximately 300 proteins. In addition, there are large numbers of ECM-modifying enzymes, ECM-binding growth factors, and other ECM-associated proteins. Thirty-four genes were analysed in this thesis.

- It would be beneficial to analyse the expression of the candidate ECM genes in a wider variety of tissue types to understand if the patterns of dysregulation are tissue/tumour specific. The potential to have panels of ECM genes which influenced cancer progression for a number of cancers would be beneficial as a support tool for patient prognosis.

- Further validate ECM genetic dysregulation from both the colorectal and breast studies by determining how changes in expression in subsets of these protein encoding genes influence the morphological behaviour of colon cancer cells. By employing a 3D cellular model to determine these behaviours, the cell-cell and cell-matrix interactions would reflect better those of tissues in the body and achieve an accurate model system to validate the ECM signature.

- It is suggested that further in-vitro tests be performed. Malignant cells, representing different tumours, should be transfected to suppress the expression of certain genes from the panel (i.e. most dysregulated). Parallel tests could be performed on analysis platforms to monitor differentiation, growth, proliferation, invasion and metastasis. Additionally, confocal and epifluorescence techniques could be used to examine how the apical-basal
axis of polarity is affected by suppression of genes identified in the candidate panel.

- It is recommended to perform interferometric analysis to study cell behaviour when the substratum of the cells is changed. By varying concentrations of key ECM constituents such as laminin, collagen and fibronectin (which were dysregulated in the expression analysis), the adhesion and spreading of cancer cells could be monitored in real time on platforms such as the xCELLigence DP instrument.

7.2.3 Pathological Recommendations

- Design a real-time method (i.e. mobile application) to allow surgeons and clinical practitioners to accurately record histological parameters and update data accordingly more efficiently that currently possible. A large amount of time was expended on retrieving individual patient data for analysis. Certain important parameters were either poorly recorded or not at all making analysis of results more difficult. If it was possible to record and document this data in a more efficient manner, the process would be simplified and allow for expedited generation of results.


ALTINAY, S 2016. *Is Extracellular Matrix a Castle Against to Invasion of Cancer Cells?, Tumor Metastasis, Dr. Ke Xu (Ed.), InTech.*


[Accessed 29 June 2016].


ARABZADEH, A. & BEAUCHEMIN, N. 2012. Stromal CEACAM1 expression regulates colorectal cancer metastasis. OncoImmunology, 1, 1205-1207.


BARRIER, A., LEMOINE, A., BOELLE, P.-Y., TSE, C., BRAULT, D.,
CHIAPPINI, F., BREITTSCHNEIDER, J., LACAINÉ, F., HOURY, S.,
HUGUER, M., VAN DER LAAN, M. J., SPEED, T., DEBUIRE, B.,
FLAHU, A. & DUDOIT, S. 2005. Colon cancer prognosis prediction by

BARRON, J. 2010. *Understanding the Colon & Intestinal Cleansing* [Online]. The

BARTSCH, M. S., EDWARDS, H. S., LEE, D., MOSELEY, C. E., TEW, K. E.,
RENZI, R. F., VAN DE VREUGDE, J. L., KIM, H., KNIGHT, D. L.,
SINHA, A., BRANDA, S. S. & PATEL, K. D. 2015. The Rotary Zone
Thermal Cycler: A Low-Power System Enabling Automated Rapid PCR.

BASERGA, R. 2000. The contradictions of the insulin-like growth factor 1 receptor.
*Oncogene*, 19, 5574-5581.

BATTIFORA, H. 1976. Spindle cell carcinoma. Ultrastructural evidence of
squamous origin and collagen production by the tumor cells. *Cancer*, 37,
2275–2282.

BEATING BOWEL CANCER. 2016. *Peritoneal metastases* [Online]. Available:
https://www.beatingbowelcancer.org/understanding-bowel-cancer/secondary-
bowel-cancer/peritoneal-metastases/ [Accessed 29 June 2016].

of Disease in Childhood. Education and Practice Edition*, 98, 236-238.

BELGRADER, P., YOUNG, S., YUAN, B., PRIMEAU, M., CHRISTEL, L. A.,
Notebook Thermal Cycler for Rapid Multiplex Real-Time PCR Analysis.
*Analytical Chemistry*, 73, 286-289.

expression and survival in human breast cancer. *European Journal of
Surgical Oncology (EJSO)*, 30, 484-489.

BERTUCCI, F., SALAS, S., EYSTIERIES, S., NASSER, V., FINETTI, P.,
GINESTIER, C., CHARAFE-JAFFRETT, E., LORIOD, B., BACHELART,
L., MONTFORT, J., VICTORERO, G., VIRET, F., OLLENDORFF, V.,
FERT, V., GIOVANNINI, M., DELPERO, J.-R., NGUYEN, C., VIENS, P.,


suppresses colorectal cancer by regulating IGF-1 mediated cell survival. Oncotarget. doi, 10.


FAKIH, M. G. & PADMANABHAN, A. 2006. CEA monitoring in colorectal cancer. What you should know. Oncology (Williston Park), 20, 579-87; discussion 588, 594, 596 passim.

(PGK1) as a reference gene for quantitative gene expression measurements in human blood RNA. *BMC Res Notes*, 4, 324.


180


quantitation of HER2 expression in FFPE breast cancer samples. Methods, 59, S20-3.


Leslie, M. 2009. When cancer cells can't let go. The Journal of Cell Biology, 185, 178.


prognostic molecular markers for cancer medicine. *Therapeutic Advances in Medical Oncology*, 2, 125-148.


PURI, D. 2014. Textbook of Medical Biochemistry, Elsevier Health Sciences APAC.


REVIEWS, C. 2015. Principles And Practice Of Toxicology In Public Health: Medicine, Medicine, Cram101.


RONG ZHANG, G.-L. X., YIN LI, LONG-JUN HE, LI-MING CHEN, GUO-BAO WANG, SHI-YONG LIN, GUANG-YU LUO, XIAO-YAN GAO, AND HONG-BO SHAN; 2013 The role of insulin-like growth factor 1 and its receptor in the formation and development of colorectal carcinoma; *Journal of International Medical Research;* 41, 1228-1235.


SHAHEEN, R., TSENG, W., VELLAGAS, R., LIU, W., AHMAD, S., JUNG, Y., REINMUTH, N., DRAZAN, K., BUCANA, C., HICKLIN, D. & ELLIS, L.


206


time reverse transcription polymerase chain reaction: normalization to rRNA or single housekeeping genes is inappropriate for human tissue biopsies. *Anal Biochem*, 309, 293-300.


BIBLIOGRAPHY


Viana Lde, S., Affonso, R. J., Jr., Silva, S. R., Denadai, M. V., Matos, D., Salinas De Souza, C. & Waisberg, J. 2013. Relationship between the expression of the extracellular matrix genes SPARC, SPP1, FN1, ITGA5 and ITGAV and clinicopathological parameters of tumor progression and colorectal cancer dissemination. Oncology, 84, 81-91.


complete genome of an individual by massively parallel DNA sequencing. *Nature*, 452, 872-6.


ZHANG C-S, L. Y.-Y., WANG H-Y. 2011. Rapid amplification and detection of foodborne pathogenic rotavirus by continuous-flow reverse transcription-
polymerase chain reaction integrated with online fluorescence analysis ; Chin J Anal Chem., 39, 645–651.


Appendix A

Expression associations with histopathological parameters