Biological Characteristics of Gastroesophageal Junction Adenocarcinomas and their Clinical Correlates

Submitted for the degree of M.D. in Histopathology and Morbid Anatomy

Trinity College Dublin

Coláiste na Tríonóide, Baile Átha Cliath

The University of Dublin

Dr. Kate Dinneen

MB BCh BAO

2020

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

_____

Dr. Kate Dinneen

March 2020

Table of Contents

List of Figures

List of Tables

List of Appendices

4.2 Significant associations between clinical outcomes and increased miRNA expression.

4.3 Significant associations between clinical outcomes and increased protein expression.

i. Summary

Gastroesophageal junction adenocarcinomas (GEJA) have increased in incidence in the Western world over the last 50 years, a trend largely attributed to lifestyle factors including obesity and gastroesophageal reflux disease. Their prognosis is poor, and treatment is often complicated by resistance to conventional anti-cancer therapies. Current drug therapies and management strategies used in GEJA have been largely inferred from studies looking at gastric and oesophageal adenocarcinomas, however, a growing school of thought exists which believes that GEJAs have a distinct molecular profile. Investigation of the molecular biology of these tumours may therefore provide us with a novel target for drug therapies, in addition to potential prognostic biomarkers for use in the clinical setting.

Cancer stem cells (CSC) have been extensively investigated across a range of solid organ and haematological malignancies due to their known role in tumorigenesis, invasion, metastasis and drug resistance, yet little is known about their role in GEJA. This thesis investigates the presence of CSC-like cells in GEJA by analysis of the expression patterns of mRNAs, miRNAs and proteins which have previously been described as CSC markers. Molecular markers of EMT were additionally investigated using the same techniques due to the known role of EMT in the regulation of CSCs. This expression data was analysed to seek a significant molecular expression pattern that may be of use in identification, prognostication and/or treatment of aggressive disease.

Chapter 1 presents a background on GEJA, with particular emphasis on advances in tumour classification and drug therapies currently used in the treatment of this disease.

Chapter 2 describes the lab techniques used throughout this work, including quantitative real time PCR and immunohistochemical analysis of tissue microarrays. Chapter 3 presents the expression data for each mRNA, miRNA and protein analysed in this study and correlates each individual marker with patient-specific clinical outcomes. Chapter 4 describes the clinical utility of a predictive model based upon the combined expression data from the previous chapter. This chapter additionally interrogates the mRNA and miRNA expression data to determine the signalling pathways involved in the regulation of CSC-like cells in GEJA. Finally, chapter 5 discusses the significance of these findings in the context of the current literature.

ii. Acknowledgements

This research project was joint funded by awards from the BDIAP, TTMI and Trinity College Dublin. I would like to thank these institutions for their support, without which this project would not have been possible. I would additionally like to thank the St. James's Hospital Upper Gastrointestinal Biobank for providing the clinical samples used in this thesis.

I would like to extend my gratitude to everyone who played a role in the completion of this project. To Orla Sheils, who has the innate ability to make any stresses melt away with her practicality, intelligence, wit and kindness. To Ciara Ryan, whose continued support and attention to detail has been instrumental to this project. To Anne Marie Baird, who I cannot thank enough for her support, help and friendship throughout this process. To Shane Brennan and Marvin Lim, who were a constant source of encouragement and fun, and whose friendship I truly value. To Paul Smyth, Julie McFadden and Marie Reidy who all went out of their way to help in this project and whose assistance will always be appreciated.

Finally, I would like to thank my family, to whom this work is dedicated. To my parents, Audrey and Sean, who taught me the value and importance of hard work and who have been encouraging and supportive of everything I have ever done. To my brother and sisters, Sarah, David and Molly, who have all turned into extraordinary people and without whom I would be half the woman I am now. And to Kevin Byrnes – my best friend and the love of my life. Thank you for everything.

iii. Awards and Funding

- BDIAP Travel Fellowship, 2018
- Ray O'Meara Research Fund Award, TCD, 2018
- Building Engagements in Health Research Award, TTMI, 2019
- BDIAP Glasgow 2020 Educational Fellowship

iv. Abstracts and Presentations

- "Biological characteristics of gastroesophageal junction adenocarcinomas and their clinical correlates: the role of cancer stem cells and epithelial mesenchymal transition" – Royal College of Physicians Faculty of Pathology Annual Symposium 2020, poster presentation
- "The role of cancer stem cells in gastroesophageal junction adenocarcinomas" – abstract accepted for 32nd Congress of the European Society of Pathology and XXXIII International Congress of the International Academy of Pathology 2020

v. Abbreviations

AC	Anatomical Cardia
ACF	Adriamycin, Cisplatin and 5-FU
ACVR1	Activin A Receptor Type 1
ACVR1C	Activin A Receptor Type 1C
ACVR2A	Activin A Receptor Type 2A
ACVR2B	Activin A Receptor Type 2B
ALDH	Aldehyde Dehydrogenase
ALDH1A1	Aldehyde Dehydrogenase 1 Family, Member A1
AMNCH	Adelaide and Meath Hospital Dublin Incorporating the National Children's Hospital
ASR	Age Standardized Rate
ATP	Adenosine Triphosphate
BDIAP	British Division of the International Academy of Pathology
BMPR1B	Bone Morphogenetic Protein Receptor Type 1B
BO	Barrett's Oesophagus
CAF	Cancer Associated Fibroblasts
CC1	Cell Conditioning Solution 1
CDH-1	Cadherin 1
CDKN1B	Cyclin Dependent Kinase Inhibitor 1B
cDNA	Complimentary DNA
CIN	Chromosome Instability
CNS	Central Nervous System
COX2	Cyclooxygenase-2
CPR	Complete Pathologic Response
CR	Complete Response
CSC	Cancer Stem Cells
C_T	Cycle Threshold
CTC	Circulating Tumour Cells
CTLA-4	Cytotoxic T-Lymphocyte Associated Antigen-4
DAB	3, 3'-Diaminobenzidine Tetrahydrochloride
DAVID	Database for Annotation, Visualization and Integrated Discovery
DFS	Disease Free Survival
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide Triphosphate
DUSP9	Dual Specificity Phosphatase 9
DTC	Disseminated Tumour Cells
EBV	Epstein Barr Virus
E-cadherin	Epithelial Cadherin

ECF	Epirubicin, Cisplatin, and Fluorouracil
ECX	Epirubicin, Cisplatin and Capecitabine
EFS	Event Free Survival
EGF	Epidermal Growth Factor
E/M	Epithelial Mesenchymal
EMT	Epithelial Mesenchymal Transition
EpCAM	Epithelial Cell Adhesion Molecule
ERK	Extracellular Signal–Regulated Kinases
EtOH	Ethanol
FISH	Fluorescent In Situ Hybridization
FFPE	Formalin Fixed Paraffin Embedded
FOXC	Forkhead Box C
FP	5-Fluorouracil
FZD3	Frizzled Class Receptor 3
FZD10	Frizzled Class Receptor 10
GAPDH	Glyceraldehyde 3-Phosphate Dehydrogenase
GAS6	Growth Arrest-Specific Protein 6
GEJ	Gastroesophageal Junction
GEJA	Gastroesophageal Junction Adenocarcinoma
GORD	Gastroesophageal Reflux Disease
GSK3β	Glycogen Synthase Kinase 3 Beta
HER2	Human Epidermal Growth Factor Receptor 2
HGF	Hepatocyte Growth Factor
HIF1α	Hypoxia Inducible Factor 1α
HMLE	Human Mammary Epithelial Cell
HRD	Homologous Recombination Deficiency
H_2O	Water
H_2O_2	Hydrogen Peroxide
H&E	Haematoxylin and Eosin
ICAM1	Intercellular Adhesion Molecule 1
ICI	Immune Checkpoint Inhibitors
ID	Inhibitor of Differentiation
IHC	Immunohistochemistry
Il-6	Interleukin 6
Il-12	Interleukin 12
IL6ST	Interleukin 6 Signal Transducer
IM	Intestinal Metaplasia
iPSC	Induced Pluripotent Stem Cells
IR	Infra-Red
ISL1	ISL LIM Homeobox 1
JARID2	Jumonji and AT-Rich Interaction Domain Containing 2
KEGG	Kyoto Encyclopaedia of Genes and Genomes
Klf-4	Kruppel-Like Factor 4

LCM	Laser Capture Microdissection
MAPK	Mitogen-Activated Protein Kinase
MAPK13	Mitogen-Activated Protein Kinase 13
Mbp	Million Base Pairs
MET	Mesenchymal Epithelial Transition
miR	Micro RNA
miRNA	Micro RNA
MMP	Matrix Metalloproteinase
mRNA	Messenger RNA
MSI	Microsatellite Instable
MUC-1	Mucin 1 Cell Surface Associated
N	Number
NAT	Neoadjuvant Treatment
N-Cadherin	Neural Cadherin
NDP	NanoZoomer Digital Pathology
NODAL	Nodal Growth Differentiation Factor
NSCLC	Non-Small Cell Lung Cancer
O_2	Oxygen
OAC	Oesophageal Adenocarcinoma
Oct-4	Octamer Binding Transcription Factor 4
ORR	Overall Response Rate
OS	Overall Survival
Pai-1	Plasminogen Activator Inhibitor-1
PAK1	p21 Activated Kinase 1
PALS1	Protein Associated with Lin-7 1
PATJ	PALS1-Associated Tight-Junction Protein
PCGF2	Polycomb Group Ring Finger 2
PCGF6	Polycomb Group Ring Finger 6
PCR	Polymerase Chain Reaction
PD-1	Programmed Death 1
PDGF	Platelet Derived Growth Factor
PDL-1	Programmed Death Ligand 1
PFS	Progression Free Survival
PI3K	Phosphoinositide 3-Kinases
POU5F1	POU Class 5 Homeobox 1
PR	Partial Response
PRC2	Polycomb Repressive Complex 2
PRRX1	Paired Related Homeobox 1
qPCR	Quantitative Polymerase Chain Reaction
RAF1	Raf-1 Proto-oncogene, Serine/Threonine Kinase
REST	RE1 Silencing Transcription Factor
RIF1	Replication Timing Regulatory Factor 1
RNA	Ribonucleic Acid
ROC	Receiver Operating Characteristic Curve
RODI	Reverse Osmosis Deionized
RQ	Relative Quantification

RT	Reverse Transcriptase
RTK	Receptor Tyrosine Kinase
SCC	Squamous Cell Carcinoma
SEER	Surveillance Epidemiology and End Results
SEM	Standard Error of the Mean
SIP1	Survival of Motor Neuron Protein Interacting Protein 1
SJH	St. James's Hospital
SKIL	SKI like Proto-Oncogene
SMAD3	SMAD Family Member 3
SMAD5	SMAD Family Member 5
snRNA	Small Nucleolar RNA
SPARC	Secreted Protein Acidic and Rich in Cys
TBST	Tris Buffered Saline with Tween
TCF	T Cell Factor
TCF3	Transcription Factor 3
TF	Transcription Factor
TGF-β	Transforming Growth Factor Beta
TIL	Tumour Infiltrating Lymphocyte
Tis	Carcinoma In Situ
TKI	Tyrosine Kinase Inhibitor
TMA	Tissue Microarray
TNF	Tumour Necrosis Factor
TRG	Tumour Regression Grade
TSP-1	Thrombospondin-1
TTMI	Trinity Translational Medicine Institute
uPAR	Urokinase-Type Plasminogen Activator Receptor
UV	Ultra-Violet
WNT5B	Wnt Family Member 5B
WNT8A	Wnt Family Member 8A
WNT9B	Wnt Family Member 9B
WNT10B	Wnt Family Member 10B
WNT16	Wnt Family Member 16
VEGFR-2	Vascular Endothelial Growth Factor Receptor-2
XP	Cisplatin Plus Capecitabine
Zic3	Zic Family Member 3
ZO-1	Zona Occludens 1
5-FU	Fluorouracil

Chapter One: Introduction

1.1 Gastroesophageal Junction Adenocarcinoma

1.1.1 Introduction

Gastroesophageal junction adenocarcinomas (GEJA) are cancers which straddle the junction between the oesophagus and stomach, sharing similar epidemiological characteristics and risk factors to oesophageal adenocarcinoma (OAC) (Bray et al., 2018). The incidence of GEJA has dramatically increased by approximately 600% since the 1970s (Rubenstein & Shaheen, 2015). The Irish 5 year survival rate is 22.6%, which is on par with the international rate of 19.9%, however the prognosis remains poor irrespective of geographic region (NCRI, 2019; SEER, 2019). Despite early advances in treatment modalities, rates of disease recurrence and resistance to anti-cancer therapies remain high (Brungs et al., 2019). Extensive research into the biological nature of GEJA is required to improve both therapeutic options and survival rates.

1.1.2 Epidemiology

Oesophageal and proximal gastric cancers are malignancies with a poor prognosis, regardless of histologic subtype, accounting for combined global deaths in excess of 1.2 million in 2018 (Ferlay et al., 2019). Gastroesophageal junction (GEJ) cancers are those which involve the lower oesophagus, proximal stomach and the junction between the two. This junction is defined histologically as the transition point between stratified oesophageal squamous epithelium and columnar gastric epithelium. The two primary GEJ cancer subtypes are adenocarcinoma and squamous cell carcinoma (SCC), with stark contrast between the two in terms of aetiology and geographic distribution. SCC accounts for 90% of oesophageal cancers worldwide, with major risk factors including alcohol and

tobacco smoking (Smyth et al., 2017; Trivers, Sabatino, & Stewart, 2008). It is predominantly observed in lower income countries including parts of Asia and Sub-Saharan Africa (Smyth et al., 2017). By comparison, adenocarcinomas represent the majority of GEJ malignancies in industrialized countries centred in Northwest Europe and North America (Batool, Khan, Akbar, & Ashraf, 2019; Bray et al., 2018; Buas & Vaughan, 2013; Kumamoto et al., 2019; Rubenstein & Shaheen, 2015) (Figure 1.1).

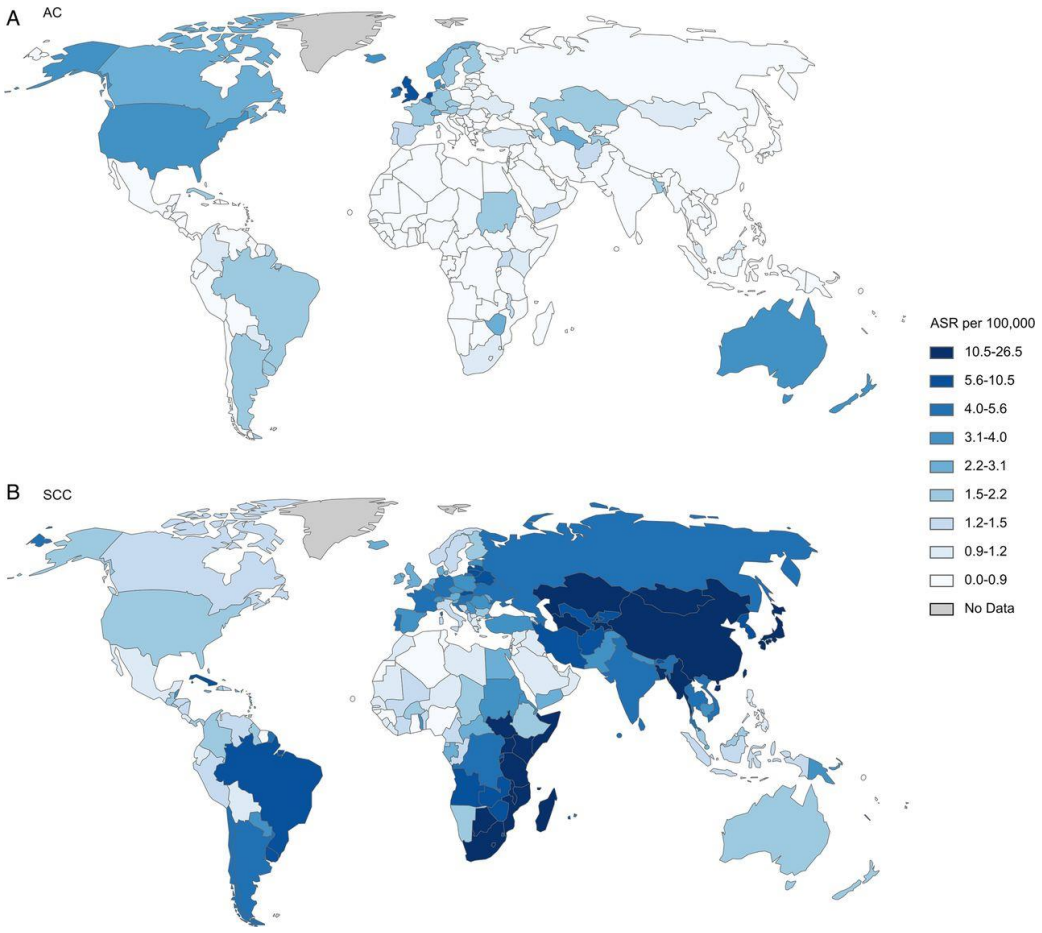


Figure 1.1: Global Age Standardized Incidence Rate (ASR) of OAC versus Oesophageal Squamous Cell Carcinoma (SCC) in 2012.

Image A shows the distribution of OAC (labelled as AC in this diagram), predominantly centred in Western countries. Image B shows the different geographic distribution patterns of oesophageal SCC, most densely located in Asia and parts of Sub-Saharan Africa. Figure taken from Arnold *et al*, 2012 (Arnold, Soerjomataram, Ferlay, & Forman, 2015).

The incidence of GEJA has markedly increased over the last 50 years. Whilst this trend is mirrored by a decrease in rates of distal gastric cancer – a trend which is predominantly attributable to recent improvements in treatment of *Helicobacter pylori* infection – distal gastric cancer nevertheless remains more prevalent than GEJA worldwide (Battaglin, Naseem, Puccini, & Lenz, 2018; Rawla & Barsouk, 2019). The epidemiological shift seen in GEJA can be partially accounted for by aspects of the Western lifestyle including diet, obesity, smoking and gastro-oesophageal reflux disease (GORD) (Buas & Vaughan, 2013). However, despite these known associations, the precise cause of the rise of GEJA remains unclear. Whilst traditionally GEJA has been subdivided into cancers of either gastric or oesophageal origin, many now believe that malignancies of the GEJ are best regarded as a separate disease entity (Hayakawa, Sethi, Sepulveda, Bass, & Wang, 2016). Much of our current knowledge about GEJA is inferred from studies conducted on oesophageal and gastric adenocarcinomas, highlighting the need for further research into the epidemiology, management, molecular biology and prognosis of these tumours as a distinct entity.

1.1.3 Risk Factors

There are many known risk factors for OAC and GEJA in addition to the lifestyle influences previously mentioned. These include male sex, advancing age and Caucasian ethnicity (Buas & Vaughan, 2013; Schneider & Corley, 2017) (Figure 1.2).

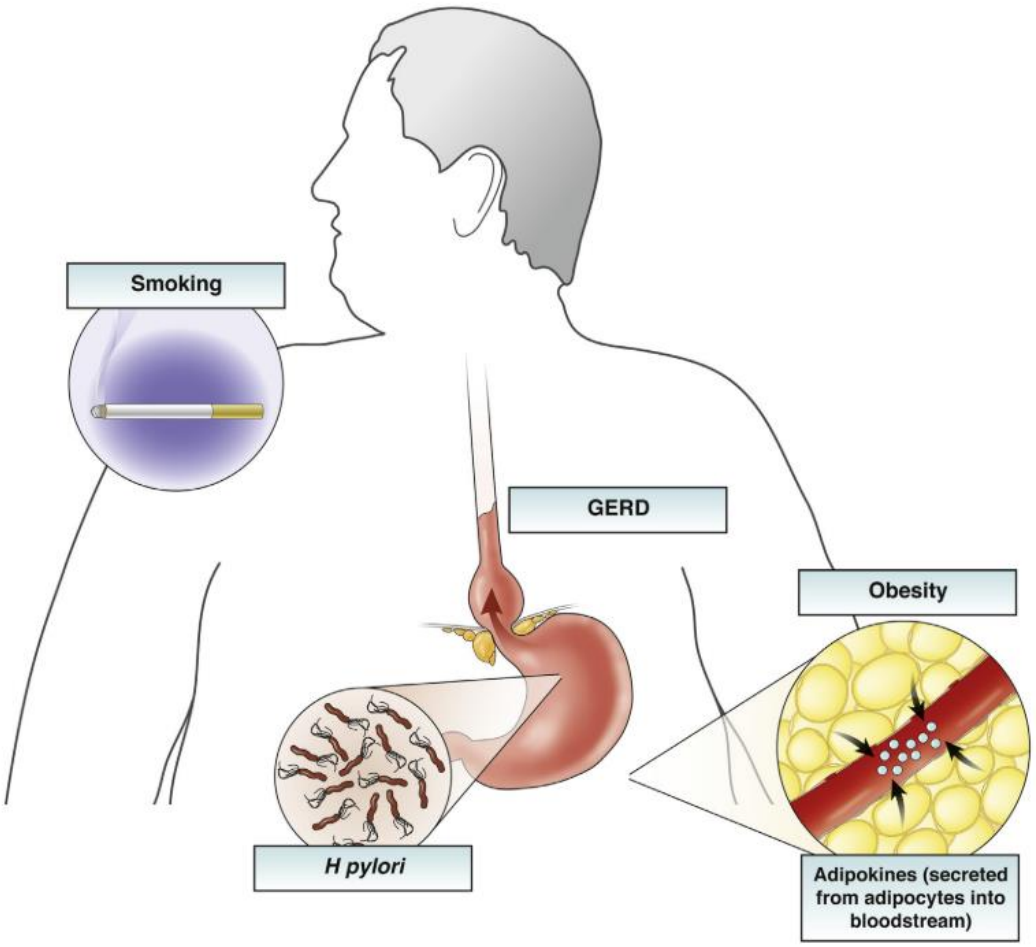


Figure 1.2: Risk Factors for Gastroesophageal Junction Adenocarcinoma.

The primary risk factors for OAC and thus GEJA are male gender, advancing age, Caucasian race, GORD (labelled as GERD in this diagram), smoking and obesity. Obesity acts both mechanically, by increasing rates of GORD, and hormonally through alterations in circulating adipokines. Infection with *H pylori* exerts a protective effect against development of both Barrett's Oesophagus (BO) and OAC/GEJA. Image taken from Rubenstein *et al*, 2015 (Rubenstein & Shaheen, 2015).

Infection with *H pylori* is the strongest known risk factor for diffuse and intestinal type distal gastric adenocarcinoma, yet an inverse relationship has been demonstrated between *H pylori* infection and rates of gastric cardia adenocarcinoma and OAC (Islami & Kamangar, 2008; Kamangar et al., 2006). This is thought to be attributable to one of two main processes: decreased acid secretion, leading to reduced GORD symptoms and oesophagitis; or the proposed ability of refluxed *H pylori* DNA to dampen the IL-12 mediated inflammatory response associated with progression from GORD to Barrett's Oesophagus (BO) (Luther et al., 2011; Moons et al., 2008; Rubenstein & Shaheen, 2015). BO is a condition characterised by metaplastic change of damaged oesophageal squamous cells to intestinal-type mucous secreting columnar cells – a process known as intestinal metaplasia (IM) (Spechler & Souza, 2014) (Figure 1.3).

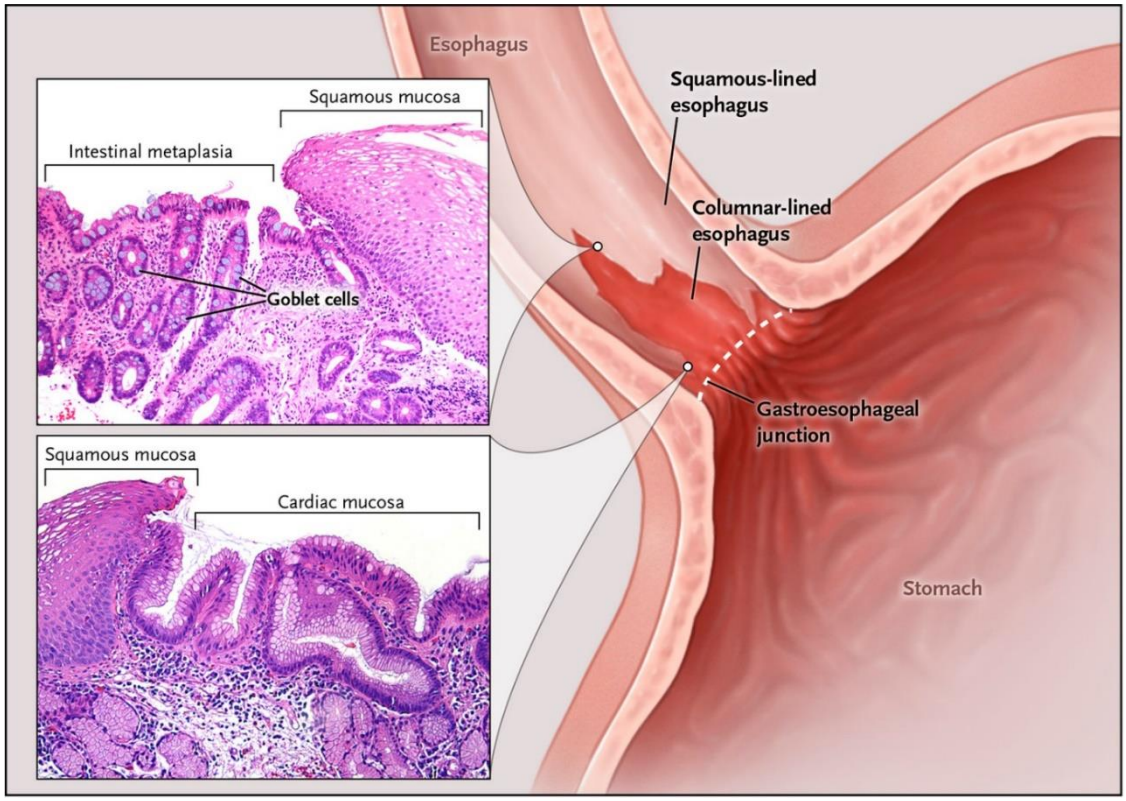


Figure 1.3: Diagnostic Features of Barrett's Oesophagus.

The diagnosis of BO is based on both endoscopic and histologic appearances. At endoscopy, tongues of salmon coloured gastric columnar mucosa extend above the gastroesophageal junction (dashed white line) and into the lower oesophagus. Biopsy specimens are taken to confirm this diagnosis. The biopsy from the area represented by the upper white dot demonstrates the junction between oesophageal stratified squamous epithelium and adjacent IM, confirming a diagnosis of BO. The biopsy from area represented by the lower white dot demonstrates junctional mucosa without IM, which indicates that IM may not be uniformly distributed throughout the entire columnar lined oesophagus. Image taken from Spechler *et al*, 2014 (Spechler & Souza, 2014).

GORD is recognised as the most important risk factor for development of adenocarcinoma of the lower oesophagus and GEJA. GORD and OAC were first linked to each other in 1995 by Chow *et al* (Chow et al., 1995); 4 years later Lagergren *et al* substantiated these findings through a population-based case-control study, which demonstrated an eight fold increased risk of OAC in patients with recurrent GORD symptoms compared to those without symptoms (Lagergren, Bergstrom, Lindgren, & Nyren, 1999). BO may develop in the setting of longstanding GORD with chronic oesophagitis and is the only known precursor to OAC and GEJA (Peters et al., 2019). The presence of IM is associated with an increased risk of dysplasia, with the attendant risk of subsequent malignant change.

Non-dysplastic BO progresses to OAC at an estimated rate of 0.12-0.6% per year (Rubenstein & Shaheen, 2015), thus it is recommended that all patients with known BO undergo regular endoscopic screening. Unfortunately, this is complicated by difficulties in the clinical detection of BO: due to its asymptomatic nature the true prevalence is unknown, with a best estimate of 5.6% produced through interrogation of the US Surveillance Epidemiology and End Results (SEER) data in 2010 (Hayeck, Kong, Spechler, Gazelle, & Hur, 2010). Blanket screening of all patients with GORD symptoms for BO would be inappropriate as most patients with GORD never develop OAC (Rubenstein & Shaheen, 2015; Runge, Abrams, & Shaheen, 2015). Interestingly, the pathogenesis of OAC and intestinal type adenocarcinoma of the gastric cardia both arise in the setting of IM, indicating a potential shared pathway between the two anatomical locations. This pathological link is further supported by genetic analysis studies, which suggest that metaplastic cells in BO originate not from squamous progenitor cells, but rather from gastric cardia progenitor cells that have migrated to the lower oesophagus

(Paulson et al., 2006; Quante et al., 2012). These findings further support treating GEJA as an individual entity.

1.1.4 Classification

The rising incidence of GEJA has led to a need for more effective treatment strategies. A unique range of challenges are encountered in the setting of locally advanced GEJA amenable to curative surgical excision. These tumours straddle the anatomical boundary between the distal oesophagus and the proximal stomach, thus they have generally been treated as either gastric or oesophageal tumours in both clinical and trial settings, rather than as a distinct entity (Lin et al., 2019). There has been much debate as to whether GEJ tumours should be considered to be of gastric or oesophageal origin (Curtis et al., 2014; Ustaalioglu et al., 2017; Zanoni et al., 2018).

First described in 1996, the Siewert scoring system is still used clinically to classify tumours of the GEJ, defining tumours by the location of their epicentre in relation to the gastric cardia: the epicentre of Siewert I tumours are 1-5cm above; Siewert II tumour epicentres lie between 1 cm above and 2 cm below, and the epicentre of Siewert III tumours lies 2-5 cm below the gastric cardia (Siewert & Stein, 1998) (Figure 1.4). Despite its proven clinical utility, this classification system has some limitations. The presence of BO and hiatus hernias can create difficulties for the endoscopist and pathologist in delineating anatomical landmarks, leading to potential for mis-classification (Curtis et al., 2014). Clinical identification of the tumour epicentre has been further complicated since the advent of neoadjuvant treatment (NAT) for locally advanced cancer due to a marked variability in tumour response (Rice, Patil, & Blackstone, 2017). As surgical management of GEJ tumours is heavily influenced by Siewert grouping, erroneous classification can lead to inappropriate surgical approaches, lymph node dissections and

can increase the risk of R1 resections, as characterised by microscopic evidence of residual tumour (Amenabar, Hoppo, & Jobe, 2013; Hermanek & Wittekind, 1994).

Pathological classification and staging of GEJ tumours are based on the TNM classification system. Whilst the 7th edition of the TNM staged all GEJ tumours as oesophageal cancers, the 8th edition was revised to treat Siewert III tumours as gastric cancers similar to the original definition (Rice et al., 2017; Zanoni et al., 2018). Despite this, it is believed that tumours of the GEJ may in fact have a distinct genetic signature, which could facilitate more accurate classification through a 'cell of origin' model in the future (Abdi, Latifi-Navid, Zahri, Yazdanbod, & Pourfarzi, 2019; Lin et al., 2019; Rice et al., 2017). Recent genetic profiling studies conducted by The Cancer Genome Atlas Research Network demonstrated genetic similarities between chromosomal unstable subtype (CIN) gastric cardia adenocarcinomas (Siewert III) and oesophageal adenocarcinoma of the GEJ (Siewert I-II) (Bass, 2014; J. Kim, Bowlby, R., Mungall, A. et al., 2017), which further supports the hypothesis that GEJAs are a separate disease entity.

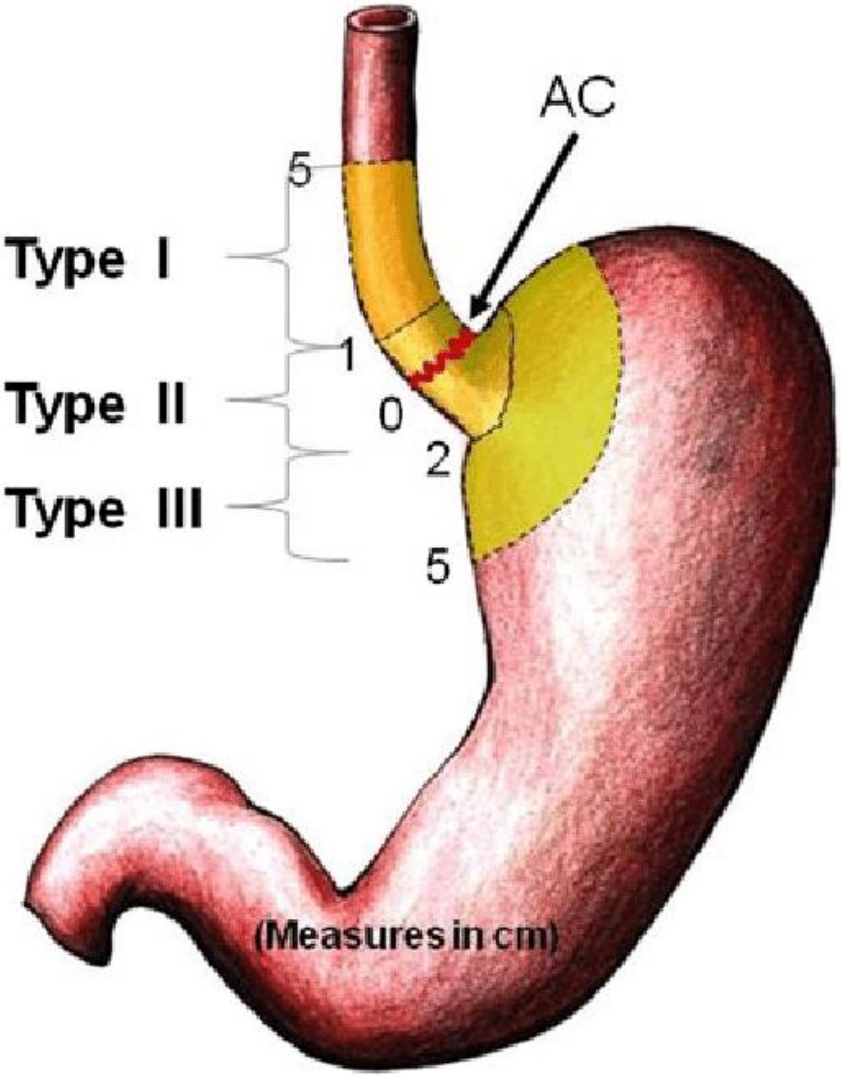


Figure 1.4: Diagram of Siewert Classification.

Siewert's classification divides GEJ tumours based on the location of their epicentre in relation to the gastric cardia (AC = anatomical cardia). Type I tumours have an epicentre 1-5 cm above the AC; type II tumours have an epicentre between 1 cm above and 2 cm below the AC; type III tumour epicentres lie 2-5 cm below the AC. Image taken from Ulla *et al*, 2010 (Ulla et al., 2010).

1.1.5 Treatment Options

Treatment options for GEJA depend on the disease stage at diagnosis. Locally advanced non-metastatic GEJA is treated with a multimodal approach, usually a combination of surgical resection with neoadjuvant, perioperative and/or adjuvant chemotherapy, with or without concomitant radiotherapy (Lin et al., 2019). In early stage (Tis, T1a) disease, minimally invasive approaches using endoscopic mucosal or submucosal resections may be possible. There remains great debate regarding the best surgical approach for frankly invasive junctional tumours however. The risk of an incomplete (R1-2) resection increases if an inappropriate surgical procedure is chosen, thus accurate pre-operative anatomical delineation is imperative. Indeed, a retrospective study of 1062 patients who underwent surgical excision of GEJ tumours showed a 5 year survival rate of 43.2% in those with negative margins (R0), compared to 11% for those with margins positive for residual tumour (R1-2) (Feith, Stein, & Siewert, 2006; Hermanek & Wittekind, 1994). Whilst consensus is usually met on performing an oesophagectomy for Siewert I tumours and total gastrectomy for Siewert III tumours, surgical management remains controversial for Siewert II tumours (Chevallay et al., 2018). Studies have demonstrated no oncological benefit of oesophagectomy versus total gastrectomy in this patient population (Blank et al., 2018; Haverkamp, Ruurda, van Leeuwen, Siersema, & van Hillegersberg, 2014), thus the surgical approach chosen varies depending on patient factors and individual surgeon preferences.

Surgical excision alone has unacceptably high rates of treatment failure; therefore, most patients also receive additional neoadjuvant or perioperative therapy. Several trial studies have examined these treatment options in lower oesophageal and gastric cancers, from which data relating to GEJ tumours has been extrapolated. The neoadjuvant CROSS regimen (Carboplatin and Paclitaxel with concomitant radiotherapy) showed a median

survival of 49.4 months when combined with surgery, compared to a 24 month median survival with surgery alone (van Hagen et al., 2012). Perioperative chemotherapy regimens MAGIC (Epirubicin, Cisplatin, and infused Fluorouracil), ACCORD (Cisplatin and Fluorouracil) and FLOT (5-Fluorouracil, Leucovorin, Oxaliplatin and Docetaxel) all showed an improvement in survival outcomes (Al-Batran et al., 2019; Cunningham et al., 2006; Ychou et al., 2011) (Table 1.1). The FLOT trial additionally demonstrated an advantage of using Docetaxel based therapies over the regimens used in MAGIC and ACCORD trials (ECF/ECX). However, despite the proven survival benefit in these large-scale trials, rates of complete pathologic response (CPR) remain poor: despite its superiority, the FLOT regimen only produced CPR rates of 15-30% (Al-Batran et al., 2016).

Table 1.1 Summary of Different Trial Outcomes.

Study Name	Intervention	Survival Data
CROSS **(van Hagen et al., 2012)**	Preoperative chemoradiotherapy	Median OS 49.4 months *vs*. 24 months for surgery alone
MAGIC **(Cunningham et al., 2006)**	Perioperative chemotherapy	5 year survival 36% *vs*. 23% for surgery alone
ACCORD 07 FNCLCC-FFCD 9703 **(Ychou et al., 2011)**	Perioperative chemotherapy	5 year survival 38% *vs*. 24% for surgery alone
FLOT **(Al-Batran et al., 2019)**	Perioperative chemotherapy	Median OS 50 months *vs*. 35 months in control ECF/ECX group

Abbreviations: OS, overall survival; ECF, Epirubicin, Cisplatin, and Fluorouracil; ECX, Epirubicin, Cisplatin, Capecitabine

Treatment options differ for patients with advanced metastatic disease. Many are surgically inoperable and are instead treated with FOLFOX (5-FU, Leucovorin and Oxaliplatin) or CAPOX (Capecitabine, Leucovorin, Oxaliplatin) along with Trastuzumab in the setting of human epidermal growth factor receptor 2 (HER2) over-expression (Smyth et al., 2017).

Despite the treatment options available, survival rates for GEJA remain extremely poor: 55-60% of patients with early stage disease who undergo primary resection with curative intent will relapse within 5 years, whilst the median OS is 11-12 months for metastatic disease (Joshi, Maron, & Catenacci, 2018). One potential reason for treatment failure in patients with metastatic disease is the differences in genomic profiling between primary and metastatic tumour deposits (Pectasides et al., 2018). The poor response to the conventional therapies outlined in Table 1.1 highlights a need for the development of more effective targeted therapies for both early and advanced stage disease.

Early advances in our understanding of the molecular biology of GEJA have already led to the development of potential new treatment options. Large scale sequencing studies have identified a number of molecularly defined GEJA subsets that may hold therapeutic relevance, including tumours related to Epstein-Barr Virus (EBV); tumours with hyper-mutation, in particular microsatellite instable tumours (MSI); and those with homologous recombination deficiency (HRD) (Y. Y. Janjigian, Sanchez-Vega, et al., 2018). Many GEJAs are of CIN subtype and frequently show amplifications in a range of receptor tyrosine kinases (RTK) – another potential therapeutic target (Bass, 2014; Cristescu et al., 2015; J. Kim, Bowlby, R., Mungall, A. et al., 2017; Secrier et al., 2016).

Immune checkpoint inhibitors (ICIs) are a class of immunotherapeutic agents that hold great promise for treatment of many different malignancies including melanoma, non-small cell lung cancer (NSCLC), urothelial carcinoma and now GEJA (Greally et al., 2019). These drugs act by blocking the immune-evasive binding of tumour surface 'immune checkpoint proteins' with their partner proteins on host immune cells. This enables the body to mount a $CD8^+$ tumour infiltrating lymphocyte (TIL) mediated destruction in response to the tumour. Susceptible tumours will regress or undergo restriction of further cancer growth (Gubin et al., 2014). Two T-cell immunomodulatory

receptors that have been targeted in drug development are Cytotoxic T-Lymphocyte Associated Antigen-4 (CTLA-4) and Programmed Death-1 (PD-1). Pembrolizumab and Nivolumab are two such anti-PD-1 antibodies that have been approved for use in the setting of chemotherapy-refractory GEJA. Indeed, Pembrolizumab has recently been approved in the United States to treat Programmed Death Ligand 1 (PDL-1) positive, MSI-high and chemotherapy refractory GEJA (Greally et al., 2019; Y. Y. Janjigian, Sanchez-Vega, et al., 2018; Le et al., 2015; Muro et al., 2016).

Many trials have reached completion (Table 1.2) and others are ongoing (Table 1.3) which investigate the efficacy of potential therapeutic agents in the management of GEJA. A number of completed trials have demonstrated a survival advantage offered by Pembrolizumab and Nivolumab in the setting of advanced or metastatic GEJA that has failed at least 1 prior line of therapy (Fuchs et al., 2018; Y. Y. Janjigian, Bendell, et al., 2018; Kang et al., 2017; Muro et al., 2016). Ongoing trials are now focusing on combinations of anti-PD-1 antibodies with adjunct therapies: Bang *et al* are investigating the impact of Pembrolizumab taken in conjunction with chemotherapy in the adjuvant/neoadjuvant setting, whilst Janjigian *et al* seek to determine the impact of Pembrolizumab plus chemotherapy and Trastuzumab in a select group of patients with HER2 amplified GEJA, as identified by immunohistochemistry (IHC) and fluorescent *in situ* hybridization (FISH) (Bang et al., 2019; Yelena Yuriy Janjigian et al., 2019).

A separate cohort of clinical trials have investigated alterative targets to ICIs. Fuchs *et al* demonstrated the survival benefit of Ramucirumab as a second line agent for patients with metastatic or unresectable locally advanced GEJA (Fuchs et al., 2014). Ramucirumab is a human IgG1 monoclonal antibody that acts as a vascular endothelial growth factor receptor-2 (VEGFR-2) antagonist. It inhibits angiogenesis through prevention of ligand binding, thus preventing activation of the receptor mediated pathway

(Fuchs et al., 2014). Angiogenesis is one of the hallmarks of cancer, hence its inhibition by Ramucirumab has an anti-tumour effect (Hanahan & Weinberg, 2011). A clinical trial by Park *et al*, currently in the recruitment phase, aims to expand upon the results of Fuchs' trial and examine the impact of Ramucirumab in conjunction with chemotherapy in treatment of metastatic GEJA. Li *et al* also investigated the impact of targeting VEGFR-2 (J. Li et al., 2016). They investigated the use of Apatinib, a TKI which acts by selectively binding to and inhibiting VEGFR-2 and its downstream angiogenic activities, in patients with GEJA who have failed at least 2 prior lines of chemotherapy. A modest survival advantage was shown: those who received Apatinib had a median OS of 6.5 months, compared to 4.7 months in the placebo group (J. Li et al., 2016).

Whilst these trials have shown some therapeutic benefits, the overall survival advantage for the patient remains low. The fact that most trials focus on patients who have disease refractory to first line therapies emphasises the fact that our available treatment options are still hampered by the issue of resistance to anti-cancer drug treatments.

Table 1.2 A Selection of Completed Trials of Targeted Therapies in GEJA.

Study	Phase	Prior Lines	Disease Types	Intervention	Results
KEYNOTE-012 **(Muro et al., 2016)**	Ib	≥2	Resistant or metastatic gastric (n=28) or **GEJ (n=11)** adenocarcinoma	Pembrolizumab	8/39 (22%) ORR
KEYNOTE-059 **(Fuchs et al., 2018)**	II	≥2	Advanced gastric (n=125) or **GEJ (n=133)** adenocarcinoma	Pembrolizumab	Objective response 11.6% (CR 2.3% + PR 9.3%)
CHECKMATE-032 **(Y. Y. Janjigian, Bendell, et al., 2018)**	I/II	≥2	Locally advanced or metastatic gastric (n=59), oesophageal (n=26) or **GEJ (n=75)** adenocarcinoma	Nivolumab alone *vs.* 2 separate dosing regimens of Nivolumab plus Ipilimumab	12 month OS: Nivolumab alone 39%; Nivolumab + Ipilimumab 35% and 24%
ATTRACTION-2 **(Kang et al., 2017)**	III	≥3	Advanced gastric (n=308) or **GEJ (n=27)** adenocarcinoma (Unknown n=32)	Nivolumab *vs.* Placebo	Median OS: 5.23 months with Nivolumab; 4.14 months with Placebo
(J. Li et al., 2016)	III	≥3	Advanced gastric (n=112) or **GEJ (n=36)** adenocarcinoma (Unknown n=13)	Apatinib *vs.* Placebo	Median OS: 6.5 months with Apatinib; 4.7 months with Placebo
REGARD **(Fuchs et al., 2014)**	III	≥2	Metastatic or unresectable locally recurrent gastric (n=265) or GEJ (n=90) adenocarcinoma	Ramucirumab *vs.* Placebo	Median OS: 5.2 months with Ramucirumab; 3.8 months with Placebo

Abbreviations: ORR, overall response rate; CR, complete response; PR, partial response

Table 1.3 A Selection of Ongoing Trials of Targeted Therapies in GEJA.

Study	Phase	Line	Disease Types	Intervention	Primary Endpoints
KEYNOTE-585 **(Bang et al., 2019)** NCT03221426	III	1	Localized gastric or **GEJ** adenocarcinoma	Pembrolizumab + Chemotherapy (FP or XP) *vs.* Placebo + Chemotherapy (FP or XP)	OS, EFS and CPR
CHECKMATE-577 **(Kelly et al., 2017)** NCT02743494	III	2	Stage II/III oesophageal or **GEJ** adenocarcinoma	Nivolumab *vs.* Placebo	OS and DFS
KEYNOTE-811 **(Yelena Yuriy Janjigian et al., 2019)** NCT02954536	II	1	HER-2 positive metastatic **GEJ** adenocarcinoma	Pembrolizumab with Chemotherapy/Trastuzumab	6 months PFS
Park *et al* NCT03141034	II	2	Metastatic gastric and **GEJ** adenocarcinoma	Irinotecan plus Ramucirumab	PFS

Abbreviations: FP, 5-fluorouracil; XP, Cisplatin plus Capecitabine; EFS, event free survival; DFS, disease free survival; PFS, progression free survival

1.1.6 Treatment Resistance

Despite modest advances in drug treatments, the issue of resistance to anti-cancer therapies persists as an obstacle to optimal clinical management and prognostication in GEJA, and indeed to malignancies of all sites. The mechanisms leading to drug resistance are complex and multifactorial. The pharmacological impact of a particular therapeutic agent depends on both intrinsic and acquired characteristics of the tumour cells (Vasan, Baselga, & Hyman, 2019). The key determinants of drug resistance in tumours are depicted in Figure 1.5.

The mechanisms of resistance vary with each determinant. For example, the interplay between the tumour and its microenvironment – that being the surrounding immune cells, stroma and vasculature – may mediate resistance through both obstruction of drug absorption by the tumour cells and stimulation of paracrine growth factors that promote tumour cell growth (Prieto-Vila, Takahashi, Usuba, Kohama, & Ochiya, 2017; Vasan et al., 2019). Physical barriers include 'sanctuary sites' such as the CNS, which are anatomical sites within which systemic therapies do not reach therapeutic concentrations (Toyokawa, Seto, Takenoyama, & Ichinose, 2015). In addition to this, many oncogenes and tumour suppressor genes have yet to be targeted by anti-cancer therapies, including TP53 and MYC: the presence of an 'undruggable genome' further contributes to the heterogeneity of tumour cells and hence drug resistance (Vasan et al., 2019).

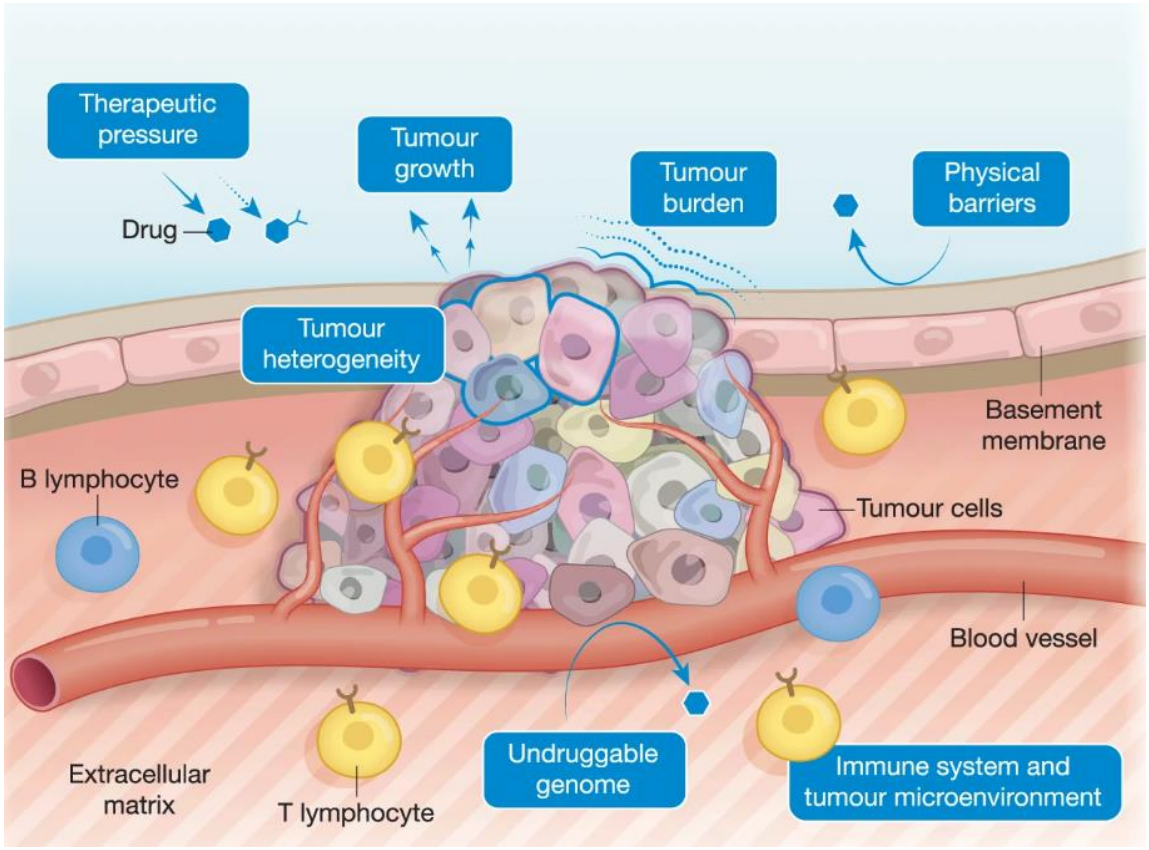


Figure 1.5: Biological Determinants of Drug Resistance. Image taken from Vasan *et al*, 2019.

Most tumours are comprised of a phenotypically diverse population of cancer cells (Prasetyanti & Medema, 2017; Shibue & Weinberg, 2017), driven by a complex array of genetic and epigenetic alterations that disrupt normal cell cycle processes. This diversity is known as intra-tumour heterogeneity and is thought to play a crucial role in the development of treatment resistance (Prasetyanti & Medema, 2017). OAC has a high mutational burden compared to other malignancies, with a median mutation frequency of 9.9 per million base pairs (mbp) (Dulak et al., 2013). Oesophageal cancer is ranked 6th out of 30 tumour types in terms of prevalence of somatic mutations, with a recent study identifying 77 driver genes and 21 non-coding driver elements within this tumour type alone (Alexandrov et al., 2013; Frankell et al., 2019). A number of driver events were found to occur exclusively in, or overlap between, dysregulated OAC pathways, indicating a significant relationship between their presence and the progression to malignancy (Frankell et al., 2019). However, the rate of acquisition of genomic alterations is highly variable, ranging from slow age-related mutations to dramatic large chromosomal alterations that may often represent a 'point of no return' in the development of drug resistance (Vasan et al., 2019). This further complicates the development of targeted therapies and highlights the need for early therapeutic intervention.

Personalised medical therapies often fail because a single biopsy may sample only one sub-population of tumour cells, thus underestimating the heterogeneity present within a tumour (Gerlinger et al., 2012). The presence of small sub-populations of cancer stem cells (CSC) within a tumour is of crucial clinical importance as CSCs are known to contribute to both resistance to anti-cancer therapies and to tumour metastasis in many solid organ malignancies, yet there remains a paucity of published literature pertaining to this subject in relation to GEJA (S. Li & Li, 2014; Nunes et al., 2018).

1.2 Epithelial Mesenchymal Transition and Cancer Stem Cells

1.2.1 Cancer Stem Cells

Cancer stem cells are a small but crucially important sub-population of tumour cells which drive tumorigenesis, metastasis and treatment resistance (Prasetyanti & Medema, 2017). They are undifferentiated and capable of limitless self-renewal, with potential for subsequent differentiation into various non-CSC cell types which have no capacity for self-renewal or migration and instead form the bulk of the tumour (Reya, Morrison, Clarke, & Weissman, 2001). CSCs hold a Darwinian survival advantage over other subclones within a single tumour due to their endogenous resistance mechanisms against chemo-radiotherapy (Eun, Ham, & Kim, 2017; Prieto-Vila et al., 2017). Additionally, their ability to generate phenotypically varied clonal populations of both CSC and non-CSC cells within a single tumour increases the likelihood of at least one group of tumour cells surviving the assault of anti-cancer treatments (Brooks, Burness, & Wicha, 2015; Eun et al., 2017). It has been proposed that the limited efficacy of conventional anti-cancer therapies is attributable to the fact that these treatments target the bulk population of non-CSCs within a tumour, allowing minor populations of CSCs to persist and propagate, leading to a clinical relapse (Reya et al., 2001; Shibue & Weinberg, 2017). CSCs are therefore one of the most clinically important contributors to intra-tumour heterogeneity and thus resistance to anti-cancer treatments.

In the same fashion as normal adult stem cells, CSCs reside in niches: these are a specialised component of the tumour microenvironment which act to regulate the fate of stem cells via specific signals and cellular interactions (Cabrera, Hollingsworth, & Hurt,

2015). These interactions result in phenotypic plasticity of CSCs, as they can interconvert between differentiated and stem-like states (Quail, Taylor, & Postovit, 2012). This plasticity greatly contributes to intra-tumour heterogeneity and treatment resistance, as cells can adopt a quiescent non-CSC state in response to signals from its niche (Batlle & Clevers, 2017) (Figure 1.6).

Two models exist with which to understand the mechanisms of tumour progression and heterogeneity: the hierarchical model and the stochastic model (Plaks, Kong, & Werb, 2015). The hierarchical model places greater weight on the role of CSCs, assuming that they represent a distinct sub-population within a tumour which drives carcinogenesis by giving rise to different cancer cells, which in turn assemble in a hierarchical fashion as they generate their own CSCs (Visvader & Lindeman, 2008). By contrast, the stochastic model considers every cell within a tumour to be equally likely to act as a cell of origin for tumorigenesis (Plaks et al., 2015). This is based on the premise that in cancer, normal quiescent adult cells can acquire a critical volume of mutations in cell cycle genes which lead to subsequent clonal expansions (Quail et al., 2012). Both models have their limitations: the hierarchical model does not consider the ability of CSCs to transition into non-CSCs, whilst the stochastic model does not take into consideration the role of microenvironmental cues in generating tumour-initiating capacity within cancer cells (Batlle & Clevers, 2017). Despite their differences however, the concept of CSC plasticity unites the two models: CSCs organised within a hierarchy can transition between states, whilst oncogenic mutations can facilitate the generation of CSCs.

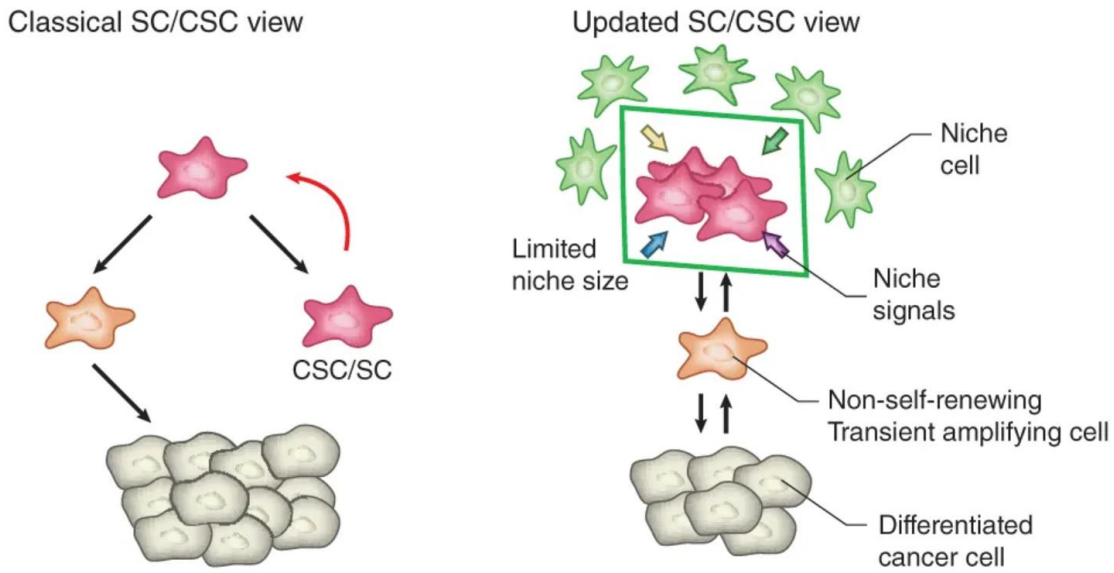


Figure 1.6. Updates in Understanding of Stem Cells and Cancer Stem Cells.

The earlier concept of CSCs revolved around the understanding that CSCs (pictured in pink) are rare cells within a tumour which undergo asymmetric division, giving rise to one CSC and one transient amplifying cell (pictured in orange). The latter is not capable of self-renewal and thus eventually undergoes terminal differentiation. By contrast, we now understand that following division, the outcome of the CSC daughters is determined by signals from the niche. Only cells that remain within the niche are stem cells; those that exit the niche undergo differentiation. Adapted from Batlle *et al*, 2017 (Batlle & Clevers, 2017).

CSCs are regulated by a number of signalling pathways associated with stemness, which include Notch, Hedgehog, Wnt/β-Catenin, JAK/STAT, and NF-κB (K. Chen, Huang, & Chen, 2013). These pathways play a role in the maintenance of stem cell properties and/or regulation of their differentiation through alteration of messenger RNA (mRNA) expression via a specific subset of transcription factors (TF) (Eun et al., 2017). In 2006 Yamanaka identified the first four TFs that were involved in inducing pluripotent stem cells from mouse embryonic and adult fibroblast cultures: OCT3/4, SOX2, c-MYC and Klf-4 (Takahashi & Yamanaka, 2006). Other TFs that play a role in the generation of induced pluripotent stem cells (iPSC) include NANOG and SALL4, which are encoded by their correspondingly named genes (Rodriguez et al., 2014; K. H. Song et al., 2017; Zeineddine, Hammoud, Mortada, & Boeuf, 2014). These TFs are thought to act in concert with each other and other complex molecular processes to establish stem cell traits in a range of cell types, including CSC traits in neoplastic cells.

MicroRNAs (miRNA) also play a role in the regulation of CSC traits. These are a class of small non-coding RNAs which are involved in regulating gene expression through either degradation of their target mRNA or through inhibition of their translation, with an overall effect of altered protein expression within cells. miRNAs are key in regulating a range of essential biological processes including proliferation, differentiation, survival and apoptosis in many different cell types (Hezova et al., 2016). They have been shown to be aberrantly expressed in many different human cancers and to play a role in the regulation of CSC characteristics (Khan et al., 2019). In their latter role, they act by targeting many of the same signalling pathways detailed above, which are involved in regulation of mRNAs associated with stemness, including Wnt/β-Catenin, JAK/STAT, and NF-κB (Khan et al., 2019). Certain miRNAs may also play a role in tumorigenesis

by regulating the cell cycle components of CSCs to inhibit apoptosis and promote cellular proliferation (Mens & Ghanbari, 2018).

miRNAs involved in CSC regulation include the miR-17-92 family, which regulates the MYC oncogene to protect CSCs against apoptosis; the let-7 family, whose decreased expression is associated with metastasis and chemoresistance; and a wide range of others including miR-21, miR-16 and miR-200 (Y. Li, Choi, Casey, Dill, & Felsher, 2014; Mens & Ghanbari, 2018). Whilst many of these miRNAs have been shown to regulate organ-specific CSCs, there is considerable overlap between the expression of miRNAs in different solid organ malignancies (Chakraborty, Chin, & Das, 2016). For example, miR-17 is downregulated in OAC and renal cell carcinoma CSCs, yet miR-17 over-expression has been demonstrated in colorectal CSCs (Lichner et al., 2015; Xi et al., 2016). This highlights the molecular complexities of CSC regulation, and thus the difficulties in identifying a suitable targeted therapeutic agent for individual malignancies.

CSCs present within a tumour are reported to be selected by a small subset of cell surface proteins which include CD133, CD44, CD24, CD34 and ALDH1A1 (T. Chen, You, Jiang, & Wang, 2017; Hermansen, Christensen, Jensen, & Kristensen, 2011; Yang, Wang, Wang, Chen, & Bai, 2018). There are no unified CSC-specific markers in all tissues and organs: their expression can vary between cancer type. For example, CD133 is expressed in liver, brain, colorectal and pancreatic CSCs, yet it is not an appropriate biomarker for melanoma and head and neck cancer CSCs (K. Chen et al., 2013; Madjd et al., 2016).

In addition to the expression of CSC cell surface markers, these cells also classically express cellular proteins associated with a mesenchymal phenotype (T. Chen et al., 2017). The expression of mesenchymal cell markers in an epithelial malignancy points towards

a shift from an epithelial to a mesenchymal phenotype – a phenomenon known as epithelial mesenchymal transition (EMT) (Greenburg & Hay, 1982). The overlapping molecular features between CSCs and EMT suggest that EMT plays a role in the promotion and regulation of CSCs. A greater understanding of the biological link between these processes would potentially facilitate the development of targeted therapies against these small sub-populations of cells, thus reducing the rates of drug resistance and metastasis and improving survival outcomes for patients.

1.2.2 Epithelial Mesenchymal Transition

Epithelial mesenchymal transition, first described in 1982 by Greenberg and Hay (Greenburg & Hay, 1982), is a process of lineage transition whereby epithelial cells lose their adhesive properties and acquire a mesenchymal cell phenotype, with changes in cell morphology and expression of surface markers (Kalluri & Weinberg, 2009). This programme involves epigenetic cellular modifications, resulting in heritable phenotypic changes without the acquisition of new genetic alterations. Epithelial and mesenchymal cells differ in their histological appearance and physiology. Epithelial cells are characterized by apical-basal polarity and tight intracellular junctions, whilst mesenchymal cells are fibroblast-like with discohesive cell-cell interactions. This change in phenotype facilitates invasion and tumour cell migration to remote sites, as seen in metastasis (T. Chen et al., 2017; Lamouille, Xu, & Derynck, 2014). The reverse of this process – mesenchymal-epithelial transition (MET) – confirms the plasticity of these cells and plays a role in the colonization of metastatic deposits at remote sites: a common CSC fate (Kalluri & Weinberg, 2009). These distant metastatic deposits no longer express the

mesenchymal markers associated with metastasizing carcinoma cells, but rather reassume an epithelial phenotype (Kalluri & Weinberg, 2009).

The development of EMT is complex and involves an array of different molecular processes including activation of TFs, changes in expression of specific miRNAs, epigenetic changes and alterations of cytoskeletal and cell surface proteins (Diepenbruck & Christofori, 2016; Kalluri & Weinberg, 2009). EMT is involved in both pathological and physiological processes and is divided into three subgroups relating to embryogenesis, tissue regeneration and cancer progression (T. Chen et al., 2017; Kalluri & Weinberg, 2009). The EMT programme was first studied in the context of embryogenesis, where it was shown that the mesodermal and endodermal cell layers arose from the transition of epiblasts (primitive epithelial cells) present within the endodermal layer to cells with a mesenchymal phenotype. Similarly, EMT facilitated the migration of neural crest cells in the dorsal neural tube, with differentiation at remote sites into a range of cell types including melanocytes and glial cells (T. Chen et al., 2017; Reya et al., 2001; Shibue & Weinberg, 2017).

Type 3 EMT – that which relates to cancer progression – occurs in neoplastic cells that have already undergone genetic and epigenetic changes in tumour suppressor genes and oncogenes (Kalluri & Weinberg, 2009). Neoplastic cells that undergo EMT are typically found at the invasive front of primary tumours and as they progress thorough the transitional process they acquire the ability to invade and metastasize – two hallmarks of cancer (Fouad & Aanei, 2017; Kalluri & Weinberg, 2009). The progression through EMT is not uniform however: different sub-populations of cells within a tumour may undergo only partial EMT, thus further contributing to intra-tumour heterogeneity (Prasetyanti & Medema, 2017).

Type 3 EMT is tightly regulated by a wide spectrum of complex cellular signalling pathways (Figure 1.7). In many carcinomas the tumour microenvironment – comprised of a large cohort of stromal cells including cancer associated fibroblasts (CAF), T-lymphocytes, macrophages and myeloid derived suppressor cells – releases a range of cytokines, chemokines and growth factors which act in a paracrine fashion to induce EMT; these include hepatocyte growth factor (HGF), epidermal growth factor (EGF), platelet derived growth factor (PDGF), Interleukin 6 (IL-6) and transforming growth factor beta (TGF-β) (Table 1.4) (Dongre & Weinberg, 2019; Kalluri & Weinberg, 2009). These mediators are involved in the activation of a group of EMT inducers, including the E-box binding protein family 'Zeb'; zinc-finger protein family 'Snails', helix-loop-helix protein family 'Twists' and forkhead box protein family 'FOXCs' (Galvan et al., 2015; Kalluri & Weinberg, 2009; Medici, Hay, & Olsen, 2008; Wei et al., 2015; J. M. Yu et al., 2015). Once activated, these TFs orchestrate the EMT programme through a series of intracellular signalling pathways which include MAPK, ERK, PI3K, Wnt/β-catenin, Ras and Smads; these interact and work in tandem with various cell surface proteins including β4 and α5 integrins (Kalluri & Weinberg, 2009; Tse & Kalluri, 2007). These signalling pathways frequently overlap and are regulated by multiple intricate cellular interactions involving miRNAs, epigenetic modulators and exogenous inducers (T. Chen et al., 2017).

Table 1.4 Mechanisms of EMT Activation by the Tumour Microenvironment.

Cellular Components of the Tumour Microenvironment (Dongre & Weinberg, 2019)	**Mechanisms of EMT Activation** (Dongre & Weinberg, 2019; Kalluri & Weinberg, 2009)
Cancer Associated Fibroblasts	- Secretion of cytokines and growth factors including TGF-β, Il-6, EGF, VEGF and HGF - Induction of vimentin expression and inhibition of E-cadherin expression - Methylation of EMT regulatory genes
Tumour Associated Macrophages	- Secretion of TGF-β - Secretion of TNF, which acts in synergy with TGF-β
Myeloid Derived Suppressor Cells	- Secretion of TGFβ and activation of COX2, EGF and HGF pathways
T-Lymphocytes	- $CD4^+$ and $CD8^+$ cytotoxic T cells activate EMT through poorly understood mechanisms

Abbreviations: TNF, tumour necrosis factor; COX2, cyclooxygenase-2

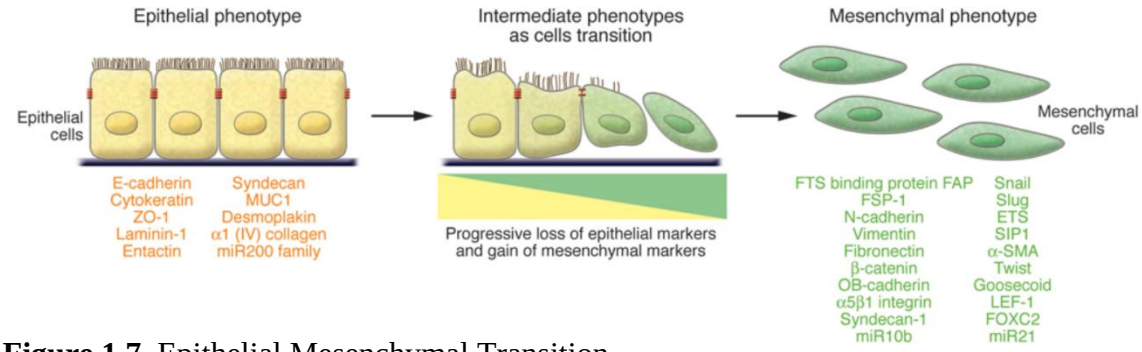


Figure 1.7. Epithelial Mesenchymal Transition.

EMT involves the transition of cells from an epithelial to a mesenchymal phenotype. The cell markers commonly used in research to differentiate between the two states are listed. During the intermediate transition phase, both epithelial and mesenchymal markers may be detected, reflecting the fact that some cells go through 'partial' rather than 'complete' EMT. *ZO-1, zona occludens 1; MUC1, mucin 1 cell surface associated; miR200, microRNA 200; SIP1, survival of motor neuron protein interacting protein 1; FOXC2, forkhead box C2.* Image taken from Kalluri *et al* 2009 (Kalluri & Weinberg, 2009).

The regulation of β-catenin in EMT initiation demonstrates this signalling complexity well. β-catenin is a protein involved in cell adhesion and gene transcription (Nelson & Nusse, 2004). It forms part of the cadherin complex, which maintains cellular integrity, and plays a central role in canonical (β-catenin dependant) Wnt signalling pathways, which lead to the transcription of genes that favour an EMT programme (Nelson & Nusse, 2004; Niehrs, 2012). The process of disruption of intercellular junctions is a key initiating event in EMT that involves the breakdown of specialised cell surface protein complexes which act to maintain epithelial cell integrity (Lamouille et al., 2014). Loss of E-cadherin through this process prevents its interaction with β-catenin, which results in accumulation of β-catenin and thus activation of the Wnt/β-catenin signalling pathway (Nelson & Nusse, 2004). Through this pathway β-catenin translocates to the nucleus, where it activates the transcription of target genes involved in EMT under the control of T cell factor (TCF) (Niehrs, 2012). However, loss of E-cadherin is not the only factor that leads to accumulation of β-catenin: activation of the Wnt signalling pathway also results in inhibition of GSK3β activity – an enzyme known to inhibit tumour migration and invasion. This inhibition prevents phosphorylation of β-catenin, thus further facilitating its regulation of gene expression involved in EMT initiation (Basu, Cheriyamundath, & Ben-Ze'ev, 2018; Kao et al., 2014; Lamouille et al., 2014).

The intricacy of EMT initiation is similarly well demonstrated by examining the role of TGF-β in this process. TGF-β is a pleiotropic molecule involved in a range of pathological and physiological processes, including two which are mutually exclusive: initiation of EMT and apoptosis (J. Song, 2007; Spender et al., 2009). In the early stages of tumorigenesis TGF-β exerts its tumour suppressor function by inducing apoptosis of pre-malignant cells, whilst in the later stages its role shifts to that of a tumour promotor as it instead becomes involved in the promotion of EMT (Hao, Baker, & Ten Dijke,

2019). The direction of action of TGF-β is regulated through activation of different signalling pathways. Studies have suggested that the TGF-β/Smad signalling pathway plays an important role in inducing apoptosis via upregulation of pro-apoptotic factors (Spender et al., 2009). By comparison, regulation of the role of TGF-β in EMT initiation is more complex: in this process a range of signalling pathways which lead to EMT are activated by TGF-β, including both TGF-β/Smad and MAPK/ERK (Hao et al., 2019; Zavadil et al., 2001). In addition to this, TGF-β also exerts its influence on EMT by affecting the activities of other EMT associated pathways such as Notch and Wnt (Gonzalez & Medici, 2014). Taken in combination, these studies clearly demonstrate the subtlety and complexity of cellular processes involved in the initiation and completion of EMT.

The role of hypoxia, which is often present in poorly vascularized areas of tumour, in the induction of EMT is also of great importance. It has been shown to induce EMT through a range of mechanisms, including upregulation of hypoxia-inducible factor-1α (HIF1α), HGF, Snail1 and Twist1; activation of Notch and NF-κB pathways; and induction of DNA hypomethylation (Gort, Groot, van der Wall, van Diest, & Vooijs, 2008; Polyak & Weinberg, 2009). Interestingly, low O_2 levels have been shown to induce EMT through inhibition of GSK3β activity, leading to accumulation of β-catenin, upregulation of Wnt/β-catenin signalling and ultimately induction of the TF Snail1 (Cannito et al., 2008). These findings not only implicate hypoxia in the induction of EMT, they also highlight the biological diversity of EMT activation and thus the need to expand our knowledge of the molecular mechanisms underlying this process.

Our current understanding of EMT can be further expanded through experimental interrogation of the cross talk between complex signalling pathways, as directed by existing knowledge of molecular markers associated with each stage of phenotypic

transformation. Cells in the process of transition from an epithelial to a mesenchymal phenotype will classically express increased levels of mesenchymal markers and decreased levels of epithelial markers. A pure epithelial phenotype will commonly express cell surface markers including E-cadherin, cytokeratin, desmoplakin and others, whilst markers expressed by cells with a pure mesenchymal phenotype include vimentin and β-catenin. Both epithelial and mesenchymal phenotypes are also associated with expression of different miRNAs. For example, the expression of miR-21, which is involved in activation of TGF-β induced EMT, is associated with the mesenchymal phenotype (Kalluri & Weinberg, 2009). By comparison, miR-200 expression is associated with the epithelial phenotype, playing an important role in inhibition of E-cadherin repressors Zeb1 and Zeb2 (Dongre & Weinberg, 2019; Kalluri & Weinberg, 2009) (Figure 1.8). Accurate identification of cells in various stages of EMT is paramount due to its involvement in such clinically applicable processes as drug resistance and tumour metastasis: traits that are shared with CSCs.

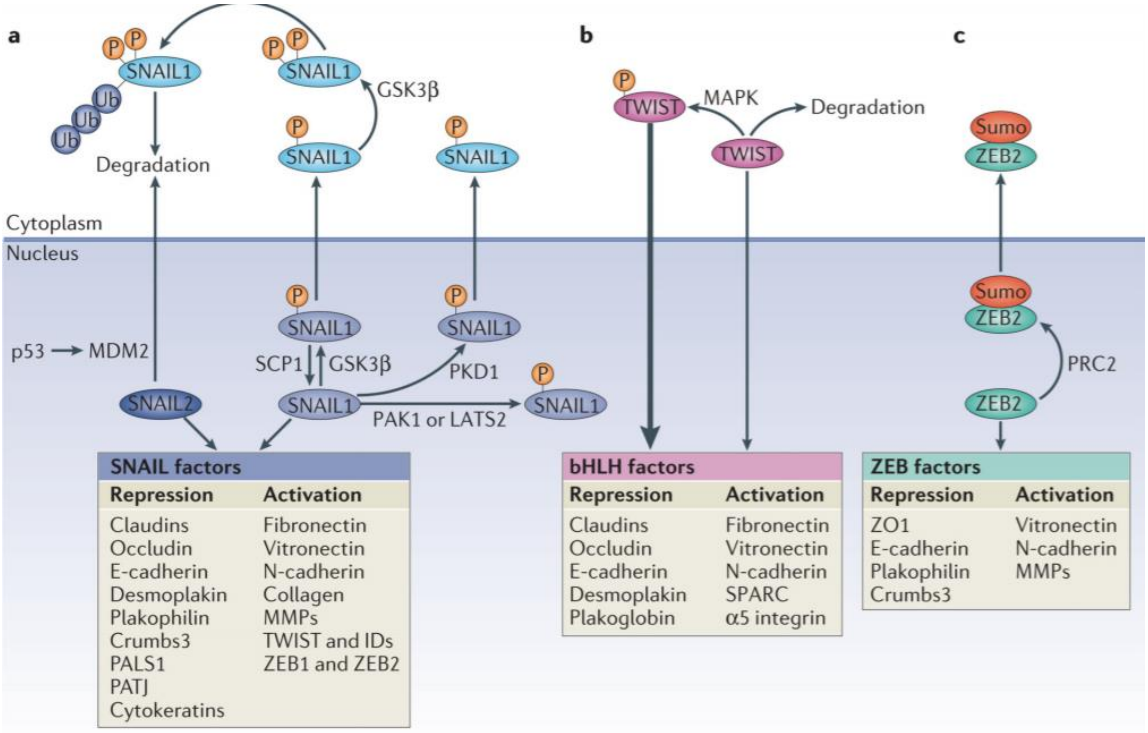


Figure 1.8. The Roles and Regulation of Major EMT Transcription Factors.

EMT is driven by several TFs including Snail1, Twist and Zeb2, which each exert their effects through repression of epithelial genes and activation of genes associated with a mesenchymal phenotype. Their action is further modified at the post-translational level through several interactions including (a) phosphorylation of Snail1 by enzymes GSK3β and p21 Activated Kinase 1 (PAK1); (b) phosphorylation of Twist by MAPK and (c) sumoylation of Zeb2 by Polycomb Repressive Complex 2 (PRC2). *E-cadherin, Epithelial Cadherin; ID, Inhibitor of Differentiation; MMP, Matrix Metalloproteinase; N-cadherin, Neural Cadherin; PALS1, Protein Associated with Lin-7 1; PATJ, PALS1-Associated Tight-Junction Protein; SPARC, Secreted Protein Acidic and Rich in Cys; ZO1, Zonula Occludens 1.* Adapted from Lamouille *et al*, 2014.

1.2.3 EMT as a Regulator of Cancer Stem Cells

The relationship between EMT and the acquisition of CSC traits has been extensively investigated. Experimental studies have demonstrated that the EMT process is associated with the acquisition of CSC properties on neoplastic cells across a wide range of human carcinomas (T. Chen et al., 2017). CSCs were first identified in the 1990s when $CD34^+$ $CD38^-$ leukemic cells were shown to have bone marrow hematopoietic stem cell characteristics (Bonnet & Dick, 1997; Lapidot et al., 1994). In 2003 Al-Hajj *et al* identified CSCs in solid tumours by demonstrating tumorigenic (stem) cells with cell surface marker profile $CD44^+CD24^{-/low}$ in breast cancer (Al-Hajj, Wicha, Benito-Hernandez, Morrison, & Clarke, 2003), followed shortly after by identification of CSC markers for other solid organ malignancies including prostate, colon, liver and lung (Eun et al., 2017; Medema, 2013).

In 2008 Mani *et al* demonstrated a direct link between EMT and CSCs by inducing EMT in human mammary epithelial cells (HMLE) and identifying both the acquisition of mesenchymal traits and the expression of stem cell markers (Mani et al., 2008). The group induced EMT in HMLEs via ectopic expression of Snail or Twist, or exposure to TGF-β stimulation, leading to acquisition of traits seen in neoplastic mammary stem cells: a $CD44^{high}/CD24^{low}$ phenotype with the ability to form a mammosphere. These cells additionally had the ability to produce a $CD44^{low}/CD24^{high}$ population of non-mammosphere forming cells, thus indicating that EMT is associated with a gain of CSC traits. Morel *et al* also demonstrated the acquisition of CSC traits in HMLEs following activation of the Ras-MAPK pathway (Morel et al., 2008). Whilst Mani and Morel have both shown that induction of EMT-TFs in HMLEs *in vitro* leads to the development of stem cell characteristics, subsequent research into this area has demonstrated that a

variety of these TFs have much broader range of influences over the process of tumorigenesis including cellular plasticity, oncogenic transformation and mobility. The versatility of these functions, which are further complicated by the cell-type specific epigenetic landscape, pose a barrier to our understanding of the molecular mechanisms underlying tumour development (Goossens, Vandamme, Van Vlierberghe, & Berx, 2017).

Additional studies have demonstrated this link between EMT and CSC, supporting their causality. Both initiation of EMT and generation of CSCs are associated with TGF-β signalling. A study conducted by Shipitsin *et al* in 2007 showed that $CD44^+$ breast cancer stem cells had higher levels of TGF-β compared with non-CSC breast cancer cells, and inhibition of this signalling pathway in the CSCs re-established an epithelial phenotype (Shipitsin et al., 2007). Other studies have shown high levels of Wnt signalling in colorectal CSCs with high expression of β-catenin at the invasive front, and demonstrated the role of Notch signalling in development of CSCs (Brabletz et al., 1998; Vermeulen et al., 2010). Furthermore, a recent study by Pistore *et al* showed that CAF-conditioned media can induce both mesenchymal and CSC-like properties in prostate cancer cells through concurrent DNA hypo- and hyper-methylation (Pistore et al., 2017).

However, despite strong evidence to support a link between EMT and the acquisition of CSC traits, studies exist which contradict this association (Celia-Terrassa et al., 2012; T. Chen et al., 2017; Liao & Yang, 2017; Xie et al., 2014). Two separate studies examined the role of Twist1, which is involved in both EMT and the promotion of CSC properties; both demonstrated that the acquisition of CSC traits does not always occur in tandem with EMT (Beck et al., 2015; Schmidt et al., 2015). This implies that whilst EMT and CSC are closely linked, EMT is not necessarily required for development of cancer stemness; their regulation may in fact be via independent functions of the same EMT-

TFs (Liao & Yang, 2017). Other groups have suggested an association between MET and stemness. For example, a study which investigated the reprogramming of mouse fibroblasts to iPSC showed that MET played a key role in the acquisition of stemness traits through suppression of Snail via OCT-4/SOX2, downregulation of TGF-β via c-Myc and induction of epithelial genes such as E-cadherin through Klf-4 (R. Li et al., 2010). In 2012 Ocana *et al* demonstrated that repression of homeobox TF Prrx1 – an EMT inducer – in human breast cancer cells caused them to simultaneously revert to an epithelial phenotype (MET) and acquire CSC traits (Ocana et al., 2012). Given the clinical relevance of the shared traits between these two processes, specifically metastasis and drug resistance, greater investigation is warranted in order to gain a better understanding of their molecular associations.

1.2.4 The Role of EMT and CSC in Tumour Metastasis

Metastasis refers to the dissemination of tumour cells to sites remote from the primary tumour (A. F. Chambers, Groom, & MacDonald, 2002). The metastatic cascade commences with invasion of primary tumour cells at the invasive tumour edge into the surrounding tissue, followed by intravasation into the blood or lymphatic system as either a single cell or a group of cells. The tumour cells are then disseminated throughout the systemic circulation as circulating tumour cells (CTC), with subsequent extravasation and metastatic outgrowth at a secondary site (Figure 1.9) (Diepenbruck & Christofori, 2016). The formation of metastases requires all stages of this metastatic cascade, yet the process is complex: it has been estimated that only 0.01% of tumour cells which enter the systemic circulation are capable of forming a secondary tumour (A. F. Chambers et al.,

2002). Both EMT and MET are crucial components involved in the establishment of metastasis (Diepenbruck & Christofori, 2016).

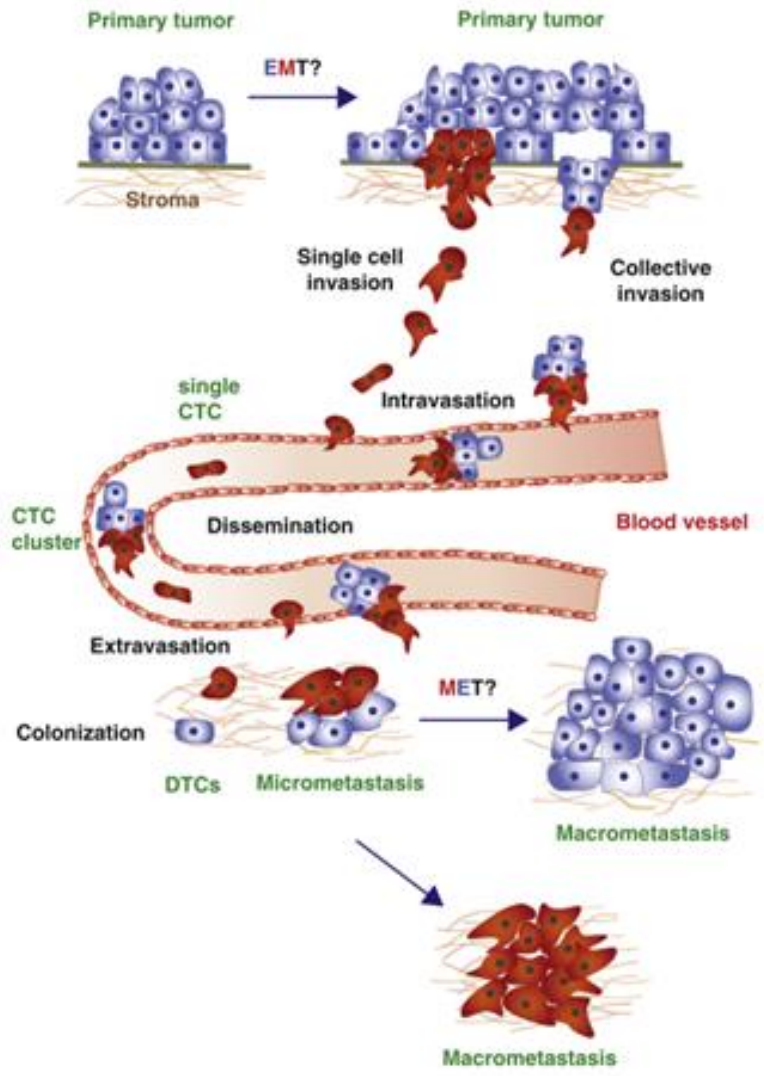


Figure 1.9: The Role of EMT in the Invasion Metastasis Cascade.

Whilst not yet fully understood, this figure depicts the proposed role of EMT in metastasis. Metastasis starts with loss of adhesive properties of tumour cells, leading to migration out from the primary tumour, invasion of the surrounding tissue and subsequent intravasation into the blood/lymphatic stream. Single CTCs mainly express mesenchymal characteristics (depicted in red), yet clusters of CTCs often consist of cells expressing both epithelial (depicted in blue) and mesenchymal traits. Once they reach their secondary site, cancer cells must adhere to the target site endothelium, extravasate and migrate into the parenchyma. At this point the tumour cells either persist in a prolonged dormant state as disseminated tumour cells (DTC), or settle as multicellular micrometastasis which grow to form macrometastasis. Image adapted from Diepenbruk *et al*, 2016 (Diepenbruck & Christofori, 2016).

Many studies have sought to determine the exact role of EMT in metastasis. Experimental models involving manipulation of TGF-β expression and interference with E-cadherin function have shown that induction of EMT promotes tumour cell invasion and migration (Diepenbruck & Christofori, 2016; Tiwari, Gheldof, Tatari, & Christofori, 2012). This relationship is supported by the observation that CTCs isolated from the bloodstream of patients with metastatic breast cancer expressed more mesenchymal markers than the primary tumour. Additionally, de-differentiated cells at the invasive tumour edge – those which are entering into the metastatic cascade – frequently express a mesenchymal phenotype and are associated with a poorer overall survival (Prat et al., 2010; M. Yu et al., 2013). Furthermore, experimental inhibition of EMT through induction of MET or genetic knockout of EMT-inducing genes has been shown to reduce metastases (Diepenbruck & Christofori, 2016).

Whilst it is largely accepted that EMT is involved in induction of metastasis, studies have demonstrated that MET is also an essential component of metastatic colonization, responsible for the process of metastatic outgrowth at the secondary site (Liao & Yang, 2017). Tsai *et al* showed that induction of Twist1 in SCC promoted the initial dissemination of tumour cells, yet metastatic outgrowth at distant sites required inactivation of Twist1 and subsequent MET (Tsai, Donaher, Murphy, Chau, & Yang, 2012). Comparable studies which investigated the role of EMT-TFs Prrx1 and Snail1 in metastasis demonstrated a similar spatiotemporal relationship (Ocana et al., 2012; Tran et al., 2014). Furthermore, the miR-200 miRNA family have been shown to promote MET, and their expression was found to increase in metastatic deposits compared with tumour cells which lack the capacity to form metastatic deposits (Dykxhoorn et al., 2009). Of additional interest, metastatic deposits frequently display a differentiated phenotype despite the fact that the cells located at the invasive edge of the primary tumour – those

which enter the metastatic cascade – are often de-differentiated (Brabletz et al., 2001). This further supports the role of MET in establishing metastases at secondary sites.

It stands to reason that in light of their unique ability to produce new tumours, metastatic deposits should be predominantly comprised of CSCs rather than non-CSCs. Indeed, studies have found that tumour cells with CSC traits develop mature adhesion plaques – macromolecular structures involved in the initial proliferation of CSCs at secondary sites – far more readily than non-CSCs (Shibue & Weinberg, 2009, 2017). However, no study has yet discussed whether metastatic deposits express higher levels of CSCs compared to the primary tumour. The question of when and how CSC traits are acquired is also as of yet unanswered.

There are many difficulties in determining which direction of transition between epithelial and mesenchymal traits is associated with the acquisition of CSC traits, as discussed in Section 1.2.3. Many studies have examined the role of both EMT and MET in this process, yet the simple fact that these experimental studies often achieved EMT or MET through forced expression of various factors, thus fixing cells in a terminal epithelial or mesenchymal phenotype, may hold the answer to this conundrum. Contrary to the pure epithelial or mesenchymal phenotypes achieved through experimental induction of EMT/MET, *in vivo* EMT/MET is believed to be a dynamic process, as reflected in a range of studies that reported dual expression of epithelial and mesenchymal markers in CTCs (Bonnomet et al., 2012; Raimondi et al., 2011; M. Yu et al., 2013). In 2016 Beerling *et al* examined epithelial mesenchymal (E/M) plasticity in metastatic breast cancer without artificial modulation of the process (Beerling et al., 2016). They determined that an EMT is required for early migration, but not for intravasation; and importantly they showed that upon arrival at their secondary metastatic site, mesenchymal cells will revert to an epithelial phenotype within a few rounds of cell

division. Any difference in stemness traits between epithelial and mesenchymal states in the process of metastasis therefore becomes irrelevant due to the inherent plasticity of the process (Beerling et al., 2016). Indeed, this hybrid E/M state has been shown to be associated with greater stemness and enhanced tumour invasive properties in metastatic prostate cancer (Ruscetti, Quach, Dadashian, Mulholland, & Wu, 2015).

These studies suggest that whilst both EMT and MET play an important role in the generation of CSCs, this process is not associated with either EMT or MET in isolation, but rather is flexible, reflecting the E/M plasticity. Further interrogation of this model of cellular plasticity is required in order to improve our understanding of cancer progression, metastasis and potentially mechanisms of resistance to anti-cancer drug therapies.

1.2.5 The Role of EMT and CSC in Drug Resistance

As discussed in Section 1.1.6, intra-tumour heterogeneity contributes to the success of anti-cancer drug therapies through acquired drug resistance, which develops as a result of both genetic and epigenetic alterations of sub-populations of cancer cells within the tumour mass (Shibue & Weinberg, 2017) (Esteller, 2008).

A growing body of evidence supports the concept that conventional anti-cancer therapies fail because they only target the bulk non-CSC population of the tumour, allowing the more resistant CSCs to survive, self-renew and generate new tumour masses (Dean, Fojo, & Bates, 2005; Eyler & Rich, 2008; Shibue & Weinberg, 2017). Indeed, studies have investigated the relative sensitivities of isolated CSC-enriched tumour sub-populations to chemotherapy, radiotherapy, immunotherapy and molecularly targeted therapies, compared with non-CSCs. These analyses demonstrated a far greater survival of CSCs than non-CSCs in all treatment modalities, across multiple different cancer types (Dallas

et al., 2009; Graham et al., 2002; Levina, Marrangoni, DeMarco, Gorelik, & Lokshin, 2008; Shibue & Weinberg, 2017).

The process of EMT, which imparts heritable phenotypic changes through both genetic and epigenetic modifications, is one of the most important processes involved in generation of CSCs in human carcinomas (Polyak & Weinberg, 2009). EMT activation confers resistance to many different types of therapeutic agents upon tumour cells through a range of mechanisms, including elevated expression of anti-apoptotic proteins, slow stem cell proliferation rates and transmembrane protein transporters that mediate drug reflux (Figure 1.10) (Shibue & Weinberg, 2017; Singh & Settleman, 2010). Additionally, EMT-inducing TFs Zeb1, Snail and Slug have been shown to confer resistance to Oxaliplatin- and Cisplatin-based chemotherapies in breast, colon, ovarian and pancreatic cancers (Dongre & Weinberg, 2019; Lim et al., 2013). Snail and Slug promote resistance to chemotherapy through antagonization of p53 mediated apoptosis and by regulation of genes involved in cell death (Dongre & Weinberg, 2019). The association between EMT and drug resistance is further compounded by the activity of the miR-200 family: these miRNAs act to reverse EMT, and accordingly have been shown to restore chemosensitivity in aggressive cancer cells (Cochrane, Howe, Spoelstra, & Richer, 2010). The role of EMT in drug resistance is further corroborated by studies that demonstrated a strong link between treatment resistance and the expression of genes associated with EMT in cancer cells (Byers et al., 2013; Farmer et al., 2009).

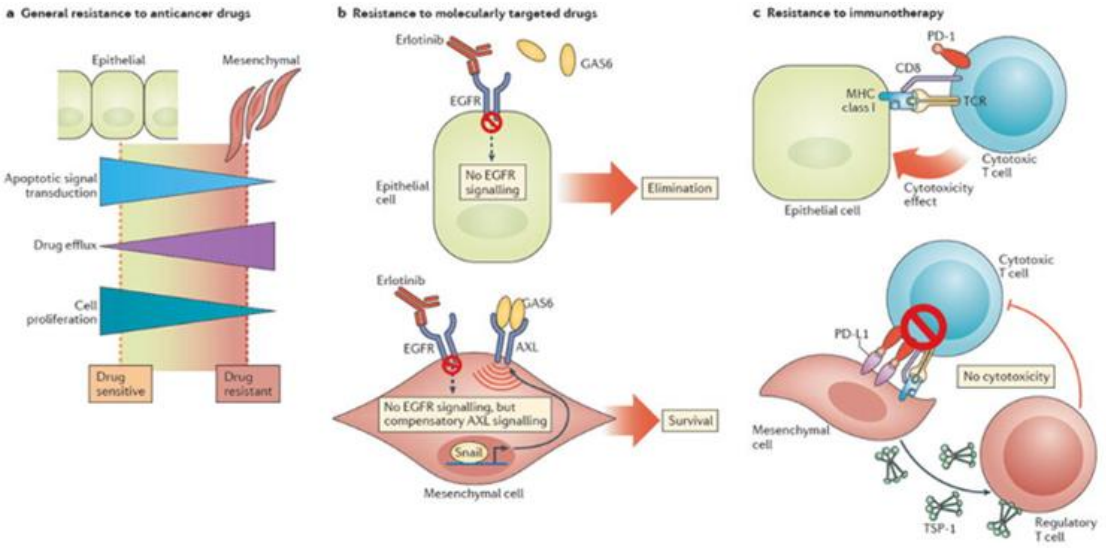


Figure 1.10: Mechanisms of EMT-Induced Anti-Cancer Treatment Resistance in Carcinomas.

Image **a** demonstrates EMT-associated downregulation of apoptotic signalling pathways, increased drug efflux and slowing of cellular proliferation. These mechanisms all contribute to general resistance to anti-cancer drugs. Image **b** depicts EMT-induced evasion of therapeutic EGFR blockade. EMT associated TF Snail1 induces surface expression of the AXL RTK; AXL signalling is triggered by binding of its ligand growth arrest-specific protein 6 (GAS6), allowing the carcinoma cell to evade the cytostatic effects of the drug. Image **c** shows how carcinoma cells, having undergone EMT, can avoid the fatal effects of cytotoxic T cells employed by immunotherapy agents. Elevated surface expression of PDL-1 binds to the PD-1 receptor on cytotoxic T cells, thus diminishing their function. Secretion of thrombospondin-1 (TSP-1) further promotes resistance through development of regulatory T cells in the tumour microenvironment that help to suppress the activity of cytotoxic T cells. Image adapted from Shibue *et al*, 2017 (Shibue & Weinberg, 2017).

The sum of these findings indicates that EMT plays a key role in regulation of CSC resistance to anti-cancer drugs, as supported by early results of clinical trials which target CSCs and their regulatory pathways (Codd, Kanaseki, Torigo, & Tabi, 2018; Yoon et al., 2014). Analysis of gastric cancer samples from a phase II clinical trial showed that patients who received chemotherapy with Vismodegib – a hedgehog inhibitor – held a survival advantage in those with a high expression of CSC marker CD44 (Yoon et al., 2014). The use of immunotherapy approaches to target CSCs are also under investigation (Codd et al., 2018). However, despite these early advances, a greater understanding of the relationship between EMT and CSCs and their mechanisms of drug resistance would undoubtedly enhance drug development and hence clinical outcomes for patients.

1.3 Cancer Stem Cells in Gastroesophageal Adenocarcinoma

1.3.1 CSC Markers

Therapeutic targeting of CSCs is limited by difficulties in characterization of CSCs across a broad range of solid and haematological malignancies. A range of markers have been identified for use in isolation of CSCs, including cell surface markers CD133, CD44, CD24 and CD66; ALDH1 activity has additionally been used as a marker of cellular metabolism (Codd et al., 2018; Prasetyanti & Medema, 2017). Unsurprisingly, given their shared characteristics, the markers used to identify CSCs overlap greatly with those used in the identification of normal adult stem cells in non-neoplastic tissues (Brungs et al., 2016). The clinical utility of these markers is somewhat hampered by the fact that expression of CSC markers is not uniform across different malignancies: indeed, CSC markers may vary between cancer subtypes and even between patients within the same subtype (Visvader & Lindeman, 2012). Heterogenous expression of CSC markers throughout a single tumour has been observed, which further complicates the process of CSC isolation (Prasetyanti & Medema, 2017). Furthermore, as discussed above, there is an inherent plasticity in the process of acquisition of CSC traits, making the isolation of CSCs more difficult. Specific CSC markers should be experimentally confirmed for individual malignancies in order to improve the accuracy of identification of this sub-population of neoplastic cells.

Several studies exist within the literature regarding the identification, regulation and clinicopathologic characteristics of CSCs and CSC-like cells in both gastric and oesophageal cancers, amongst a wide range of other malignancies. Whilst studies pertaining specifically to CSCs in GEJA are sparse, it must be remembered that tumour

samples used in studies which investigate the role of CSCs in both OAC and gastric cardia adenocarcinomas will include a proportion of GEJAs. Here we describe some of the most common CSC markers used in gastric and oesophageal malignancies.

CD133

CD133, also known as Prolamin-1, is a five transmembrane glycoprotein plasma membrane protein that has been used to identify putative CSCs in a range of tumours including colon, pancreas, brain, lung, melanoma, prostate, stomach and oesophagus (Brungs et al., 2016). It plays a role in regulation of the lipid component of the plasma membrane, yet its precise function remains unknown (Codd et al., 2018). Whilst frequently used as a marker of CSCs, CD133 is not a CSC-specific antigen as it is also expressed in a number of differentiated epithelial cells in various organs (Y. Wu & Wu, 2009). It is important also to recognise that the use of different CD133 clones complicates comparisons between studies, leading to poor reducibility and potential for erroneous results (Hermansen et al., 2011). Despite this, the utility of CD133 as a target for anti-CSC therapies has been investigated in ovarian mouse models, demonstrating inhibition of cellular growth and suppression of tumour progression (Skubitz et al., 2013).

CD44

CD44 is a transmembrane glycoprotein that is expressed on both differentiated adult cells, including endothelial cells and hepatocytes, and CSCs. It has a wide range of physiological roles including adhesion, migration, differentiation, growth and survival (Ponta, Sherman, & Herrlich, 2003). It serves as a putative CSC marker in a range of

malignancies including colon, brain, stomach, oesophagus and – when interpreted in combination with decreased CD24 expression – prostate and breast (Brungs et al., 2016). However, as with CD133, we cannot regard CD44 as a CSC-specific antigen.

Interestingly, CD44 is encoded by the 20 exon CD44 gene, which is subject to alternative splicing (Lau et al., 2014). It has been proposed that CD44 variants (CD44v) are more specific in their identification of cells with tumorigenic potential when compared to the standard isoform (CD44s) (Thapa & Wilson, 2016). Moreover, whilst we know that cancer cells which undergo EMT express increased levels of CD44, the functional significance is unclear, with great overlapping between the different isoforms (C. Chen, Zhao, Karnad, & Freeman, 2018). Brown *et al* demonstrated that induction of EMT in breast cancer required a switch from CD44v to CD44s isoform expression (Brown et al., 2011). These findings were supported by a number of other studies that identified CD44v in metastatic deposits from a range of solid organ malignancies, which were associated with a poorer prognosis (Kaufmann et al., 1995; Mulder et al., 1994; Ni et al., 2014; Ozawa et al., 2014). These findings identify specific CD44 isoforms as potential targets for anti-cancer therapies, and indeed early studies have already investigated the potential for therapeutic targeting of CD44 positive breast CSCs (Aires et al., 2016).

ALDH1

Within the human genome, the aldehyde dehydrogenase (ALDH) family comprises a reported 19 functional genes which encode enzymes involved in the oxidative metabolization of endogenous and exogenous aldehyde substrates, including lipids and amino acids (Tomita, Tanaka, Tanaka, & Hara, 2016). ALDH1 has 3 isoforms (ALDH1A1, ALDH1A2 and ALD1A3) and is a marker of both stem cells and CSCs,

with expression seen in colon, pancreas, breast and prostate cancers (Brungs et al., 2016; Tomita et al., 2016). Katsuno *et al* demonstrated CSC properties of self-renewal and increased tumorigenicity in isolated $ALDH1^+$ cells from gastric cancer cell lines (Katsuno et al., 2012), and in other studies high ALDH expression has been correlated with poor clinical outcomes in pancreatic, ovarian and prostate cancers (Fitzgerald & McCubrey, 2014; Kuroda et al., 2013; Le Magnen et al., 2013). Furthermore, acquired drug resistance in tumour cells is associated with transcriptional activation of ALDH1 expression (Yoshida, Dave, Han, & Scanlon, 1993). Early studies have investigated the utility of anti-CSC therapies targeting ALDH1 in breast, ovary and NSCLC (Duan et al., 2014; H. Z. Li, Yi, & Wu, 2008; MacDonagh et al., 2017; Schech, Kazi, Yu, Shah, & Sabnis, 2015; Y. H. Wu et al., 2015). A phase II trial investigated the effect of administering Disulfiram – a potent ALDH inhibitor – in addition to standard chemotherapy to patients with NSCLC, demonstrating good drug tolerance and a prolonged survival (Nechushtan et al., 2015). Thus, in cancer therapy, ALDH1 holds great potential as a CSC target.

Other Potential CSC Markers

There is a vast array of additional markers that have been used to identify CSCs across a range of malignancies, used either on their own or in combinations. These include CD24, EpCAM, CD49f, CD54, CD90, CD117, CD166, CD177, CD29, Lgr5 and AFP (Brungs et al., 2016; Codd et al., 2018). The evidence supporting their utility as CSC markers is either weak or inconsistent, thus further studies are required to determine their suitability for this role.

1.3.2 Gastric Adenocarcinoma CSC Markers

Gastric CSC markers are well described in the published literature, with strong evidence to support CD44, CD133 and ALDH1 as reliable cell surface and metabolic markers for this sub-population (L. Lu et al., 2016; Tomita et al., 2016; Zavros, 2017). Expression of all three markers has been consistently associated with poor clinicopathological features and overall survival outcomes in gastric cancer (Abdi et al., 2019; S. Chen et al., 2013; L. Lu et al., 2016; Wakamatsu et al., 2012). Indeed, CD133 and CD44 were each found to be independent predictors of lower DFS and OS, prompting Lu *et al* to suggest that their expression interpreted in combination may be of use as a prognostic tool in gastric cancer (S. Chen et al., 2013; Lee, Seo, An, Kim, & Jeon, 2012; L. Lu et al., 2016; Mayer et al., 1993).

CD44v6 expression in gastric cancer resection specimens has been shown to be associated with poorer clinical outcomes including distant metastasis, lymph node metastasis and depth of invasion (S. Chen et al., 2013; Y. J. Liu, Yan, Li, & Jia, 2005). Additionally, $CD44^+$ CTCs in patients with gastric cancer were shown to correlate with the clinicopathologic characteristics of the resected tumour specimens, yet $CD44^-$ CTCs did not (Watanabe et al., 2017). These findings suggest that CD44 is useful as a marker of stemness in neoplastic cells and as a predictor of patient outcomes. However, whilst many studies point towards the utility of CD44 as a prognostic tool and potential therapeutic target in gastric cancer, it is important to bear in mind the impact of heterogeneity in patient populations, tumour sampling and experimental procedures: this is demonstrated in a case series by Kim *et al*, which did not identify an association between CD44 expression and any clinicopathological factors examined (J. Y. Kim, Bae, Kim, Shin, & Park, 2009; Y. J. Liu et al., 2005).

A meta-analysis conducted by Wen *et al* in 2013 investigated the correlation between $CD133^+$ gastric cancers and clinicopathological characteristics and survival outcomes in 773 patients (Wen et al., 2013). The study found worse accumulative 5 year OS rates in $CD133^+$ patients (21.4%) as compared with $CD133^-$ patients (55.7%), in addition to a close correlation between CD133 over-expression and poor clinicopathological features, including TNM stage and lymphovascular invasion (Wen et al., 2013). A more recent study demonstrated higher levels of $CD133^+$ CTCs in blood samples from gastric cancer patients compared to unmatched normal controls, the presence of which correlated with poor prognosis (Xia, Song, Liu, Wang, & Xu, 2015). These findings are in keeping with the use of CD133 as a marker for putative gastric CSCs.

Wakamatsu *et al* showed no association between ALDH1 and survival outcomes, yet deeper analysis of the individual ALDH isoforms by Li *et al* pointed towards ALDH1A3 and ALDH1L1 as potential prognostic markers and therapeutic targets in gastric cancer (K. Li et al., 2016; Wakamatsu et al., 2012).

The epithelial molecular adhesion molecule (EpCAM) is a transmembrane glycoprotein present in most epithelial tissues that plays a role in cell adhesion, migration and differentiation (Imano et al., 2013). EpCAM is over-expressed in gastric cancer, with one study demonstrating CSC characteristics within $EpCAM^+$ tumour population and not in $EpCAM^-$ tumour cell (Wenqi et al., 2009). Furthermore, Imano *et al* showed that peritoneal metastases of gastric cancer express higher levels of EpCAM, as compared with biopsy samples of the primary tumour, indicating that only gastric cancer cells with high EpCAM expression may metastasize to the peritoneum (Imano et al., 2013). Despite this, most gastric cancers are $EpCAM^+$, thus it must be used in conjunction with other more specific markers in identification of gastric CSCs (Brungs et al., 2016).

Many other cell surface markers have been identified which may hold promise as gastric CSC-like markers, however as of yet studies are conflicting. One study showed that decreased expression of CD54, also known as intercellular adhesion molecule 1 (ICAM1), is associated with a poorer prognosis and increased risk of lymphatic spread, yet a separate group correlated $CD44^+/CD54^+$ gastric cancer cells with stemness traits, as compared to $CD44^-/CD54^-$ cells (T. Chen et al., 2012; Yashiro, Sunami, & Hirakawa, 2005). Other cell surface markers that hold similar potential as markers of gastric CSC include CD90, CD71 and CD49f (Brungs et al., 2016).

Interestingly, studies have examined both the expression levels of stemness genes and their related proteins in gastric cancer. TFs including Sox2, Nanog and Oct-4 have been shown to be expressed in gastric cancer cells and their co-expression is proposed to identify gastric CSCs (J. Liu et al., 2013). Despite this, the results of studies looking at the expression of these markers in gastric cancer have thus far produced conflicting results, demanding further experimental interrogation to confirm their utility as CSC markers (N. Li et al., 2015; Matsuoka et al., 2012; Otsubo, Akiyama, Yanagihara, & Yuasa, 2008).

A number of miRNAs have been linked to the expression of gastric CSCs. miR-196a-5p has been shown to be upregulated in $CD44^+$ gastric CSCs, and to play a key role in EMT and invasion through targeting of the Smad4 signalling pathway (Pan et al., 2017). High miR-501-5p levels were associated with poor OS and were shown to induce a CSC like phenotype in gastric cell lines through activation of Wnt/β-catenin signalling pathways (Fan, Ren, Yang, Liu, & Zhang, 2016). Furthermore, upregulation of miR-132 in gastric CSCs was linked to chemoresistance (L. Zhang et al., 2017). These miRNAs, amongst many others, hold great promise as a targetable molecule in the treatment of gastric

cancer, yet extensive work is required to validate their prognostic significance and mechanisms of action.

1.3.3 Oesophageal Adenocarcinoma CSC Markers

Oesophageal CSC markers have also been widely investigated, with particular emphasis on their role in the BO to OAC transition. Similar to gastric cancer, ALDH1, CD44 and CD133 are well recognised as putative CSC markers, in addition to EpCAM (Islam, Gopalan, & Lam, 2018; Mokrowiecka et al., 2017).

Honing *et al* described a correlation between loss of CD44 expression and poor survival outcomes in patients with OAC. Ajani *et al* showed that $ALDH1^+$ tumour cells from OAC and GEJA resection specimens were more resistant to chemoradiotherapy, as compared to tumour cells with low ALDH1 expression (Ajani et al., 2014; Honing et al., 2014). A similar study by Sun *et al* demonstrated that resistance to treatment with Adriamycin, Cisplatin and 5-FU (ACF) was associated with an increase in EpCAM expression, which coincided with expression of CSC marker CD90 (Sun et al., 2018). By comparison, Driemel *et al* concluded that EpCAM expression is dynamic throughout tumour progression, with high expression correlating with proliferative stages and low/negative expression associated with migration, invasion and dissemination (Driemel et al., 2014). Alakhova *et al* also investigated the role of OAC CSCs in drug resistance, examining the ability of SP1049C – a pluronic-based micellar formulation of Doxorubicin that has demonstrated safety and efficacy in patients with advanced OAC and GEJA in a phase II trial – to deplete CSCs and decrease cancer cell tumorigenicity in vivo (Alakhova, Zhao, Li, & Kabanov, 2013). Using mouse models, they found that $CD133^+$ CSC populations

were decreased and tumorigenicity was suppressed following SP1049C treatment. These findings all suggest a link between OAC CSCs and resistance to anti-cancer treatments.

The potential for CSCs to serve as prognostic markers for the progression from BO to OAC has also been investigated. Mokrowiecka *et al* identified higher expression of CD44, CD133 and EpCAM in BO, early OAC and advanced OAC compared with normal gastric cardia, suggesting that these markers may hold potential as markers of progression from BO to OAC (Mokrowiecka et al., 2017). In 2012 Tomizawa *et al* investigated expression of CSC marker CD133 and EMT-TFs Snail, Slug and Twist in early OACs using IHC. They found abundant expression of each marker at the invasive tumour edge, indicating that early stage cancers contain cells with metastatic potential (Tomizawa, Wu, & Wang, 2012).

miRNAs have been implicated in the regulation of CSC traits in OAC tumour cells. Downregulation of miR-17-5p in OAC tumour cells with CSC traits was shown to produce a radioresistant phenotype (Lynam-Lennon et al., 2017). Similarly, over-expression of miR-221 in OAC was associated with resistance to 5-FU based chemotherapeutic regimens; experimental knockdown in resistant cells resulted in dysregulation of CD44 in addition to other Wnt/β-catenin signalling target genes (Y. Wang et al., 2016). These findings, taken in conjunction with protein and potential mRNA CSC markers, merit greater interrogation as the co-expression of different molecular markers may hold great promise as targets for anti-cancer therapies.

1.3.4 Gastroesophageal Junction Adenocarcinoma CSC Markers

There is a distinct lack of studies in the current literature which investigate the role of CSCs in GEJA as a distinct entity. Brungs *et al* examined the significance of the expression of CD133, CD44 and ALDH1 in metastatic deposits of GEJA, which they defined as 'gastroesophageal and gastric cancers' (Brungs et al., 2019). CD44 and ALDH1 expression were both significantly associated with poorer OS; CD44 was identified as an independent prognostic marker, whilst CD133 was not. Both CD44 and ALDH1 were also significantly associated with urokinase-type plasminogen activator receptor (uPAR) expression. The uPAR system is a proteolytic pathway that is involved in invasion of tumour cells into the surrounding normal tissue early in the metastatic cascade: its expression is an independent prognostic factor for GEJA and is thought to play a role in CSC signalling (Brungs et al., 2017; Gilder et al., 2018). These findings support the use of CD44 and ALDH1 as CSC markers in GEJA.

As described in Section 1.3.3, two studies by Ajani and Alakhova specifically stated that the cancer cells used in their studies were sourced from both OAC and GEJA specimens. It is likely that GEJ tumours accounted for a proportion of the tumour samples used in all studies discussing the role of CSCs in OAC and gastric cancers, yet this assumption cannot be proven. Regardless, in light of the growing belief that GEJ tumours are best regarded as a disease entity in their own right, more focused attention is required to determine the specific molecular characteristics of GEJA, rather than grouping them with non-junctional OAC and gastric cancers in study designs.

1.4 Study Aims and Objectives

The study hypothesis of this thesis states that the presence of rare populations of CSCs are associated with more clinically aggressive GEJA, and that these cells most likely to be located in poorly differentiated areas of tumour, and/or in the infiltrative or pushing edge of the tumour, here deemed 'higher grade' areas. By comparison, it is expected that tumour cells in better differentiated and/or less infiltrative areas of tumour – here deemed 'lower grade' areas – will be associated with better clinical outcomes and a non-CSC phenotype. This question was devised in order to address the gap in the literature regarding the identification, function and clinical impact of CSCs in GEJA.

The fundamental aim of this project is to perform a large scale expression analysis of GEJA resection specimens using a panel of associated mRNAs, miRNAs and proteins. The panel of markers was chosen to reflect the interplay between EMT and CSCs, with individual markers identified through a meta-analysis of the relevant literature. It is hoped that this analysis will identify a molecular signature that is of prognostic significance for patients with GEJA. The objectives of our study are:

- To examine the expression of genes relating to EMT and 'stemness' using quantitative PCR (qPCR), comparing the difference in expression patterns between 'lower grade' and 'higher grade' areas of the same tumour, where possible.
- To determine the expression of miRNAs known to play a role in EMT and in the acquisition of 'stemness' traits using qPCR, comparing the difference in expression patterns between 'lower grade' and 'higher grade' areas of the same tumour, where possible.

- To study the downstream protein expression in the same ‘lower grade’ and ‘higher grade’ areas of tumour, focusing on proteins associated with EMT and CSCs, through use of tissue microarrays (TMA) and IHC.
- To correlate the molecular signature of different areas of tumour with clinicopathological characteristics.

Chapter Two: Materials and Methods

2.1 Gene Expression Analysis

2.1.1 Sample Selection

Seventy nine formalin fixed (10% buffered formalin) paraffin embedded (FFPE) tissue samples were identified through the St. James's Hospital Upper Gastrointestinal biobank and retrieved from the main hospital archive. All samples were stored following written informed consent, as approved by the St. James's Hospital (SJH)/Adelaide and Meath Hospital Dublin incorporating National Children's Hospital (AMNCH) research ethics committee (REC Reference 041113/10804 and 2018-08 Chairman's action (10)). FFPE tissue from resection specimens of patients who underwent surgery with curative intent for gastroesophageal junction (GEJ) adenocarcinomas was used. Each tumour was scored according to the TNM 8th Edition. Exclusion criteria were age under 18 years; an inability to provide informed consent; (y)pT0 and (y)pT1a disease, as the latter is limited to the mucosal layer; gastric or oesophageal malignancies that did not involve the GEJ. Patients with only biopsy specimens available were not included. All confidentiality regulations were adhered to and data was irreversibly anonymised following collection. The clinical course for each patient was not affected by sample collection or subsequent data collection.

2.1.2 Histological Assessment

One - two FFPE tissue blocks containing tumour were selected per case to include two specific areas of tumour, based on histological appearance. Each tumour block was assessed to identify two histologically distinct areas of tumour, categorised as 'higher'

and ‘lower’ histologic grade. Higher grade areas were characterised by the presence of tumour cells with a lack of gland formation, solid growth pattern, clusters and single tumour cells. These cells were frequently located at the infiltrative/pushing tumour border; however, they could often be found scattered throughout the tumour body. Lower grade areas had more discernible gland formation, as compared to their matched pair. These cell populations were most commonly identified in less infiltrative areas of the tumour body. Due to variability in response to NAT, not all cases included a distinct ‘lower grade’ area of tumour, however a ‘higher grade’ tumour sample was taken from each case. Tumours composed of a high volume of architecturally similar tumour cells classified their paired samples based on location of neoplastic cells within the body of the tumour.

2.1.3 Laser Capture Microdissection

Two 7 µm sections were cut from each tissue block using a microtome and mounted on uncharged glass slides. The slides were air dried and each section was stained with a standard haematoxylin and eosin (H&E) stain. The slides were not cover-slipped. Laser capture microdissection (LCM) was performed using the Arcturus® system and Arcturus® CapSure® Macro LCM Caps (Applied Biosystems™ by Thermo Fisher Scientific, Waltham, MA, USA) (Figure 2.1).

The LCM system utilised an Eclipse™ Ti-E microscope and an ArcturusXT™ Instrument. The LCM caps and slides were loaded onto the stage. An overview image of each slide to be microdissected was captured on the computer. The cells of interest were marked on-screen for infrared (IR) and/or ultraviolet (UV) laser microdissection using

freehand and defined circle drawing tools. Tumour cells were selected from each histologically defined area in turn, using a separate cap for each.

Once the area of interest was marked out, the IR and/or UV lasers were tested to confirm correct location of the laser spot. The LCM cap was loaded onto the slide. The IR and/or UV lasers were used to microdissect the marked areas and these isolated cells of interest were collected onto the transfer film of the LCM caps. Repeat microdissection was performed as indicated in order to collect the maximum number of cells of interest available on the slide. Once complete, the cap was unloaded from the stage, coded and stored in a cool dry room prior to RNA extraction.

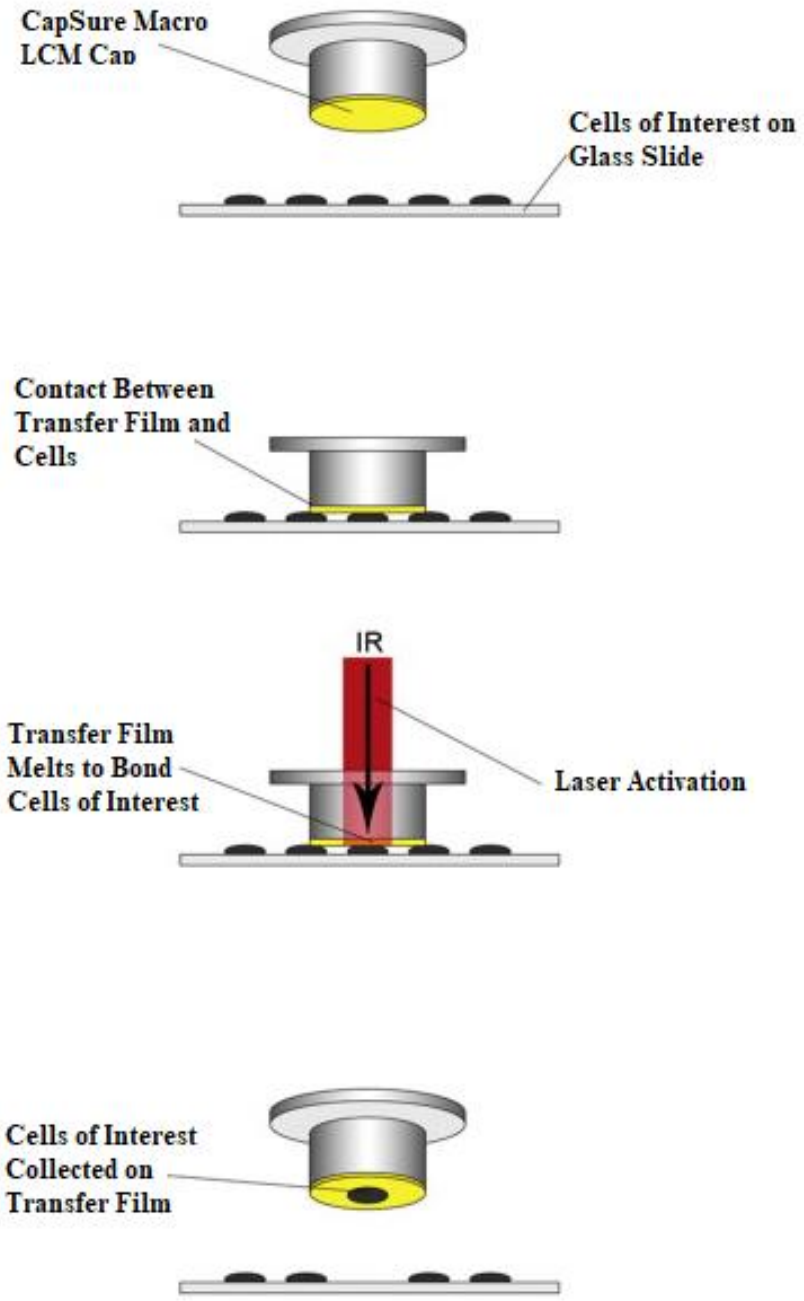


Figure. 2.1 The Laser Capture Microdissection Process. Adapted from Vandewoestyne *et al*, 2013 (Vandewoestyne et al., 2013).

2.1.4 RNA Isolation

RNA isolation was performed using the Arcturus™ PicoPure™ RNA Isolation Kit (Applied Biosystems™ by Thermo Fisher Scientific) according to manufacturer's instructions, as briefly described below.

RNA Extraction

RNA was first extracted by pipetting 50 µL extraction buffer into a 0.5 mL microcentrifuge tube. The CapSure® Macro LCM Cap was inserted onto the tube and this assembly was inverted to ensure all the extraction buffer covered the cap. The cap-microcentrifuge tube assembly was then incubated at 42°C for 30 min, following which it was centrifuged at 800 x *g* for 2 min to collect the cell extract into microcentrifuge tube. The cap was then removed, and the cell extract was either directly frozen at -80˚C or used immediately.

RNA Isolation

RNA isolation first involved pre-conditioning of the RNA purification column: 250 µL conditioning buffer was pipetted onto the purification column filter membrane and incubated for 5 min at room temperature before centrifugation at 16,000 x *g* for one min. Once complete, 50 µL 70% EtOH was pipetted into the cell extract from the first step of this process. This cell extract/EtOH mixture (combined volume approx. 100 µL) was pipetted into the pre-conditioned purification column and centrifuged for 2 min at 100 x *g* to bind RNA to the column. This was immediately followed by centrifugation at 16,000 x *g* for 30 sec to remove flow-through. A further 100 µL 'wash buffer 1' was then pipetted

into the purification column and centrifuged at 8,000 x *g* for one min, followed by pipetting 100 µL 'wash buffer 2' into the purification column and centrifuging at 8,000 x *g* for 1 min. Subsequently, an additional 100 µL 'wash buffer 2' was pipetted into the purification column and centrifuged at 16,000 x *g* for 2 min.

The purification column was then transferred to a new 0.5 mL microcentrifuge tube and 15 µL elution buffer was directly pipetted onto the membrane of the purification column by gently touching the tip of the pipette to the surface of the membrane as the buffer was dispensed, thus ensuring maximum absorption of the elution buffer onto the membrane. The purification column was then incubated at room temperature for 1 min, followed by centrifuging for 1 min at 1,000 x *g* to distribute the elution buffer in the column. Immediately after this, the column was centrifuged for 1 min at 16,000 x *g* to elute the RNA. Samples were then either directly frozen at -80˚C or used immediately.

2.1.5 Reverse Transcription

Reverse transcription (RT) was performed using a Thermo Fisher Scientific High-Capacity cDNA Reverse Transcription Kit and RNase inhibitor (Applied Biosystems™ by Thermo Fisher Scientific), according to manufacturer's instructions, as briefly described below.

Preparation of 2X Reverse Transcription Master Mix

The 2X RT master mix components were thawed on ice and the master mix was prepared on ice, using the volume components in Table 2.1. Volume calculated was that required per 20 µL reaction. The 2X master mix was then placed on ice and mixed gently.

Table 2.1: Volume components for reverse transcription master mix.

Component	Volume/Reaction (µL)
10X RT Buffer	2.0
25X dNTP Mix (100 mM)	0.8
10X RT Random Primers	2.0
MultiScribe™ Reverse Transcriptase	1.0
RNase Inhibitor	1.0
Nuclease-Free H_2O	8.2
Total Per Reaction	15

Preparation of cDNA Reverse Transcription Reactions

The cDNA RT reactions were prepared. Fifteen µL 2X RT master mix was pipetted into each well of a 96 well reaction plate, followed by 5 µL RNA sample which was gently pipetted to mix the sample thoroughly. The plate was sealed using MicroAmp™ Optical Adhesive Film (Applied Biosystems™ by Thermo Fisher Scientific) and then briefly centrifuged to spin down the contents and eliminate any air bubbles.

Reverse Transcription Reaction

RT was performed using the Veriti Dx 96 well thermal cycler (Applied Biosystems™ by Thermo Fisher Scientific) under the conditions outlined in Table 2.2.

Table 2.2: Thermal cycler conditions required for reverse transcription.

	Step 1	Step 2	Step 3	Step 4
Temperature (°C)	25	37	85	4
Time	10 min	120 min	5 sec	∞

2.1.6 Preamplification

Preamplification was performed prior to PCR amplification, following the TaqMan® PreAmp Master Mix (2X) Kit (Applied Biosystems™ by Thermo Fisher Scientific) protocol, according to manufacturer's instructions, as briefly described below.

TaqMan® Assays

The TaqMan® Gene Expression Assays (20X) (Applied Biosystems™ by Thermo Fisher Scientific) used in this experiment were: POU5F1 (Hs04260367_gH), NANOG (Hs02387400_g1), Vimentin (Hs00958111_m1), CDH1 (Hs01023895_m1), Serpine-1 (Hs00167155_m1) and GAPDH (Hs03929097_g1). Prior to use they were thawed on ice, vortexed and centrifuged briefly. The assays were first pooled by combining equal volumes of each, followed by dilution using 1X TE buffer. This ensured that the final concentration of each assay was 0.2X.

Preamplification

Ten preamplification cycles were used in this experiment due to the low number of pooled assays. These conditions produced material sufficient for fifty 20 µL PCR amplification reactions. Each preamplification reaction was prepared in a 0.2 mL microcentrifuge tube, using the volume components described in Table 2.3.

Table 2.3: Volume components required for preamplification reaction.

Component	Volume (µL/Reaction)	Final Concentration
TaqMan® PreAmp Master Mix (2X)	25.0	1X
Pooled Assay Mix	12.5	0.05X (each assay)
1-250 ng cDNA Sample + Nuclease-Free H_2O	6 µL cDNA + 6.5 µL nuclease free H_2O	0.02-5.0 ng/µL
Total	50.0	-

The reactions were mixed by gentle inversion followed by centrifugation. The tubes were placed in the thermal cycler and run using the conditions outlined in Table 2.4.

Table 2.4: Thermal cycler conditions required for preamplification PCR.

	Enzyme Activation	Preamplification PCR		
	HOLD	**CYCLE**		**HOLD**
		Denature	**Anneal/Extend**	
Temperature	95°C	95°C	60°C	4°C
Time	10 min	15 sec	4 min	∞

As soon as the run was complete, the plate was placed on ice. The preamplification products were then diluted according to Table 2.5 and the samples were either directly frozen at -20˚C or used immediately.

Table 2.5: Dilution of preamplification products.

Number of Preamplification Cycles	Dilution Factor of Preamplification Products with 1X TE Buffer	Final Volume of Diluted Preamplification Product
10	1:5 (50 µL + 200 µL TE)	250 µL

2.1.7 PCR Amplification

The TaqMan® Gene Expression assays (20X) listed above were used in this experiment. All PCR runs were performed using the 7500 Fast Real-Time PCR System (Applied Biosystems™ by Thermo Fisher Scientific). The reactions were as outlined in Table 2.6.

Table 2.6: Volume components for PCR amplification.

Component	Volume (µL/Reaction)
TaqMan® Gene Expression Assay (20X)	1.0
Diluted Preamplified cDNA Products (Diluted 1:5)	5.0
TaqMan® Gene Expression Master Mix	10.0
Nuclease-Free H_2O	4.0
Total Volume	20.0

The 96 well fast plate was sealed with MicroAmp™ Optical Adhesive Film, then briefly vortexed and centrifuged. The cycling conditions are outlined in Table 2.7.

Table 2.7: Thermal cycling conditions for PCR amplification.

Step	UDG Activation	AmpliTaq Gold® Enzyme Activation	PCR	
	HOLD	**HOLD**	**CYCLE (40 CYCLES)**	
			Denature	Anneal/Extend
Temperature	50°C	95°C	95°C	60°C
Time	2 min	10 min	15 sec	1 min

2.2 miRNA Analysis

2.2.1 Sample selection

RNA samples used in this study were acquired as described in Sections 2.1.1-2.1.4. Total RNA was isolated using the Arcturus™ PicoPure™ RNA Isolation Kit and was thus suitable for miRNA analysis due to the collection of small RNAs.

2.2.2 cDNA Template Preparation

cDNA was synthesised using the TaqMan® Advanced miRNA cDNA Synthesis Kit (Applied Biosystems™ by Thermo Fisher Scientific) protocol, according to manufacturer's instructions, as briefly described below.

Poly(A) Tailing Reaction

The RNA samples and cDNA synthesis reagents were thawed on ice, then vortexed and centrifuged briefly. The Poly(A) reaction mix was then prepared in a 1.5 mL microcentrifuge tube, using the volumes in Table 2.8. The reaction mix was vortexed and centrifuged to mix samples and eliminate air bubbles. A 2 µL RNA sample was added to each well of a 96 well plate, followed by 3 µL Poly(A) reaction mix. Total volume: 5 µL per well. The plate was then sealed, vortexed and centrifuged briefly.

Table 2.8: Volume components for Poly(A) reaction mix.

Component	Volume (µL)
10X Poly(A) Buffer	0.5
ATP	0.5
Poly(A) Enzyme	0.3
RNase Free H_2O	1.7
Total Poly(A) Reaction Mix Volume	3.0

The plate was placed in the thermal cycler under the conditions in Table 2.9. Once complete, the adaptor ligation reaction commenced immediately.

Table 2.9: Thermal cycler conditions required for poly(A) tailing reaction.

Step	Temperature (°C)	Time
Polyadenylation	37	45 min
Stop Reaction	65	10 min
Hold	4	Hold

Adaptor Ligation Reaction

First the ligation reaction mix was prepared in a 1.5 mL microcentrifuge tube using volumes indicated in Table 2.10, ensuring that 50% PEG 8000 solution was at room temperature. The reaction mix was vortexed and centrifuged to mix samples and eliminate air bubbles.

Table 2.10: Volume components for adaptor ligation reaction mix.

Component	Volume (µL)
5X DNA Ligase Buffer	3
50% PEG 8000	4.5
25X Ligation Adaptor	0.6
RNA Ligase	1.5
RNase Free H_2O	0.4
Total Ligation Reaction Mix Volume	10

Ten µL adaptor ligation reaction mix was transferred to each well of the plate containing the poly(A) tailing reaction product, with a total volume of 15 µL per well. The plate was sealed, vortexed and centrifuged before placing it in the thermal cycler under the conditions in Table 2.11. Once complete, the RT reaction immediately commenced.

Table 2.11: Thermal cycler conditions required for adaptor ligation reaction.

Step	Temperature (°C)	Time
Ligation	16	60 min
Hold	4	Hold

Reverse Transcription

The RT reaction mix was prepared in a 1.5 mL microcentrifuge tube, using volumes indicated in Table 2.12. The reaction mix was vortexed and centrifuged to mix samples and eliminate air bubbles.

Table 2.12: Volume components for reverse transcription reaction mix.

Component	Volume (µL)
5X RT Buffer	6
dNTP Mix (25mM Each)	1.2
20X Universal RTPprimer	1.5
10X RT Enzyme Mix	3
RNase Free H_2O	3.3
Total RT Reaction Mix Volume	15

Fifteen µL RT reaction mix was transferred to each well of the plate containing the adaptor ligation reaction product. Total volume: 30 µL per well. The plate was sealed, vortexed and centrifuged before placing it in the thermal cycler under the conditions in Table 2.13. Once complete, samples were either directly frozen at -20˚C or used immediately.

Table 2.13: Thermal cycler conditions required for reverse transcription.

Step	Temperature (°C)	Time
Reverse Transcription	42	15 min
Stop Reaction	85	5 min
Hold	4	Hold

miR-Amp Reaction

First the miR-Amp reaction mix was prepared in a 1.5 mL microcentrifuge tube, using volumes indicated in Table 2.14. The reaction mix was vortexed and centrifuged to eliminate air bubbles.

Table 2.14: Volume components for miR-Amp reaction mix.

Component	Volume (µL)
2X miR-Amp Master Mix	25
20X miR-Amp Primer Mix	2.5
RNase Free H_2O	17.5
Total miR-Amp Reaction Mix Volume	45

A total of 45 µL miR-Amp reaction mix was transferred to each well of a new 96 well plate. Five µL RT reaction product was then added to each well with a total volume of 50 µL per well. The plate was sealed, vortexed and centrifuged before being placed in the thermal cycler under the conditions in Table 2.15, with incubation under MAX ramp speed and standard cycling. Upon completion, samples were either directly frozen at -20˚C or used immediately.

Table 2.15: Thermal cycler conditions required for miR-Amp reaction.

Step	Temperature (°C)	Time	Cycle
Enzyme Activation	95	5 min	1
Denature	95	3 sec	14
Anneal/Extend	60	30 sec	
Stop Reaction	99	10 min	1
Hold	4	Hold	1

2.2.3 PCR Amplification

TaqMan® Advanced miRNA assays (Applied Biosystems™ by Thermo Fisher Scientific) used in this experiment are listed in Table 2.16. TaqMan® Fast Advanced Master Mix (Applied Biosystems™ by Thermo Fisher Scientific) was also used. All PCR was undertaken using the 7500 Fast Real-Time PCR System.

Table 2.16: miRNAs used in expression analysis study.

miRNA	Assay ID
hsa-miR-224-3p	478780_mir
hsa-miR-221-3p	477981_mir
hsa-miR-21-5p	477975_mir
hsa-miR-17-5p	478447_mir
hsa-miR-10b-5p	478494_mir
hsa-miR-16-5p	477860_mir
hsa-miR-141-5p	478712_mir
hsa-miR-203a-3p	478316_mir
hsa-miR-103a-3p	478253_mir
hsa-miR-223-3p	477983_mir
hsa-miR-200a-3p	478490_mir
hsa-miR-133b	480871_mir

First the miRNA assays were thawed on ice, then vortexed and centrifuged briefly. A 1:10 dilution was prepared of cDNA templates produced in 2.2.2 (Table 2.17).

Table 2.17: Dilution required to produce a 1:10 dilution of preamplification product.

Dilution Factor of cDNA Templates with 0.1X TE Buffer	Final Volume of Diluted Preamplification Product
1:10 (20 µL + 180 µL TE)	200 µL

A PCR reaction mix was prepared in a 1.5 mL microcentrifuge tube, using volumes indicated in Table 2.18. The reaction mix was vortexed and centrifuged to mix samples and eliminate air bubbles.

Table 2.18: Volume components for PCR reaction mix.

Component	Volume (µL)
TaqMan® Fast Advanced Master Mix (2X)	10
TaqMan® Advanced miRNA Assay (20X)	1
RNase Free H_2O	4
Total PCR Reaction Mix Volume	15

Fifteen µL PCR reaction mix was transferred to each well of a 96 well fast plate, then 5 µL diluted cDNA template was added to each well with a total volume per well of 20 µL. The plate was sealed, vortexed and centrifuged before using the PCR thermal cycling conditions outlined in Table 2.19.

Table 2.19: Thermal cycling conditions for PCR amplification.

Step	Temperature (°C)	Time	Cycles
Enzyme Activation	95	20 sec	1
Denature	95	3 sec	40
Anneal/Extend	60	30 sec	

2.3 Immunohistochemical Analysis

2.3.1 Tissue Selection

The same FFPE tissue blocks chosen for RNA extraction (Sections 2.1.1-2.1.4) were used for IHC analysis. Where possible, each case had two separately identifiable areas based on histological appearances: an ‘higher grade’ area, featuring poor tumour differentiation and/or an infiltrative/pushing boarder; and an ‘lower grade’ area, as defined by better differentiation and/or a less infiltrative portion of the tumour. Internal controls chosen were liver, kidney and normal GEJ.

2.3.2 Tissue Micro-Array Construction

Five tissue micro-arrays (TMA) were constructed using a Beecher TMA instrument using 1 mm punches (Estigen, Tartu, Estonia) (Figure 2.2).

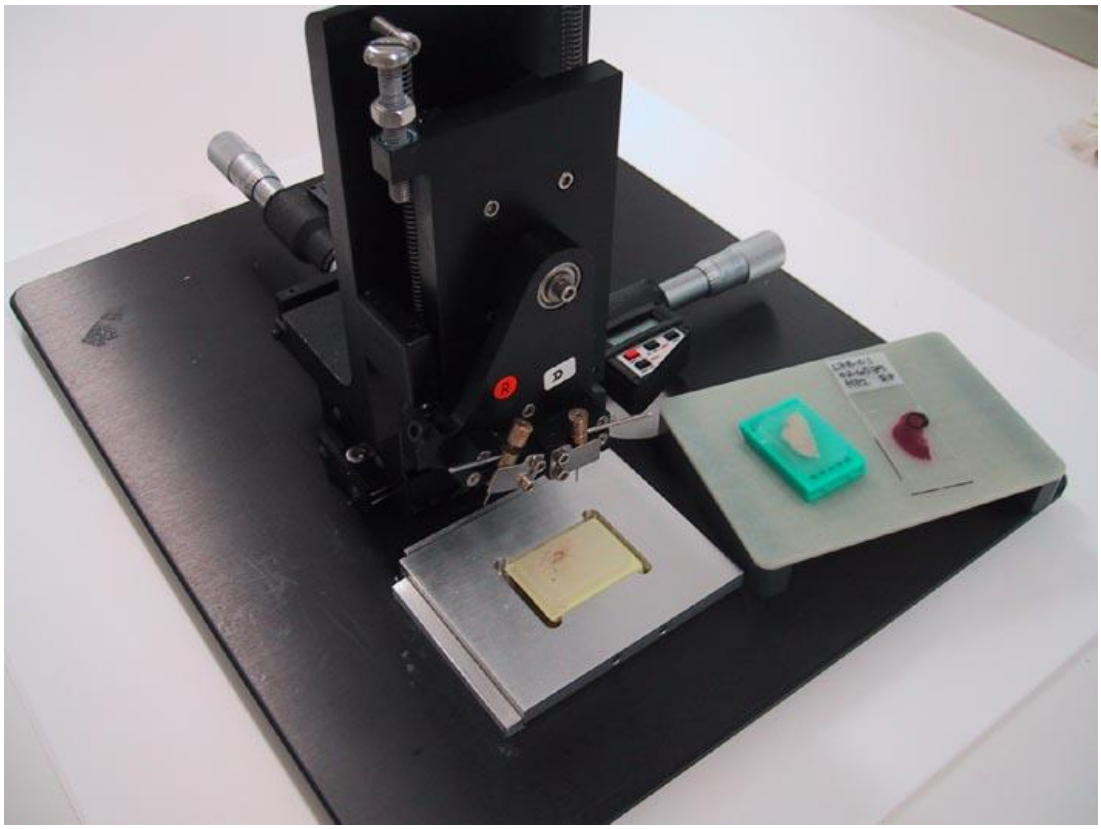


Figure 2.2 Beecher Tissue Microarray Equipment.

Source: https://www.ous-research.no/cytometry/?k=cytometry%2FMethods+and+development&aid=1452

Prior to construction, areas of ‘higher grade’ and ‘lower grade’ tumour were marked on the corresponding H&E slide. These areas were used as a guide to select areas for sampling. Tissue samples were taken from the same areas of tumour which were sampled by LCM. A recipient block was prepared by pouring melted paraffin into a mould of 7 mm depth, cooling it on a cold plate and facing the block with a microtome to ensure a flat surface. The empty block was placed into the holder and secured with clamping screws. Using the recipient punch, a core of blank wax was removed from the recipient block. A tissue core of equal depth was removed from the donor block using the donor punch and inserted into the empty core in the recipient block. The adjustment knobs were used to space each core 1.5 mm apart. Tissue samples from the same area of tumour were aligned next to each other. Internal control tissues were interspersed throughout the recipient block and were used as reference markers. Care was taken to ensure that tissue sampling did not compromise the volume of residual tumour for potential future diagnostic purposes.

Following completion of an entire recipient block, the block was placed upside down on a glass slide in a 60°C oven for 10-15 min, or until the top layer of paraffin was warmed and slightly softened. This promoted adherence of the tissue cores to the walls of the holes in the TMA block. The block was then levelled to ensure that the maximum number of sections would include all cores (Figure 2.3). A ‘map’ of each TMA was constructed using Microsoft Excel to ensure correct identification of each tissue core when scoring the IHC using light microscopy. Where possible, 3 tissue cores were taken from both tumour areas.

Figure 2.3 Example of a completed TMA block.

2.3.3 Tissue Micro-Array Sectioning

Each TMA block was placed in Mollifex prior to sectioning to increase the ease and accuracy of cutting. A microtome was used to cut sections of 4 µm thickness, which were mounted on charged glass slides before baking for 2 h at 60°C.

2.3.4 Immunohistochemistry (IHC) on the Roche Ventana Benchmark Ultra Autostainer System

IHC is a technique by which antigens on the cell surface can be localised using specific antibodies. The site of this antigen-antibody reaction is demonstrated by direct labelling of the antibody or using a secondary labelling method. The site of interaction is marked by a chromogenic reaction, which can be visualised using a microscope. A research collaboration was established with the IHC laboratory in SJH, through which 6 immunostains were performed using the Benchmark Ultra IHC/ISH System (Roche, Ventana, Arizona, USA). Antibody optimization for CD133 is detailed in Section 2.3.4.

TMA sections were prepared as described in Section 2.3.2. A unique protocol was designed for each antibody, as determined by prior antibody optimization efforts (Table 2.20). Unique barcoded labels corresponding to each IHC protocol were printed and attached to each slide. The slides were inserted into the instrument along with the required reagents. The detection method used for all antibodies was the Optiview DAB IHC detection Kit (Ventana). This kit is used to detect mouse and rabbit primary antibodies which are bound to a specific antigen in FFPE tissue. This primary antibody is then located using a secondary antibody bound to an enzyme-labelled tertiary antibody, which in turn is visualised with hydrogen peroxide (H_2O_2) and 3, 3'-diaminobenzidine

tetrahydrochloride (DAB) (Figure 2.4). DAB produces a brown precipitate that is visualised on light microscopy.

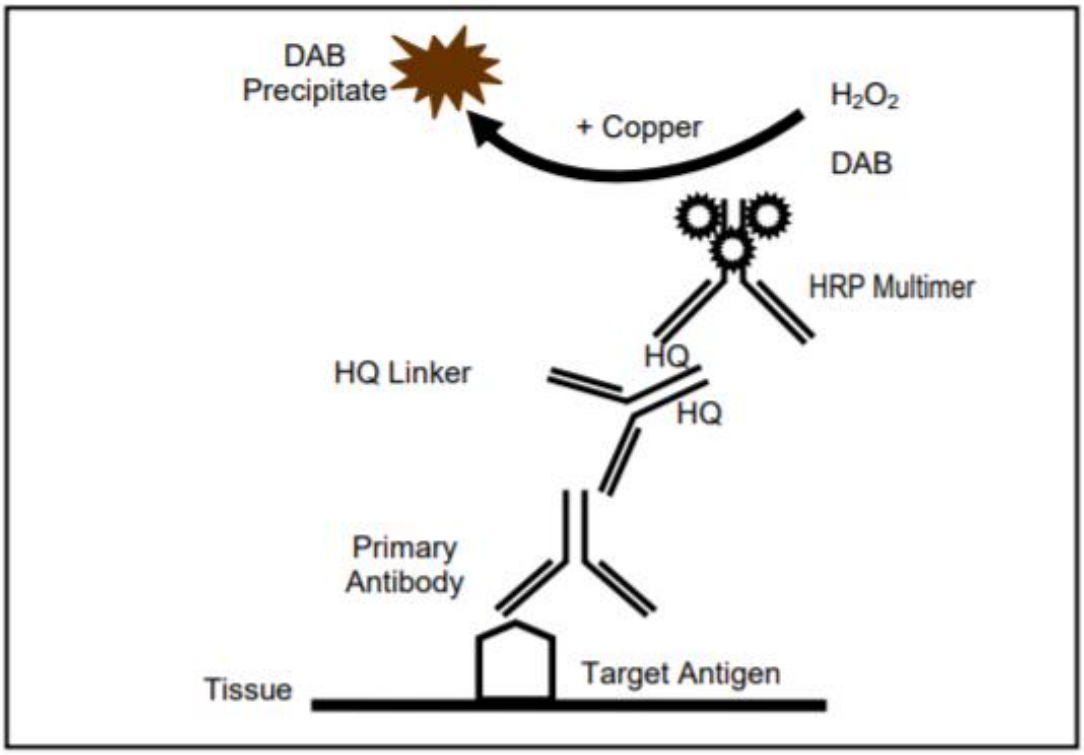


Figure 2.4 Mechanism of Antigen Detection Using the Optiview DAB Detection Kit. Adapted from the Optiview DAB detection kit product specification kit by Ventana.

All protocols included a pre-treatment step with cell conditioning solution 1 (CC1). Once the staining was complete, slides were rinsed in lukewarm water mixed with weak detergent before being placed in reverse osmosis deionized (RODI) H_2O. They were then washed 3 times in EtOH followed by 3 washes in Xylene, before cover slipping and microscopic analysis. Antibody-specific control tissues were stained in addition to the TMA sections for validation purposes (Table 2.21).

Table 2.20: Primary antibodies.

Primary Antibody Name	Pre-Treatment	Primary Antibody Incubation Time	Company	Reference Code
Anti-Vimentin (V9) Mouse mAb	CC1 48 min	16 min	Cell Marque	790-2917
Anti-CD34 (QBEnd/10) Mouse mAb	CC1 16 min	16 min	Ventana	790-2927
Anti-Ep-cam (Ber-EP4) Mouse mAb	CC1 32 min	15 min	Cell Marque	CMC43830050
Anti-Oct-4 (MRQ-10) Mouse mAb	CC1 32 min	16 min	Cell Marque	CMC30921040
Anti-E-cadherin (36) Mouse mAb	CC1 40 min	12 min	Ventana	790-4497

Table 2.21: Primary antibody control tissues.

Primary Antibody Name	Control Tissue
Anti-Vimentin (V9) Mouse mAb	Colon
Anti-CD34 (QBEnd/10) Mouse mAb	Colon
Anti-EpCAM (Ber-EP4) Mouse mAb	Colonic adenocarcinoma
Anti-Oct-4 (MRQ-10) Mouse mAb	Seminoma
Anti-E-cadherin (36) Mouse mAb	Breast DCIS
Anti-CD133 Rabbit mAb	Kidney

2.3.5 Antibody Optimization

Manual antibody optimization was performed for two primary antibodies: CD133 and ALDH1A1. Optimization was performed using known positive and negative control tissues for each antibody (Table 2.22). The antibody protocols provided by Cell Signalling Technology were followed and adapted as required.

Table 2.22: Primary antibody positive and negative control tissues.

Primary Antibody Name	Positive Control Tissue	Negative Control Tissue
Anti-CD133 Rabbit mAb	Kidney	Oesophagus
Anti-ALDH1A1 Rabbit mAb	Stomach	Appendix

Deparaffinization and Rehydration

Four µm sections from a block containing antibody specific positive and negative control tissues were cut using a microtome, mounted on charged glass slides and baked for 2 h at 60°C. The sections were deparaffinized and rehydrated by washing 3 times in xylene for 5 min each, followed by two 10 min washes in 100% EtOH, two 10 min washes in 95% ethanol and finally two 5 min washes in RODI H_2O.

Antigen Retrieval

The slides were submerged in 1X citrate unmasking solution (Cell Signalling Technology, MA, USA) until boiling, followed by 10 min at sub-boiling temperatures (95-98°C) before cooling on the bench at room temperature for 30 min.

Peroxide Block

Sections were washed in RODI water three times for 5 min each, before incubating in 3% H_2O_2 for 10 min. The sections were then washed twice in RODI H_2O for 5 min, followed by washing in Tris Buffered Saline with Tween (TBST) (Cell Signalling Technology) for 5 min.

Protein Block

The tissue on the slide was outlined with a Dako hydrophobic pen (Agilent, CA, USA) before blocking with 100-400 µL 1X animal free blocking solution (Cell Signalling Technology) for 1 h at room temperature.

Primary Antibody Incubation

The blocking solution was removed and 100-400 µL primary antibody diluted in SignalStain antibody diluent (Cell Signalling Technology) was applied to each section as required (Table 2.23). Rabbit Monoclonal Negative Control (Ventana) was applied to one section of control tissue in place of a primary antibody for validation purposes. These sections were incubated with the primary antibody or negative control overnight at 4°C.

Secondary Antibody Detection and Visualization

The following day the antibody solution was removed, and the sections were washed in TBST three times for 5 min. Each section was covered with 1-3 drops of SignalStain boost detection reagent (Cell Signalling Technology) and incubated in a humidified chamber for 30 min at room temperature, before three 5 min washes in TBST. Next, 100-400 µL SignalStain DAB (Cell Signalling Technology) was applied to each section, and closely monitored for 1-10 minutes until the chromogenic reaction became visible.

Counterstaining, Dehydration and Cover Slipping

The slides were then immersed in RODI water and counterstained for 30 sec in haematoxylin followed by two subsequent 5 min washes in RODI water. Finally, the slides were dehydrated by incubating twice in 95% EtOH for 10 sec, then twice in 100% EtOH for 10 sec, then twice in xylene for 10 sec. The sections were mounted, and cover slipped using Vectamount (Vector Laboratories, CA, USA).

Table 2.23 Antibody optimization.

	CD133			ALDH1A1	
Variable	**Dilution**	**Antigen Retrieval**	**Antibody Incubation**	**Dilution**	**Antibody Incubation**
Optimization Run 1	1:300 1:400 1:500	Heat in 1X citrate unmasking solution until boiling, followed by 10 min at 95-98°C	Overnight at 4°C	1:100 **1:200** 1:300	Overnight at 4°C
Optimization Run 2	1:400		1 h at room temp		
Optimization Run 3	**1:100** 1:150 1:200	CC1 32 min at 100°C **CC1 32 min at 100°C** CC1 64 min at 100°C	16 min at 36°C **32 min at 36°C** 16 min at 36°C		

ALDH1A1 staining was optimized at 1:200 dilution with antibody incubation overnight at 4°C. A chromogenic response with SignalStain DAB was observed after approx. 2 min. CD133 optimization was achieved at 1:100 dilution, CC1 32 min and antibody incubation time 32 min (Table 2.22). Optimization runs 1 and 2 were performed manually; run 3 was conducted on the Benchmark Ultra IHC/ISH System using programmed protocols. In run 3 each antibody dilution was trialled under the three different pre-treatment/antibody incubation settings listed in Table 2.23: a total of 9 optimization tests. The Optiview DAB IHC detection Kit reagents described in Section 2.3.4 were used in run 3 instead of the

reagents in the protocol detailed earlier in this section. Following optimization, TMA sections were cut as described in Section 2.3.3 and stained with ALDH1A1 and CD133 according to their optimized protocols.

2.3.6 TMA Scoring

Once stained and cover slipped, all H&E and IHC TMA sections were scanned using the NanoZoomer 2.0RS digital scanner and NanoZoomer Digital Pathology (NDP) scanning software (Hamamatsu, Japan). The slides were viewed using the NDP.view2 viewing software (Hamamatsu). TMA cores were identified and marked on the system, directed by the TMA maps compiled at the time of TMA construction (Section 2.3.2). A semi-quantitative scoring system was used for each IHC marker, as described in Chapter 3, and each IHC was independently assessed by two Histopathologists, with discordant cases co-reviewed to reach consensus.

2.4 Statistical Analysis

The data collected was checked for consistency and completeness before being entered into the database software. IBM SPSS Statistics Version 25 was used for univariate analyses. Data summarization and analysis was performed using descriptive statistics. Paired samples were compared using a Wilcoxon matched pairs test. Analyses of nominal and ordinal data were performed using Mann-Whitney U and Kruskal-Wallis tests. Continuous data was analysed using the Pearson correlation coefficient and Spearman's rank correlation.

Multivariate analyses were performed using programming software R. Survival outcomes were analysed using a Cox proportional hazard model and Kaplan Meier curves. A predictive analysis was performed using a random forest model to incorporate all expression data.

Chapter Three: Expression Profiling in Histologically Defined Areas of Gastroesophageal Junction Adenocarcinoma

3.1 Study Population

The patient cohort for this study initially comprised 80 formalin fixed (10% buffered formalin) paraffin embedded (FFPE) gastrectomy and oesophagectomy specimens excised as per surgical management of gastroesophageal junction adenocarcinoma (GEJA) – as defined endoscopically as Siewert stage 1-3 – in SJH between 2011-2018. One case was excluded from the study cohort due to an insufficient sample (Figure 3.1). Samples were included in the study irrespective of whether they received NAT (87.3% of patients (69/79)) or not. Each sample contained viable tumour cells and was at least stage (y)pT1b.

Histology slides from each case were reviewed to select two separate areas within the same tumour: one sample was deemed 'lower grade', as defined by better histological differentiation and/or less infiltrative portion of tumour; and the other sample 'higher grade', defined as poor histological differentiation and/or location within the infiltrative or pushing tumour edge (Figure 3.2). The infiltrative tumour edge was chosen for 'higher grade' samples because these cells are most likely to undergo EMT. The post-operative pathology reports and clinical charts were reviewed to collect relevant clinicopathological data for each patient. The clinical characteristics of the final 79 patient samples used in this study are listed in Table 3.1.

Paired samples were taken from 58 cases. Due to a variability in response to NAT not all tumours had two histologically distinct areas of tumour: 21 cases, all of which were post-NAT, had an unpaired tumour sample taken which was classified as 'higher grade' due to its poor differentiation status. Three cases – 2 paired and 1 unpaired – were inadequately captured using LCM extraction of tumour cells but were successfully sampled within the study TMAs. These cases were therefore only included in the protein

analysis portion of the study. Accordingly, 56 paired and 20 unpaired cases were used in the mRNA and miRNA analyses.

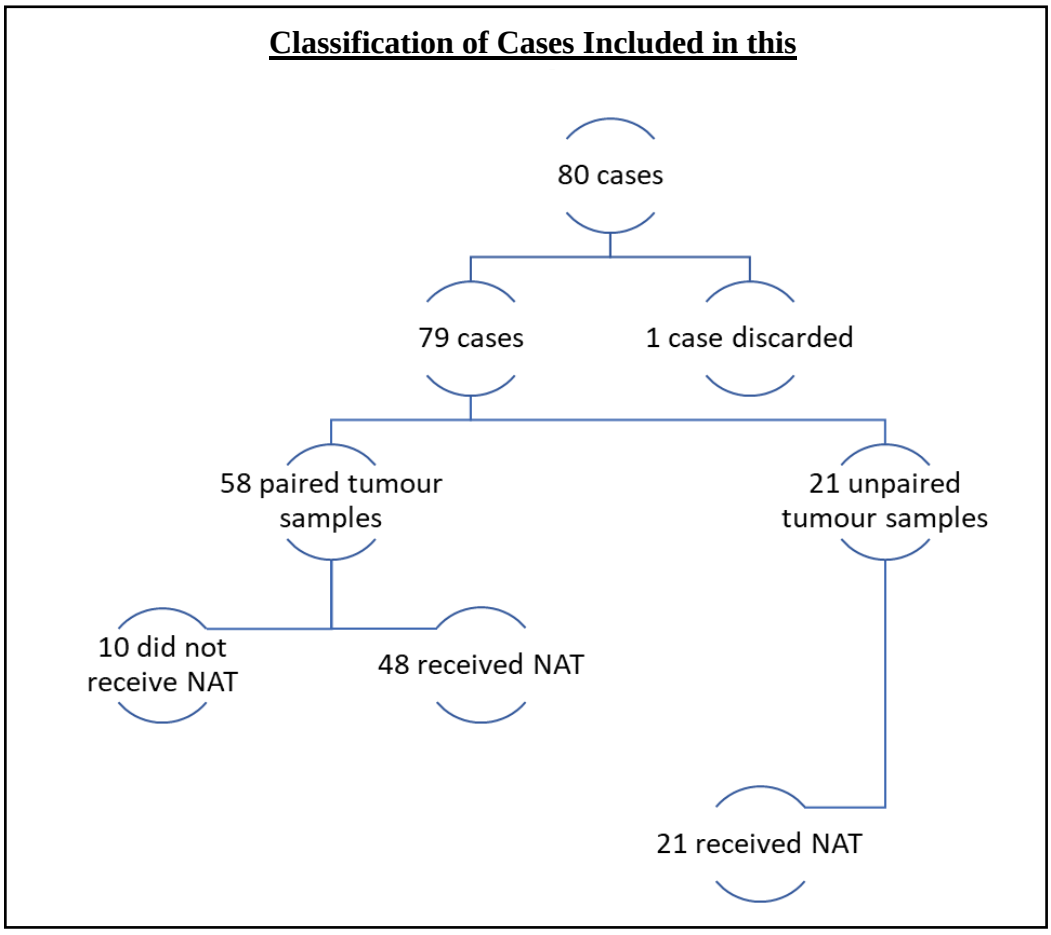


Figure 3.1 Flow chart depicting the classification of tumour samples included in the study.

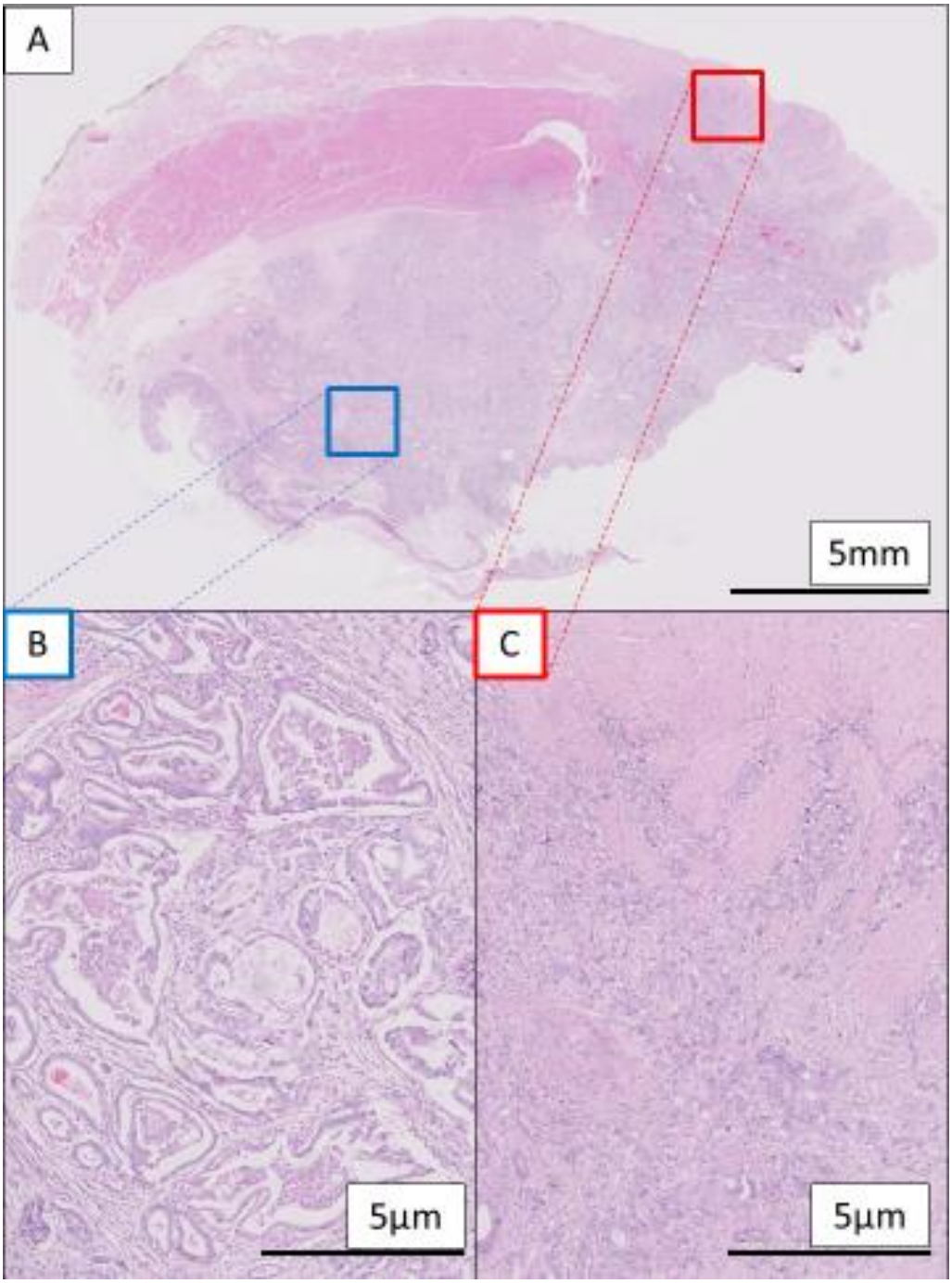


Figure 3.2: Sampling of histologically defined areas of tumour. Image A is a representative section of tumour demonstrating areas sampled for LCM and construction of TMA (0.61X magnification). Image B (blue square) highlights better differentiated and less infiltrative area of tumour, categorized as 'lower grade' throughout the study (5X magnification). Image C (red square) highlights poorly differentiated tumour located at the infiltrative edge, categorized as an 'higher grade' area of tumour throughout the study (5X magnification). Sections were stained with H&E.

Table 3.1 Clinicopathological Characteristics of Study Population.

Patient Characteristic	***n* (%)**
Age at Diagnosis	Range 38-84 years Average 66 years Median 67 years
Sex Male Female	 69 (87.3) 10 (12.7)
Disease Recurrence Yes No	 45 (57) 34 (43)
Disease Free Survival Post Primary Surgery	Range 45-1347 days Average 378 days Median 216 days
Overall Survival	Range 150-1575 days Average 614 days Median 510 days
Macroscopic Tumour Size	Range 8-145 mm Average 41.7 mm Median 40 mm
Lymphovascular Invasion Yes No	 56 (70.9) 23 (29.1)
Perineural Invasion Yes No	 13 (16.5) 66 (83.5)
Serosal Involvement Yes No	 24 (30.4) 55 (69.6)
Neoadjuvant Therapy Yes No	 69 (87.3) 26 MAGIC 18 CROSS 3 EOX 2 FLOT 20 Other 10 (12.7)
Adjuvant Therapy Yes No	 29 (36.7) 50 (63.3)

Table 3.1 Clinicopathological Characteristics of Study Population.

Patient Characteristic	***n* (%)**
Pathological T Stage	
(y)pT1	0 (0)
(y)pT2	12 (15.2)
(y)pT3	58 (73.4)
(y)pT4a	8 (10.1)
(y)pT4b	1 (1.3)
Pathological N Stage	
(y)pN0	23 (29.1)
(y)pN1	18 (22.8)
(y)pN2	21 (26.6)
(y)pN3	12 (15.2)
(y)pN3a	5 (6.3)
(y)pN3b	0 (0)
Pathological M Stage	
(y)pMX	75 (94.9)
(y)pM0	1 (1.3)
(y)pM1	3 (3.8)
Siewert Classification	
1	24 (30.4)
2	26 (32.9)
3	26 (32.9)
Unknown	3 (3.8)
Tumour Regression Grade	
1	0 (0)
2	3 (3.8)
3	16 (20.3)
4	34 (43)
5	14 (17.7)
Unknown	2 (2.5)
Not Applicable	10 (12.7)
HER2 Status	
Positive	9 (11.4)
Negative	31 (39.2)
Unknown	39 (49.4)
Resected Tumour Differentiation by Predominant Area	
Well Differentiated	5 (6.3)
Moderately Differentiated	26 (32.9)
Poorly Differentiated	47 (59.5)
Not Applicable (Mucinous)	1 (1.3)

3.2 Gene Expression Analysis

3.2.1 Validation of Endogenous Control Gene for mRNA Expression Analysis

A review of the literature was performed to identify the most commonly used endogenous control genes in mRNA expression analyses of carcinoma specimens. No information was found regarding a specific housekeeper gene for GEJA. Glyceraldehyde 3-Phosphate Dehydrogenase (GAPDH) was identified as a robust housekeeper across a range of human malignancies (ThermoFisher, 2019). A previous study performed in our institution used Cyclin Dependent Kinase Inhibitor 1B (CDKN1B) as an endogenous control for qPCR studies to great effect (Denning et al., 2007). Accordingly, both GAPDH and CDKN1B were tested as candidate endogenous control genes.

Ten tumour samples, which included representative samples from both 'lower grade' and 'higher grade' categories, were trialled with CDKN1B and GAPDH in addition to the target genes for the study: Cadherin-1 (CDH1), NANOG, POU Class 5 Homeobox 1 (POU5F1), Vimentin and Serpine-1 (Methods Section 2.1.6). The raw cycle threshold (C_T) average was compared across the samples to demonstrate the variability in gene expression, both within and between samples. CDKN1B expression was highly variable between samples, with no expression observed in 9 out of 10 cases. By comparison, GAPDH was expressed across all samples, with comparable C_T values (Figure 3.3). Therefore, GAPDH was chosen as the endogenous control gene for this study.

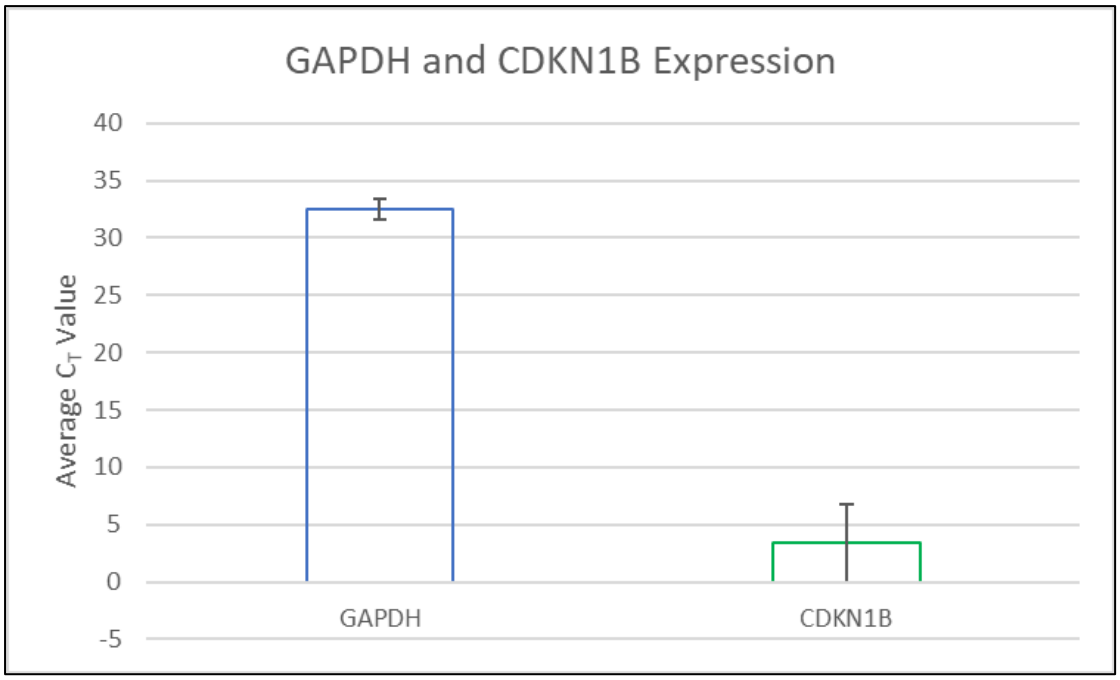


Figure 3.3 Amplification plot and box and whisker plot showing expression of GAPDH and CDKN1B across ten RNA samples. Data graphed as mean +/- SEM.

3.2.2 Differential Expression of Genes in Paired Samples

qPCR was used to examine the expression of a panel of genes in GEJA, which are associated with EMT and stemness. Samples were normalized to GAPDH to correct for systematic variables. One pair failed QC and was therefore excluded, thus only 55 pairs were analysed. Where present, the direction of change in expression was recorded from 'lower grade' to 'higher grade' samples (Appendix 1.1). Pairwise analysis was performed for each gene using a Wilcoxon matched pair's test. No significant change in expression was observed in any gene (Figures 3.5 and 3.6).

Table 3.2 demonstrates the patterns of gene expression in paired samples. Vimentin and CDH1 were expressed in only a minority of paired samples. Vimentin was expressed in 21 pairs, 38.1% (8/21) of which showed an increase in expression. CDH1 was expressed in only 8 pairs, of which 87.5% (7/8) showed a decrease in expression. The expression patterns of NANOG and POU5F1 were comparable within paired samples: NANOG was expressed in 53 pairs, with upregulation in expression seen in 56.6% (30/53), whilst POU5F1 was expressed in 55 pairs, of which 56.4% (31/55) showed an increase in expression. Serpine-1 expression was seen in only 2 pairs: 1 increased and 1 decreased. Figure 3.4 depicts the degree of fold change between paired samples. Fold change values for each case are available in Appendix 1.2. Of note, fold change could not be calculated for pairs with expression recorded in only one sample.

Table 3.2 Patterns of Gene Expression in Paired Samples.

	No Expression in Both Samples	Increase in Expression	Decrease in Expression	One or Both Samples Not Analysed
Vimentin	34	8	13	3
CDH1	47	1	7	3
NANOG	2	30	23	3
POU5F1	0	31	24	3
Serpine-1	49	1	1	7

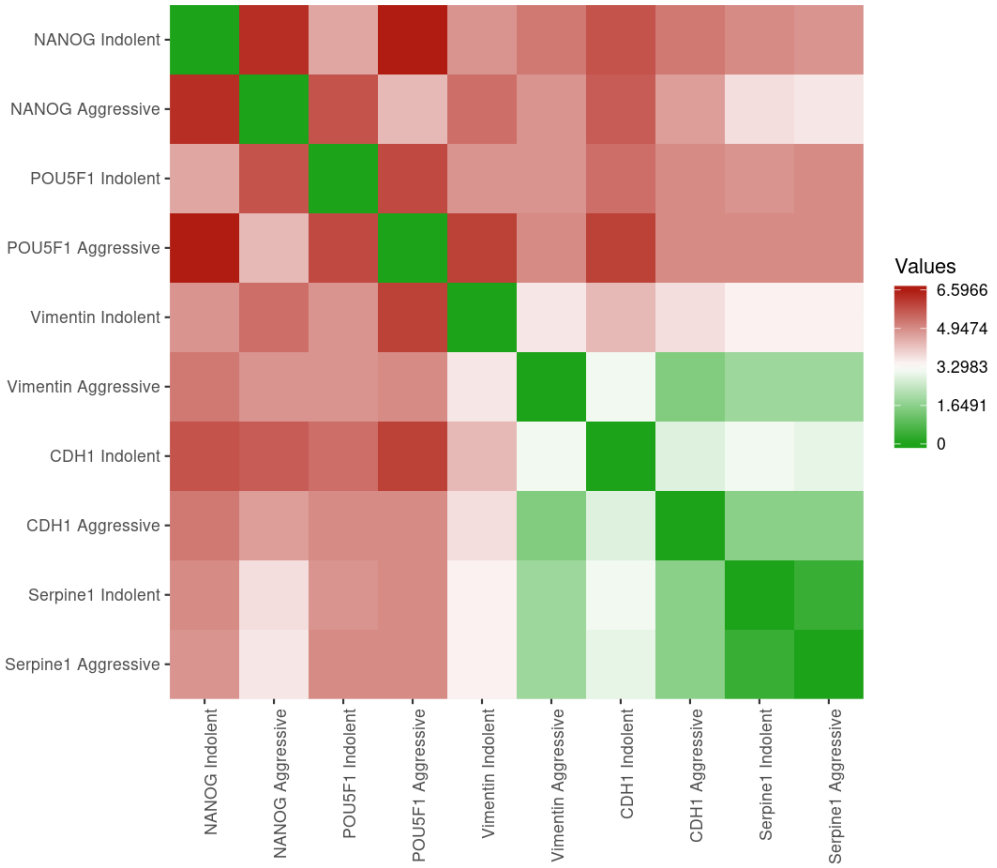
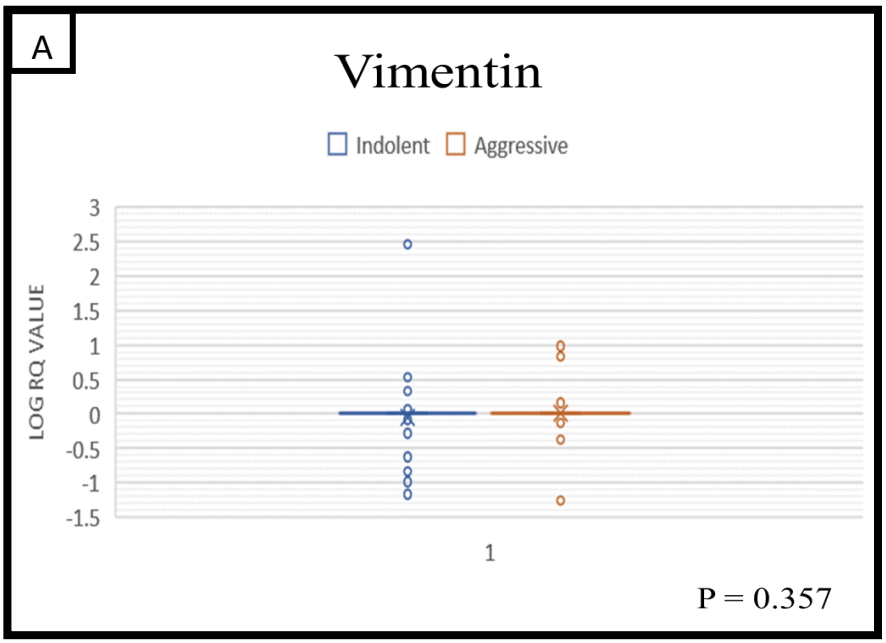


Figure 3.4 Heatmap showing the fold change in gene expression between paired 'lower grade' and 'higher grade' samples. Data is expressed as log RQ values in a distance matrix. Image generated using the online web tool 'Heatmapper'.



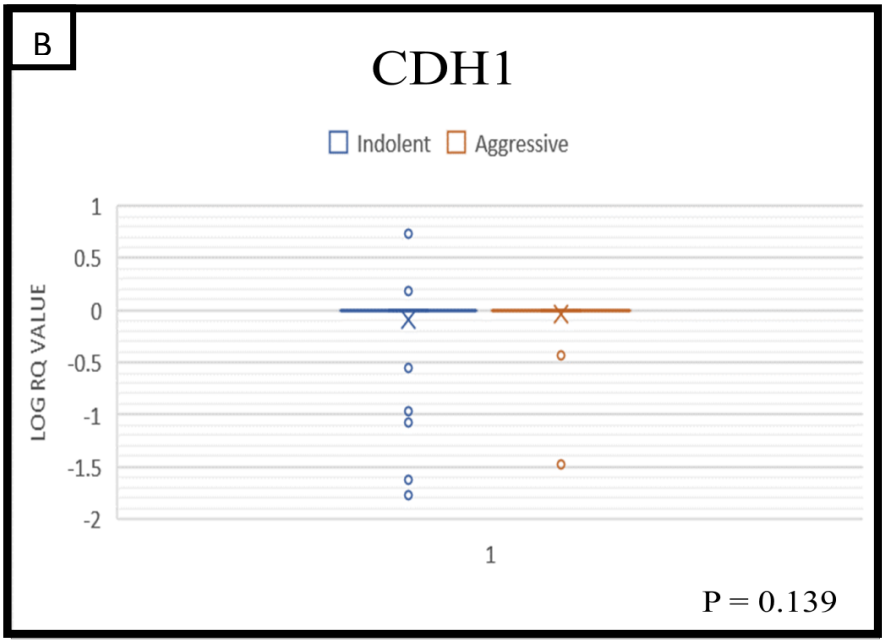


Figure 3.5 Box and whisker plots showing the difference in log RQ values of gene expression associated with EMT between paired lower grade and higher grade areas of tumour. A = Vimentin, B = CDH1. No statistically significant differential expression was observed in either gene. Statistical significance: Wilcoxon matched pairs test, $p<0.05$. Data graphed as mean, median, range and interquartile range. X represents the mean. The central horizontal line represents the median. Where present, the top line represents the first quartile and the bottom line represents the third quartile. The whiskers denote the maximum and minimum values.

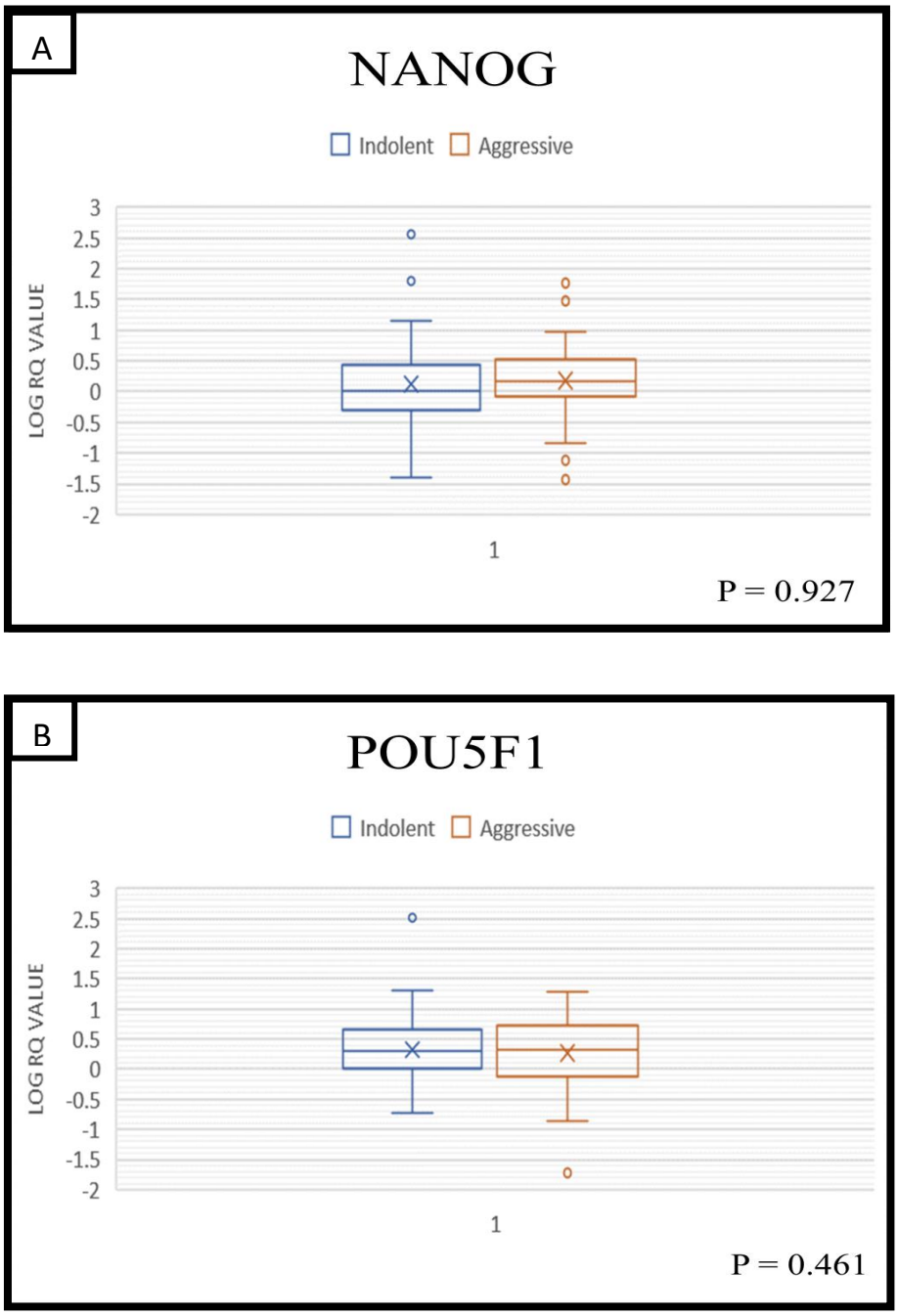


Figure 3.6 Box and whisker plots showing the difference in log RQ values of gene expression associated with stemness between paired lower grade and higher grade areas of tumour. A = NANOG, B = POU5F1. No statistically significant differential expression was observed in either gene. Statistical significance: Wilcoxon matched pairs test, $p<0.05$. Data graphed as mean, median, range and interquartile range. X represents the mean. The central horizontal line represents the median. Where present, the top line represents the first quartile and the bottom line represents the third quartile. The whiskers denote the maximum and minimum values.

3.2.3 Association Between Clinicopathological Characteristics and Gene Expression in 'Higher Grade' Tumour Samples

'Higher grade' samples were analysed from each patient to determine the presence of any significant correlation between gene expression patterns and clinicopathological characteristics. Statistical analyses of nominal and ordinal data were performed using Mann-Whitney U and Kruskal-Wallis tests. Continuous data was analysed using the Pearson correlation coefficient and Spearman's rank correlation.

Vimentin expression was shown to decrease with increasing tumour regression grade ($p=0.023$, Table 3.3). Its expression was also shown to have a significant lack of association with age at diagnosis ($p=0.029$, $r= -0.257$, Table 3.5). No other gene in this panel showed a significant association with any clinicopathological feature. A strong association was identified between NANOG expression levels and tumour involvement of the serosa ($p=0.096$) and lymph node metastases ($p=0.093$) (Table 3.4).

Table 3.3 Correlation Between Clinicopathological Characteristics and Gene Expression Associated with EMT and Metastasis.

	Vimentin			CDH1			Serpine-1		
	Cases (*n*)	**Median (IQR)****	***p***	**Cases (*n*)**	**Median (IQR)****	***p***	**Cases (*n*)**	**Median (IQR)****	***p***
Disease Recurrence									
Yes	42	0 (0-0.064)	0.569	42	0 (0-0)	0.731	38	0 (0-0)	0.864
No	30	0 (0-0.033)		30	0 (0-0)		29	0 (0-0)	
TRG*									
2	2	2.783 (2.621-2.944)	**0.023**	2	0 (0-0)	0.668	2	0 (0-0)	0.788
3	15	0 (0-0.808)		15	0 (0-0)		14	0 (0-0)	
4	30	0 (0-0)		30	0 (0-0)		29	0 (0-0)	
5	13	0 (0-0)		13	0 (0-0)		13	0 (0-0)	
LVI									
Yes	50	0 (0-0)	0.120	50	0 (0-0)	0.175	48	0 (0-0)	0.370
No	22	0 (0-0.94625)		22	0 (0-0)		19	0 (0-0)	
PNI									
Yes	10	0 (0-0.0248)	1.000	10	0 (0-0)	0.499	10	0 (0-0)	0.170
No	62	0 (0-0.064)		62	0 (0-0)		57	0 (0-0)	
Serosal Involvement									
Yes	22	0 (0-0)	0.337	22	0 (0-0)	0.841	21	0 (0-0)	0.336
No	50	0 (0-0.280)		50	0 (0-0)		46	0 (0-0)	
HER2 Status									
Positive	8	0 (0-0)	0.121	8	0 (0-0)	0.902	7	0 (0-0)	0.888
Negative	30	0 (0-0.656)		30	0 (0-0)		28	0 (0-0)	
Siewert Classification			0.189			0.343			0.564
I	21	0 (0-0)		21	0 (0-0)		21	0 (0-0)	
II	23	0 (0-0)		23	0 (0-0)		23	0 (0-0)	
III	25	0 (0-0.735)		25	0 (0-0)		20	0 (0-0)	
T Stage									
T0	0	0 (0-0)	0.745	0	0 (0-0)	0.525	0	0 (0-0)	0.727
Tis, T1-2	10	0 (0-0.033)		10	0 (0-0)		10	0 (0-0)	
T3-4	62	0 (0-0.064)		62	0 (0-0)		57	0 (0-0)	
N Stage									
N0	21	0 (0-0.017)	0.284	21	0 (0-0)	0.839	20	0 (0-0)	0.353
N+	51	0 (0-0.417)		51	0 (0-0)		47	0 (0-0)	

* Cases with TRG 1 were excluded from the study as they fall under the category of (y)pT0. ** Interquartile Range (Quartile 1 – Quartile 3).

Table 3.4 Correlation Between Clinicopathological Characteristics and Gene Expression Associated with Stemness.

	NANOG			POU5F1		
	Cases (*n*)	**Median (IQR)****	***p***	**Cases (*n*)**	**Median (IQR)****	***p***
Disease Recurrence						
Yes	42	0.972 (0.399-2.784)	0.163	42	1.458 (0.388-4.212)	0.465
No	30	1.861 (0.711-4.021)		30	1.898 (0.965-4.332)	
TRG*						
2	2	0.641 (0.350-0.931)	0.142	2	0.854 (0.673-1.034)	0.304
3	15	1 (0.319-3.369)		15	1.318 (0.491-4.01)	
4	30	0.594 (0.187-2.495)		30	1.861 (0.41-3.877)	
5	13	2.669 (1.516-2.966)		13	3.234 (0.953-5.503)	
LVI						
Yes	50	1.336 (0.462-2.885)	0.961	50	1.861 (0.61-4.273)	0.980
No	22	1.763 (0.436-2.701)		22	1.705 (0.777-4.271)	
PNI						
Yes	10	2.472 (0.594-4.022)	0.454	10	2.351 (1.334-3.445)	0.416
No	62	1.336 (0.436-2.842)		62	1.599 (0.521-4.359)	
Serosal Involvement						
Yes	22	1.878 (0.731-3.247)	0.096	22	2.107 (0.69-5.958)	0.549
No	50	1.111 (0.236-2.751)		50	1.705 (0.61-4.130)	
HER2 Status						
Positive	8	0.95 (0.476-2.545)	0.661	8	2.357 (0.3-4.337)	0.902
Negative	30	0.788 (0.158-2.774)		30	0.9 (0.422-4.201)	
Siewert Classification						
I	21	1.939 (0.858-2.87)	0.804	21	2.07 (1-4.3)	0.349
II	23	1.516 (0.504-2.762)		23	1.745 (0.886-4.27)	
III	25	0.832 (0.322-2.966)		25	1.256 (0.311-3.948)	
T Stage						
T0	0	0 (0-0)	0.803	0	0 (0-0)	0.782
Tis, T1-2	10	1.116 (0.520-2.573)		10	1.831 (0.87-3.858)	
T3-4	62	1.506 (0.436-2.947)		62	1.688 (0.633-4.273)	
N Stage						
N0	21	2.263 (0.832-4.519)	0.093	21	1.917 (0.605-4.191)	0.926
N+	51	1.221 (0.405-2.801)		51	1.384 (0.641-4.377)	

* Cases with TRG 1 were excluded from the study as they fall under the category of (y)pT0. ** Interquartile Range (Quartile 1 – Quartile 3).

Table 3.5 Correlation Between Age, Tumour Size and Gene Expression.

	NANOG	POU5F1	Vimentin	CDH1	Serpine1
Age at Diagnosis					
Pearson correlation (*r*)	0.006	-0.062	-0.257	-0.183	0.040
Number (*n*)	72	72	72	72	67
Significance (*p*)	0.960	0.608	**0.029**	0.123	0.745
Macroscopic Tumour Size					
Spearman correlation (ρ)	0.078	0.049	-0.081	-0.103	-0.081
Number (*n*)	65	65	65	65	60
Significance (*p*)	0.535	0.698	0.520	0.412	0.538

3.3 MicroRNA Expression Analysis

3.3.1 Determining a Method of Normalization for miRNA Expression Analysis

There is no consensus as to the most reliable normalization strategy for use in quantification of miRNA transcripts. Reference small nucleolar RNAs (snRNA) such as U6 and 5S have been commonly used as normalizers until recently, yet they were not chosen for this study due to a growing body of evidence suggesting that miRNAs should be normalized using the same class of biomarker (Peltier & Latham, 2008; Vandesompele et al., 2002). Furthermore, these snRNAs were unavailable for use with the TaqMan® Advanced miRNA Assays.

As directed by the technical guide for TaqMan® Advanced miRNA Assays and a thorough search of the published literature, a panel of 12 candidate miRNAs was chosen, of which 11 were known to play a role in regulating EMT and/or stemness. This list included 3 miRNAs which have been shown to be consistently expressed across a range of human tissues, with the assumption that one or all would be consistently expressed across all tumour samples: hsa-miR-16-5p, hsa-miR-103a-3p and hsa-miR-17-5p (Landgraf et al., 2007; Peltier & Latham, 2008).

Following qPCR analysis, none of the miRNAs in the chosen panel were expressed in all tumour samples. Therefore, alternative options for data normalization were considered: to use GAPDH expression data from the mRNA analysis as a normalizer, given that the same RNA samples were analysed in both studies; or to employ global normalization, in which the median C_T values of miRNA assays are used as a normalization factor on an individual sample basis.

GAPDH was deemed an inappropriate normalizer for miRNAs because it is in a different biomarker class. Additionally, the expression levels of GAPDH were not truly comparable between the mRNA and miRNA studies because the methods of generating cDNA differ in each. Thus, global normalization was chosen as the optimal means of normalization. Whilst it is preferable to use global normalization in studies which analyse large numbers of miRNAs, previous studies have used this method in the absence of a stable reference gene (Faraldi et al., 2019).

3.3.2 Differential Expression of miRNAs in Paired Samples

qPCR was used to investigate the expression of miRNAs in GEJA, which are known to be upregulated or downregulated in association with EMT and/or acquisition of stemness traits across a range of epithelial malignancies. Five miRNAs were expected to be upregulated in 'higher grade' samples: miR-10b, miR-21, miR-221 miR-223 and miR-224 (Chai et al., 2019; Z. Chen et al., 2013; Knoll et al., 2014; B. Liu et al., 2018; Ma et al., 2015; Sheedy & Medarova, 2018). By comparison, 6 miRNAs were expected to be downregulated in the same samples: miR-16, miR-17, miR-203a, miR-200a, miR-133b and miR-141 (Z. Chen et al., 2013; L. L. Fang et al., 2017; Lynam-Lennon et al., 2017; Taube et al., 2013; T. Wang et al., 2017). MiR-103a was included in the study as a potential internal control.

Global normalization was used to normalize all samples. A total of 56 paired samples were suitable for miRNA expression analysis. No expression was detected in any sample for miR-103a, miR-203a and miR-133b, thus they were excluded from the following analysis. Where present, the direction of change in expression was recorded from 'lower

grade' to 'higher grade' samples (Appendix 1.3). Pairwise analysis was performed for each miRNA using a Wilcoxon matched pairs test.

miR-221 was shown to be significantly upregulated in 'higher grade' samples compared to 'lower grade' ($p=0.002$). No other miRNA in this panel showed a significant change in expression within paired samples (Figure 3.8). miR-221 was expressed in 36 paired samples, of which 66.7% (24/36) were upregulated. miR-224 was expressed in 37 pairs and was upregulated in 54.1% (20/37), whilst miR-16 was downregulated in 57.7% of pairs with differential expression (15/26). miR-10b, miR-21, miR-223 and miR-17 were each upregulated and downregulated in equal proportion within pairs with differential expression. miR-141 and miR-200a were both upregulated in the pairs which had differential expression: miR-141 increased in 52.5% of cases (21/40) and miR-200a increased in 53.3% (24/45) (Table 3.6). Figure 3.7 depicts the degree of fold change between paired samples. Fold change values for each case are available in Appendix 1.4. Of note, fold change could not be calculated for pairs with expression recorded in only one sample.

Table 3.6 Patterns of miRNA Expression in Paired Samples.

	No Expression in Both Samples	Equal Positive Expression	Increase in Expression	Decrease in Expression	One or Both Samples Not Analysed
miR-10b	45	0	5	5	3
miR-21	9	1	23	23	2
miR-223	48	0	4	4	2
miR-224	19	0	20	17	2
miR-221	19	1	24	12	2
miR-16	30	0	11	15	2
miR-17	36	0	10	10	2
miR-141	14	0	21	19	4
miR-200a	10	1	24	21	2

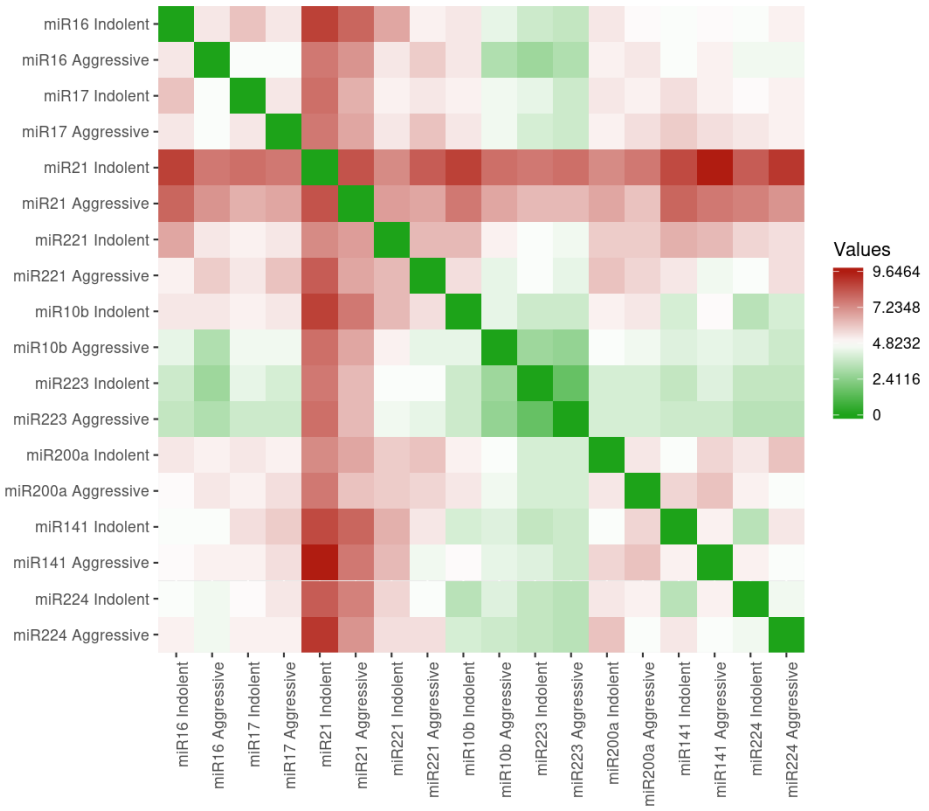
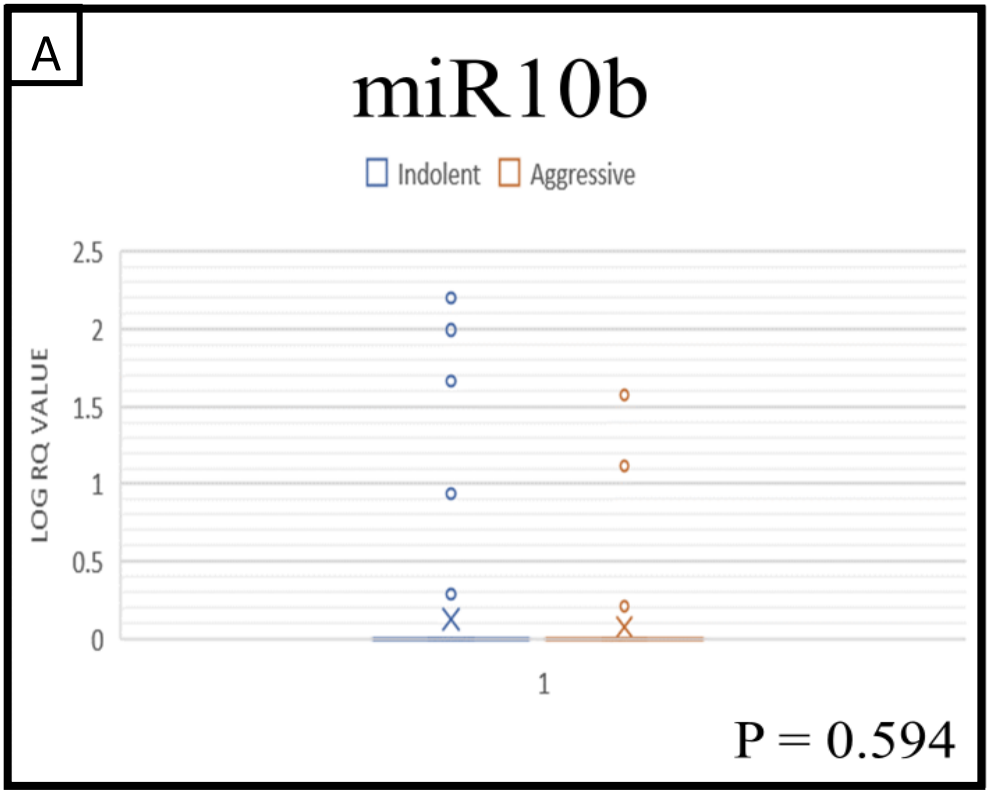
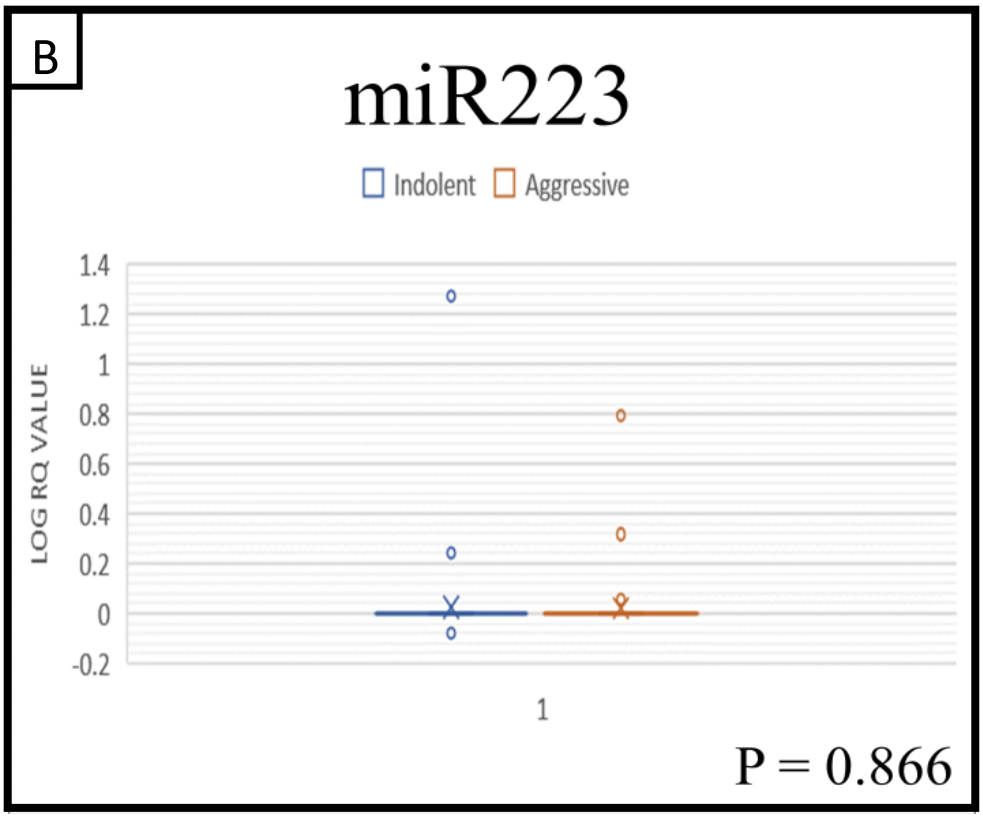
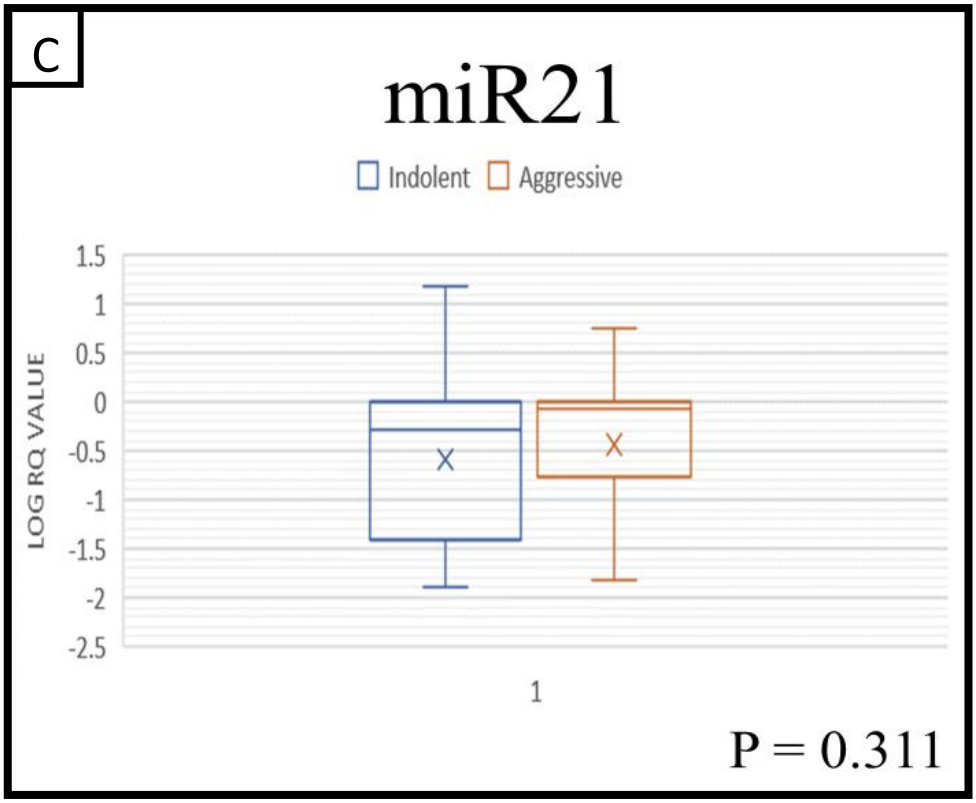
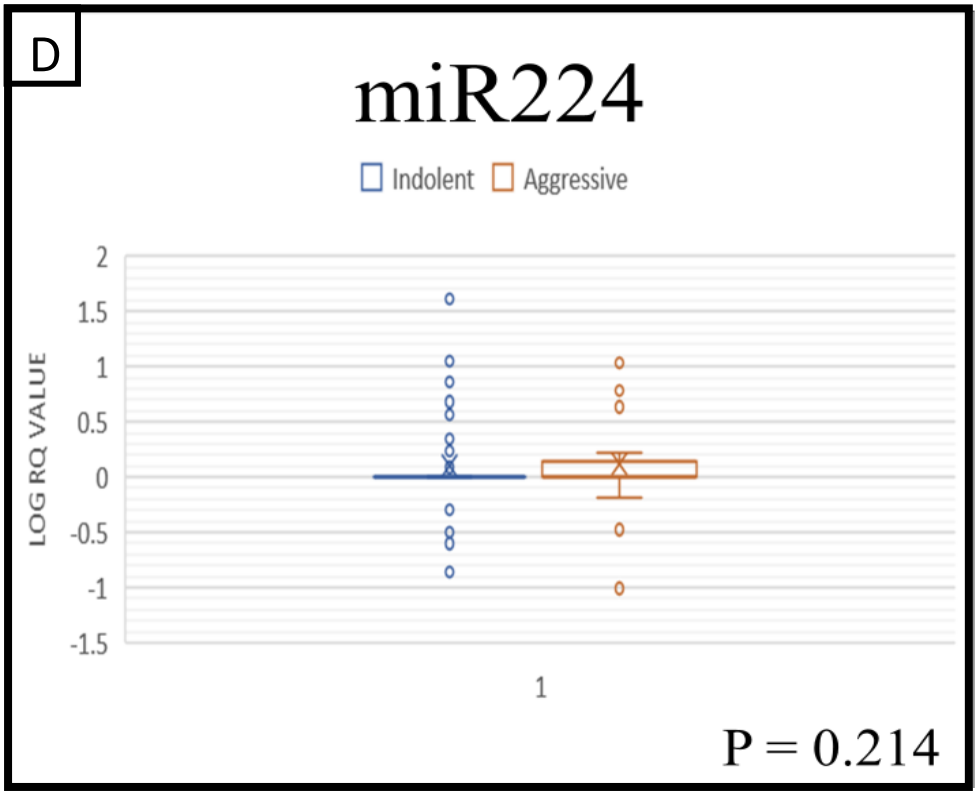


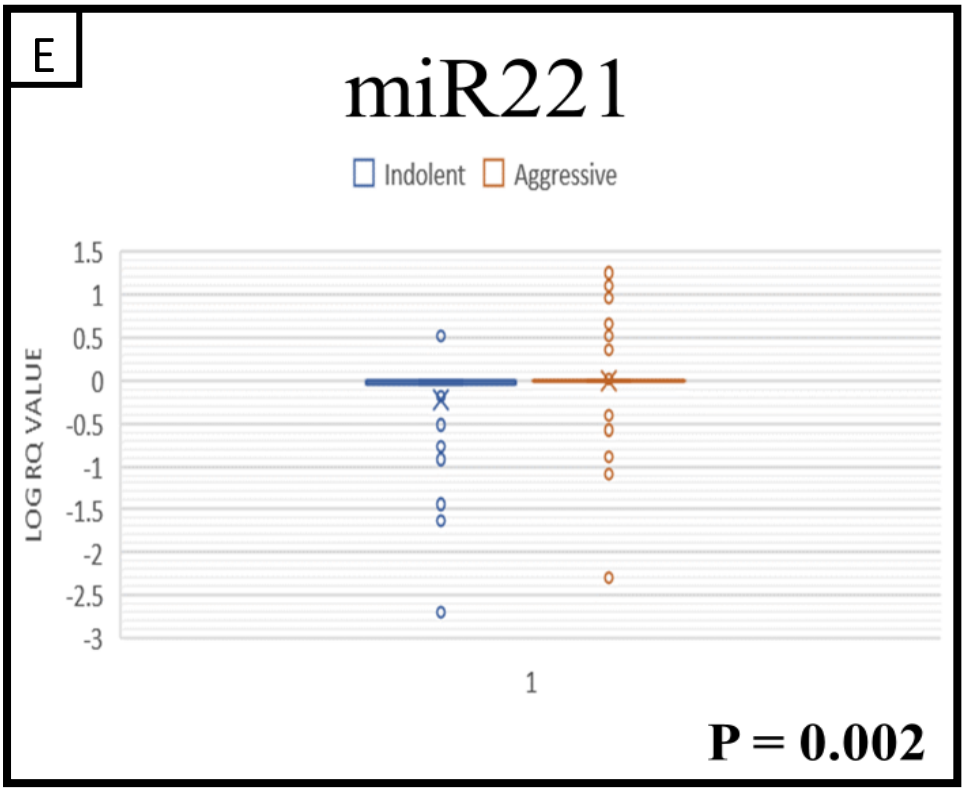
Figure 3.7 Heatmap showing the fold change in miRNA expression between paired 'lower grade' and 'higher grade' samples. Data is expressed as log RQ values in a distance matrix. Image generated using the online web tool 'Heatmapper'.

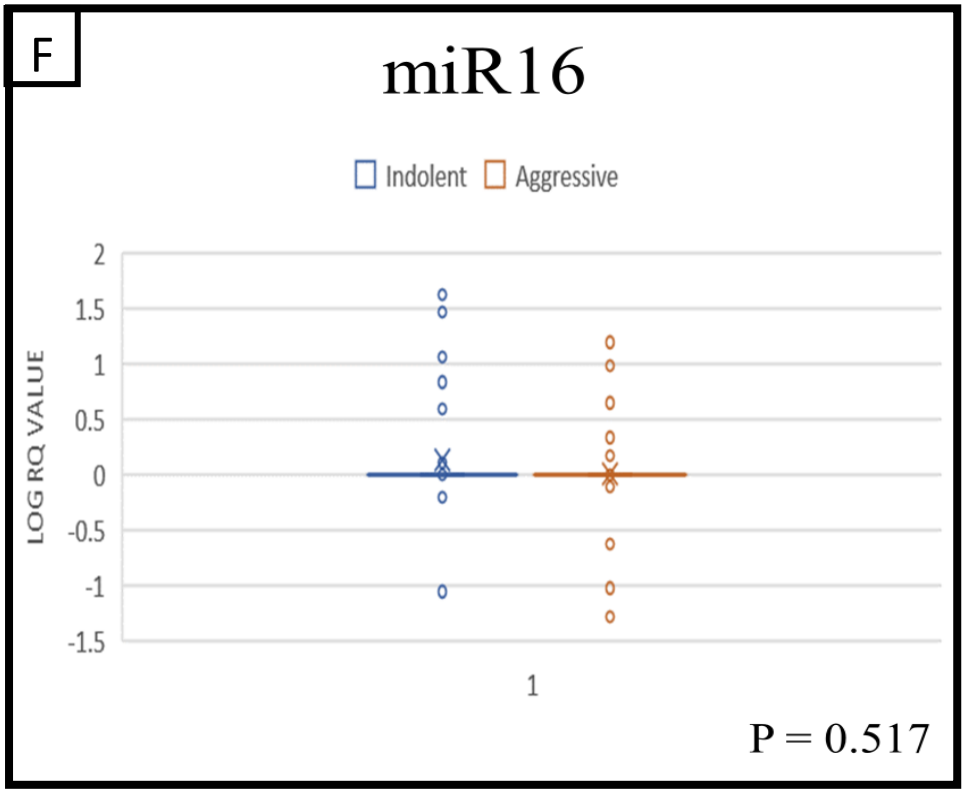

A
miR10b
Indolent
Aggressive
LOG RQ VALUE
2.5
2
1.5
1
0.5
0
1
P = 0.594

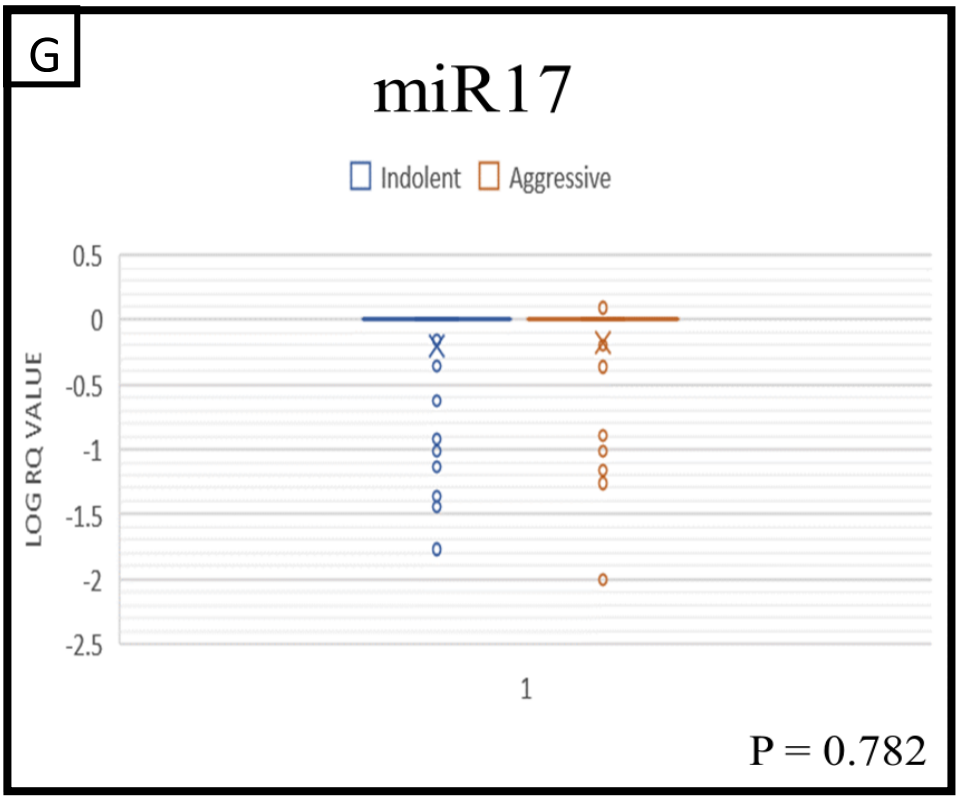

B
miR223
Indolent
Aggressive
LOG RQ VALUE
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
1
P = 0.866

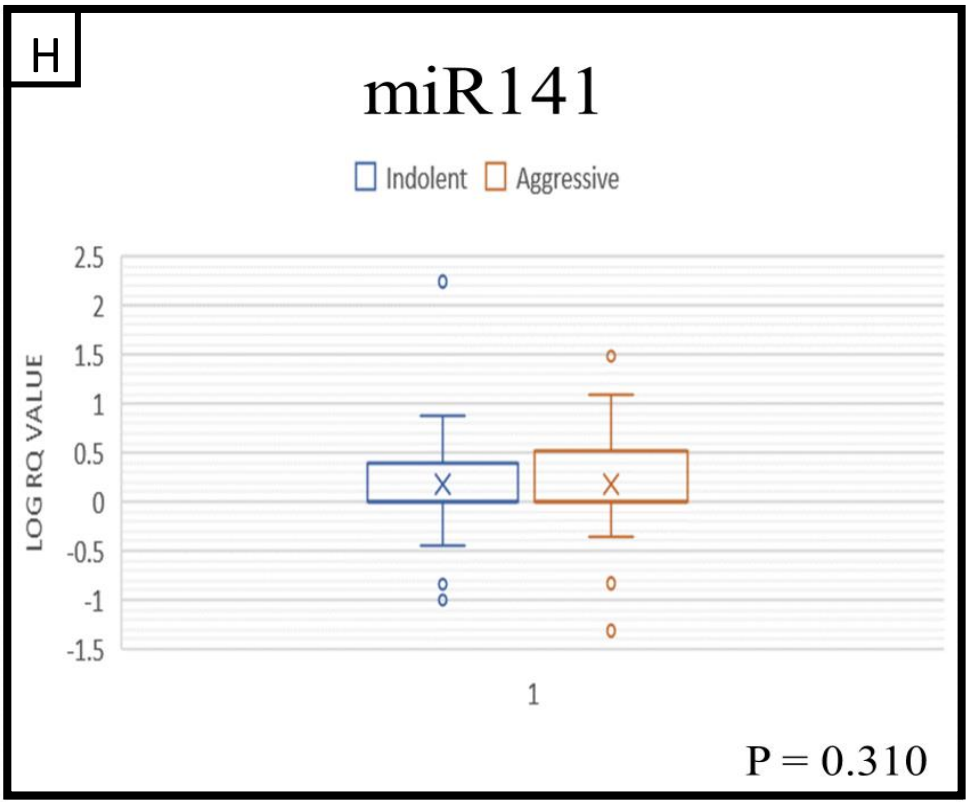

C
miR21
Indolent
Aggressive
LOG RQ VALUE
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
1
P = 0.311


D
miR224
Indolent
Aggressive
LOG RQ VALUE
2
1.5
1
0.5
0
-0.5
-1
-1.5
1
P = 0.214


E
miR221
Indolent
Aggressive
LOG RQ VALUE
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
1
P = 0.002


F
miR16
Indolent
Aggressive
LOG RQ VALUE
2
1.5
1
0.5
0
-0.5
-1
-1.5
1
P = 0.517


G
miR17
Indolent
Aggressive
LOG RQ VALUE
0.5
0
-0.5
-1
-1.5
-2
-2.5
1
P = 0.782


H
miR141
Indolent
Aggressive
LOG RQ VALUE
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
1
P = 0.310

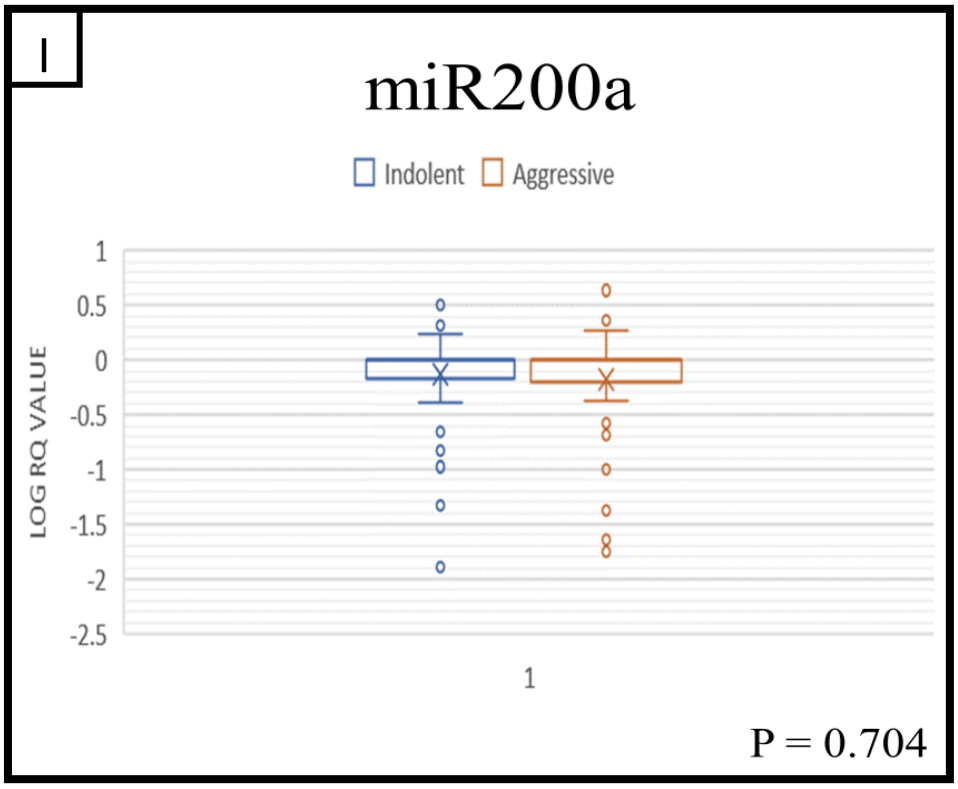


Figure 3.8 Box and whisker plots showing the difference in log RQ values of miRNA expression between paired lower grade and higher grade areas of tumour. A = miR-10b, B = miR-223, C = miR-21, D = miR-224, E = miR-221, F = miR-16, G = miR-17, H = miR-141, I = miR-200a. miR-221 showed a statistically significant change in expression between paired samples ($p=0.002$). No other miRNA showed a significant difference in expression. Wilcoxon matched pairs test, $p<0.05$. Data graphed as mean, median, range and interquartile range. X represents the mean. The central horizontal line represents the median. Where present, the top line represents the first quartile and the bottom line represents the third quartile. The whiskers denote the maximum and minimum values.

3.3.3 Association Between Clinicopathological Characteristics and miRNA Expression in 'Higher Grade' Tumour Samples

'Higher grade' samples were analysed from all cases to determine the presence of any significant correlation between miRNA expression and clinicopathological characteristics. Statistical analysis was performed as outlined in Section 3.2.3.

miR-21 was significantly associated with disease recurrence (p=0.001) and a higher tumour regression grade (TRG) (p=0.002). miR-221 also showed a significant correlation with disease recurrence (p=0.042). miR-224 expression was significantly associated with disease recurrence (p=0.021) and strongly correlated with higher TRG groups (p=0.052). (Table 3.7). Higher miR-141 expression levels were associated with HER2 positivity (p=0.05) (Table 3.8). No other miRNA in this panel showed a significant association with any clinicopathological characteristic.

Table 3.7 Correlation Between Clinicopathological Characteristics and miRNAs Upregulated in EMT and Stemness.

	miR-223			miR-224			miR-21			miR-221			miR-10b		
	Cases (*n*)	**Median (IQR)****	***p***	**Cases (*n*)**	**Median (IQR)****	***p***	**Cases (*n*)**	**Median (IQR)****	***p***	**Cases (*n*)**	**Median (IQR)****	***p***	**Cases (*n*)**	**Median (IQR)****	***p***
Disease Recurrence															
Yes	44	0 (0-0)	0.204	44	0.049 (0-1.413)	**0.021**	44	0.2 (0.015- 0.967)	**0.001**	44	0.008 (0-1.014)	**0.042**	44	0 (0-0)	0.292
No	32	0 (0-0)		32	0 (0-0.027)		32	0.015 (0-0.089)		32	0 (0-0.031)		32	0 (0-0)	
TRG*															
2	3	0 (0-0)	0.455	3	0 (0-0)	0.052	3	0.187 (0.093-0.25)	**0.002**	3	0 (0-0)	0.144	3	0 (0-0)	0.942
3	15	0 (0-0)		15	0 (0-0)		15	0 (0-0)		15	0 (0-0.0834)		15	0 (0-0)	
4	32	0 (0-0)		32	0.112 (0-1.367)		32	0.115 (0.015-0.298)		32	0.152 (0-1.095)		32	0 (0-0)	
5	14	0 (0-0)		14	0 (0-0.494)		14	0.227 (0.015-0.942)		14	0 (0-0.146)		14	0 (0-0)	
LVI															
Yes	53	0 (0-0)	0.204	53	0 (0-1.18)	0.878	53	0.031 (0-0.295)	0.633	53	0 (0-0.491)	0.696	53	0 (0-0)	0.626
No	23	0 (0-0)		23	0 (0-0.901)		23	0.172 (0-0.385)		23	0 (0-0.152)		23	0 (0-0)	
PNI															
Yes	13	0 (0-0)	0.941	13	0 (0-0.126)	0.295	13	0.068 (0.015-0.178)	0.983	13	0.068 (0.015-0.18)	0.806	13	0 (0-0)	0.835
No	63	0 (0-0)		63	0 (0-1.204)		63	0.043 (0-0.372)		63	0.043 (0-0.372)		63	0 (0-0)	
Serosal Involvement															
Yes	23	0 (0-0)	0.476	23	0 (0-0.727)	0.493	23	0.147 (0.008-0.304)	0.708	23	0 (0-0.152)	0.575	23	0 (0-0)	0.130
No	53	0 (0-0)		53	0 (0-1.18)		53	0.031 (0-0.399)		53	0 (0-0.386)		53	0 (0-0)	
HER2 Status						0.986									
Positive	8	0 (0-0)	0.932	8	0.049 (0-0.559)		8	0.136 (0.03-0.636)	0.772	8	0.193 (0-3.650)	0.746	8	0 (0-0)	0.986
Negative	31	0 (0-0)		31	0 (0-0.981)		31	0.187 (0-0.387)		31	0 (0-0.814)		31	0 (0-0)	
Siewert Classification															
I	22	0 (0-0)	0.086	22	0 (0-1.135)	0.456	22	0.159 (0-0.568)	0.387	22	0.073 (0-0.502)	0.258	22	0 (0-0)	0.694
II	25	0 (0-0)		25	0 (0-1.648)		25	0.015 (0-0.178)		25	0 (0-0.152)		25	0 (0-0)	
III	26	0 (0-0)		26	0 (0-0.277)		26	0.173 (0.014-0.356)		26	0 (0-0.118)		26	0 (0-0)	
T Stage															
T0	0	0 (0-0)	0.693	0	0 (0-0)	0.205	0	0 (0-0)	0.463	0	0 (0-0)	0.944	0	0 (0-0)	0.136
Tis, T1-2	11	0 (0-0)		11	0 (0-1.378)		11	0.015 (0-0.175)		11	0 (0-0.293)		11	0 (0-0)	
T3-4	65	0 (0-0)		65	0 (0-0.961)		65	0.098 (0-0.37)		65	0 (0-0.307)		65	0 (0-0)	
N Stage															
N0	22	0 (0-0)	0.215	22	0 (0-1.085)	0.851	22	0.029 (0-0.278)	0.762	22	0.003 (0-0.146)	0.794	22	0 (0-0)	0.143
N+	54	0 (0-0)		54	0 (0-0.990)		54	0.084 (0-0.393)		54	0 (0-0.465)		54	0 (0-0)	

* Cases with TRG 1 were excluded from the study as they fall under the category of (y)pT0. ** Interquartile Range (Quartile 1 – Quartile 3).

Table 3.8 Correlation Between Clinicopathological Characteristics and miRNAs Downregulated in EMT and Stemness.

	miR-16			miR-17			miR-200a			miR-141		
	Cases (*n*)	**Median (IQR)****	***p***	**Cases (*n*)**	**Median (IQR)****	***p***	**Cases (*n*)**	**Median (IQR)****	***p***	**Cases (*n*)**	**Median (IQR)****	***p***
Disease Recurrence												
Yes	44	0 (0-0.098)	0.379	44	0 (0-0)	0.301	44	0.446 (0.022-1)	0.162	43	0.838 (0-3.093)	0.205
No	32	0 (0-0.769)		32	0 (0-0)		32	0.012 (0-1)		32	0.074 (0-2.523)	
TRG*												
2	3	0 (0-0)	0.671	3	0 (0-0)	0.653	3	0 (0-0.5)	0.579	3	0.402 (0.201-1.082)	0.154
3	15	0 (0-1.84)		15	0 (0-0)		15	0.315 (0-1)		15	0 (0-0.502)	
4	32	0 (0-0.241)		32	0 (0-0)		32	0.282 (0.0134-1.041)		32	1.1025 (0-3.702)	
5	14	0 (0-0.741)		14	0 (0-0.0075)		14	0.180 (0-0.922)		13	0 (0-0.697)	
LVI												
Yes	53	0 (0-0.235)	0.754	53	0 (0-0)	0.947	53	0.237 (0-1)	0.340	52	0.149 (0-3.308)	0.778
No	23	0 (0-0.419)		23	0 (0-0)		23	0.472 (0.021-1)		23	0.793 (0-2.066)	
PNI												
Yes	13	0 (0-0.909)	0.819	13	0 (0-0)	0.624	13	0.1 (0-0.59)	0.413	13	0 (0-4.282)	0.661
No	63	0 (0-0.165)		63	0 (0-0)		63	0.315 (0-1)		62	0.587 (0-2.650)	
Serosal Involvement												
Yes	23	0 (0-0.048)	0.573	23	0 (0-0)	0.847	23	0.59 (0-1)	0.737	22	0.595 (0-2.525)	0.724
No	53	0 (0-0.774)		53	0 (0-0)		53	0.3 (0-1)		53	0.15 (0-2.929)	
HER2 Status												
Positive	8	0 (0-0.265)	0.905	8	0 (0-0)	0.670	8	0.286 (0.017-1.044)	0.851	8	2.9845 (2.179-3.911)	**0.050**
Negative	31	0 (0-0.955)		31	0 (0-0)		31	0.421 (0-1)		30	0.466 (0-1.878)	
Siewert Classification	22	0 (0-0.75)	0.963	22	0 (0-0)	0.449	22	0.222 (0-0.670)	0.351	22	0.275 (0-1.740)	0.782
I	25	0 (0-0.462)		25	0 (0-0)		25	0.746 (0-1)		24	0.595 (0-3.819)	
II	26	0 (0-0.087)		26	0 (0-0)		26	0.273 (0-1)		26	0.818 (0-3.339)	
III												
T Stage												
T0	0	0 (0-0)	0.677	0	0 (0-0)	0.596	0	0 (0-0)	0.681	0	0 (0-0)	0.578
Tis, T1-2	11	0 (0-1.075)		11	0 (0-0)		11	0.3 (0-0.736)		11	0 (0-0.846)	
T3-4	65	0 (0-0.095)		65	0 (0-0)		65	0.307 (0-1)		64	0.466 (0-3.020)	
N Stage												
N0	22	0 (0-1.220)	0.226	22	0 (0-0)	0.980	22	0.669 (0.006-1)	0.455	22	0.996 (0-3.345)	0.473
N+	54	0 (0-0.071)		54	0 (0-0)		54	0.258 (0-0.949)		53	0.402 (0-2.677)	

* Cases with TRG 1 were excluded from the study as they fall under the category of (y)pT0. ** Interquartile Range (Quartile 1 – Quartile 3).

Table 3.9 Correlation Between Age, Tumour Size and miRNA Expression.

	miR-16	miR-17	miR-21	miR-221	miR-10b	miR-223	miR-200a	miR-141	miR-224
Age at Diagnosis									
Pearson correlation (*r*)	-0.164	-0.190	0.156	0.009	-0.006	0.017	0.077	0.127	0.074
Number (*n*)	76	76	76	76	76	76	76	75	76
Significance (*p*)	0.157	0.099	0.179	0.937	0.962	0.884	0.508	0.277	0.526
Macroscopic Tumour Size									
Spearman correlation (ρ)	-0.022	-0.153	0.118	-0.030	-0.101	-0.015	0.060	0.135	-0.087
Number (*n*)	69	60	69	69	69	69	69	68	69
Significance (*p*)	0.858	0.210	0.333	0.804	0.408	0.904	0.626	0.274	0.477

3.4 Protein Expression Analysis

3.4.1 Establishing a Scoring System for Immunohistochemical Markers

Protein expression was determined using 7 immunohistochemical markers. Of these, 5 had been previously optimized and validated for use in clinical practice: Epithelial cell adhesion molecule (EpCAM), Oct-4, CD34, E-cadherin and Vimentin. Aldehyde dehydrogenase 1 family, member A1 (ALDH1A1) and CD133 were optimized as part of this study and a semi-quantitative scoring system was used for each, as directed by validated scoring systems identified within published literature (Fedchenko & Reifenrath, 2014; Yang et al., 2018; Zeppernick et al., 2008) (Table 3.10). CD133 staining was graded in two separate categories: the percentage of tumour cells stained and the topology of the positive staining cells, which reflects the arrangement of tumour cells as single cells or in clusters. All IHC stains were double scored by two Histopathologists. Discordant results were co-reviewed by both pathologists to reach a consensus. A Consultant Histopathologist reviewed approx. 10% of slides for quality control purposes.

Table 3.10 IHC Scoring Systems.

Primary Antibody	**Staining Pattern**	**Scoring System**
EpCAM	Membranous	0 = no staining 1 = weak/incomplete staining 2 = moderate staining 3 = strong staining
E-cadherin	Membranous	0 = no staining 1 = positive staining
Vimentin	Cytoplasmic	0 = no staining 1 = positive staining
Oct-4	Nuclear	0 = no staining 1 = positive nuclear staining
CD34	Membranous	0 = no staining 1 = positive staining

Table 3.10 IHC Scoring Systems.

CD133 **(Zeppernick et al., 2008)**	Membranous / Perinuclear Luminal staining of glandular structures interpreted as negative	**Percentage Area:** 0 = <1% 1 = 1-5% 2 = 5-10% 3 = 10-25% 4 = 25-50% 5 = >50% **Topology:** 0 = no staining 1 = single cells 2 = single cells in clusters (≥5 cells)
ALDH1A1 **(Yang et al., 2018)**	Cytoplasmic **Both scores are added together to give a final value.**	**Intensity:** 0 = no staining 1 = weak staining 2 = moderate staining 3 = strong staining **Percentage Area:** 0 = <5% 1 = 5-25% 2 = 26-50% 3 = 51-75% 4 = >75%

3.4.2 Differential Expression of Proteins in Paired Samples

IHC was employed to assess the expression of markers associated with EMT and stemness in GEJA. A panel of 7 IHC markers was chosen, of which 2 are associated with EMT (E-cadherin and Vimentin) and 5 are known or proposed markers of CSCs: EpCAM, ALDH1A1, CD133, Oct-4 and CD34.

IHC was performed on 5 TMAs, which included tissue cores from 58 paired tumour samples, 21 unpaired 'higher grade' tumour samples and 3 control tissues: liver, kidney and background normal gastroesophageal junction (GEJ). Where possible, 3 cores were taken from each area of tumour. When there was heterogeneity in staining intensity across cores from the same tumour area, the average staining intensity was recorded. Due to sample depletion, not all cores contained tumour cells for analysis for each IHC marker.

Protein expression was recorded for each paired sample, as per the scoring systems devised (Table 3.10, Figures 3.9-3.14). CD34 was not included in the analysis due to no expression in any sample. Where present, the direction of change in expression was recorded from 'lower grade' to 'higher grade' samples (Appendix 1.5). Pairwise analysis was performed for each protein using a Wilcoxon matched pairs test.

EpCAM was shown to be significantly upregulated in 'lower grade' samples compared to 'higher grade' ($p<0.001$). No other protein in this panel showed a significant change in expression (Figure 3.16). EpCAM expression decreased in 92.9% (26/28) of pairs with differential expression. In cases suitable for analysis, 72.7% (8/11) showed a decrease in Oct-4 staining, CD133 percentage staining decreased in 57.1% (16/28) and 56% (14/25) showed a decrease in CD133 topology (Table 3.11). ALDH1A1 expression showed equal rates of increase and decrease in expression within cases with differential expression patterns. Vimentin expression increased and E-cadherin expression decreased in paired

samples with differential expression, however the numbers were too small to hold any significance. Figure 3.15 depicts the degree of fold change between paired samples. Fold change values for each case are listed in Appendix 1.6. Of note, fold change could not be calculated for pairs with expression recorded in only one sample.

Table 3.11 Patterns of Protein Expression Paired Samples.

	No Staining in Both Samples	Equal Positive Expression	Increase in Expression	Decrease in Expression	One or Both Samples Not Analysed
EpCAM	0	16	2	26	14
OCT-4	31	7	3	8	9
E-Cadherin	0	47	0	2	9
Vimentin	43	0	3	0	12
ALDH1A1	1	15	16	16	10
CD133 %	12	10	12	16	8
CD133 Topology	12	13	11	14	8

Interestingly, upon scoring the TMAs, it was observed that Oct-4 expression was consistently increased in benign glandular tissue, glandular tissue with intestinal metaplasia (Figure 3.17) and dysplastic glandular tissue. However, these tissue samples were not formally scored.

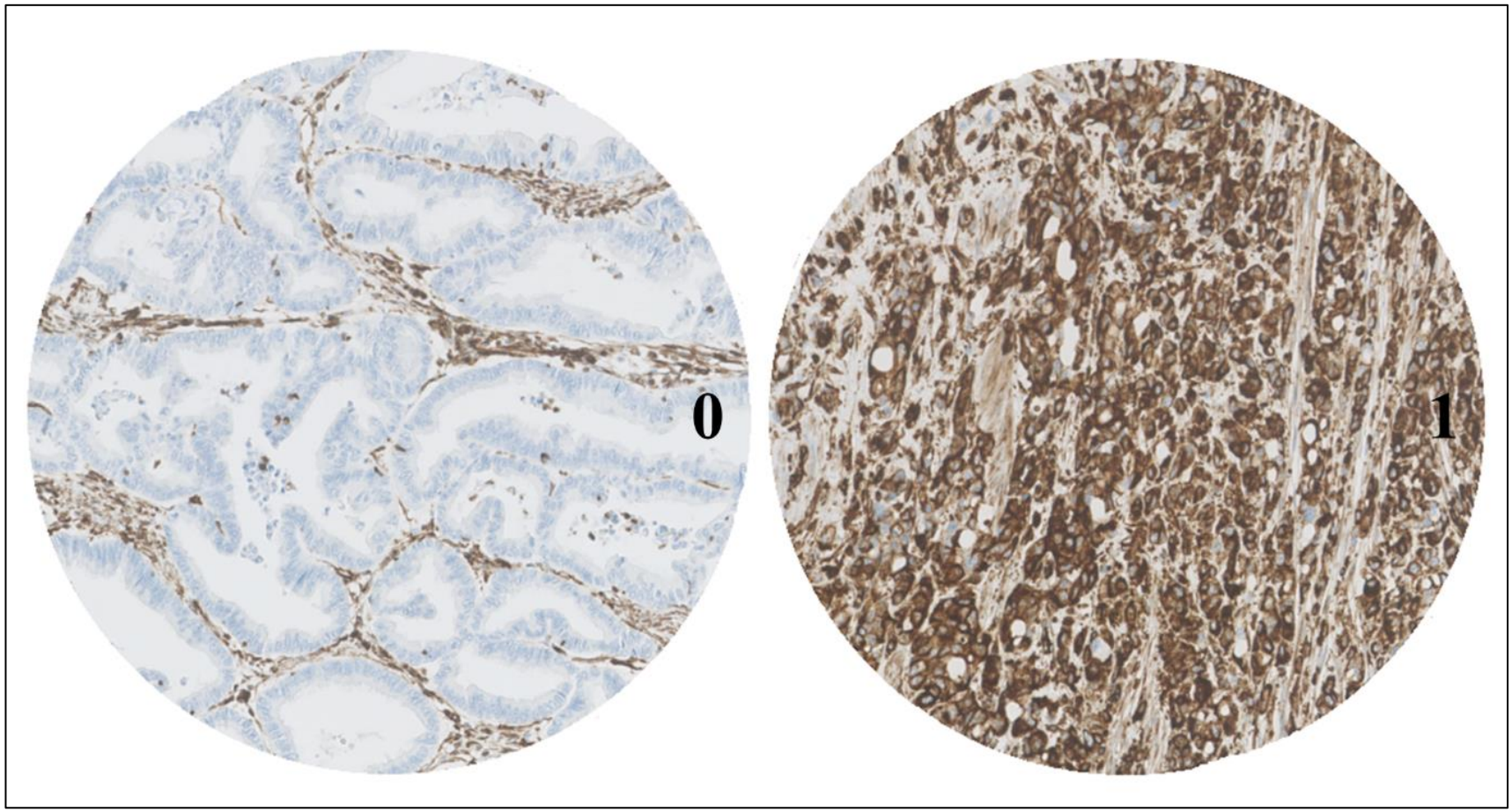


Figure 3.9 Representative images of Vimentin staining intensity. 0 = no staining, 1 = positive staining. (40X magnification)

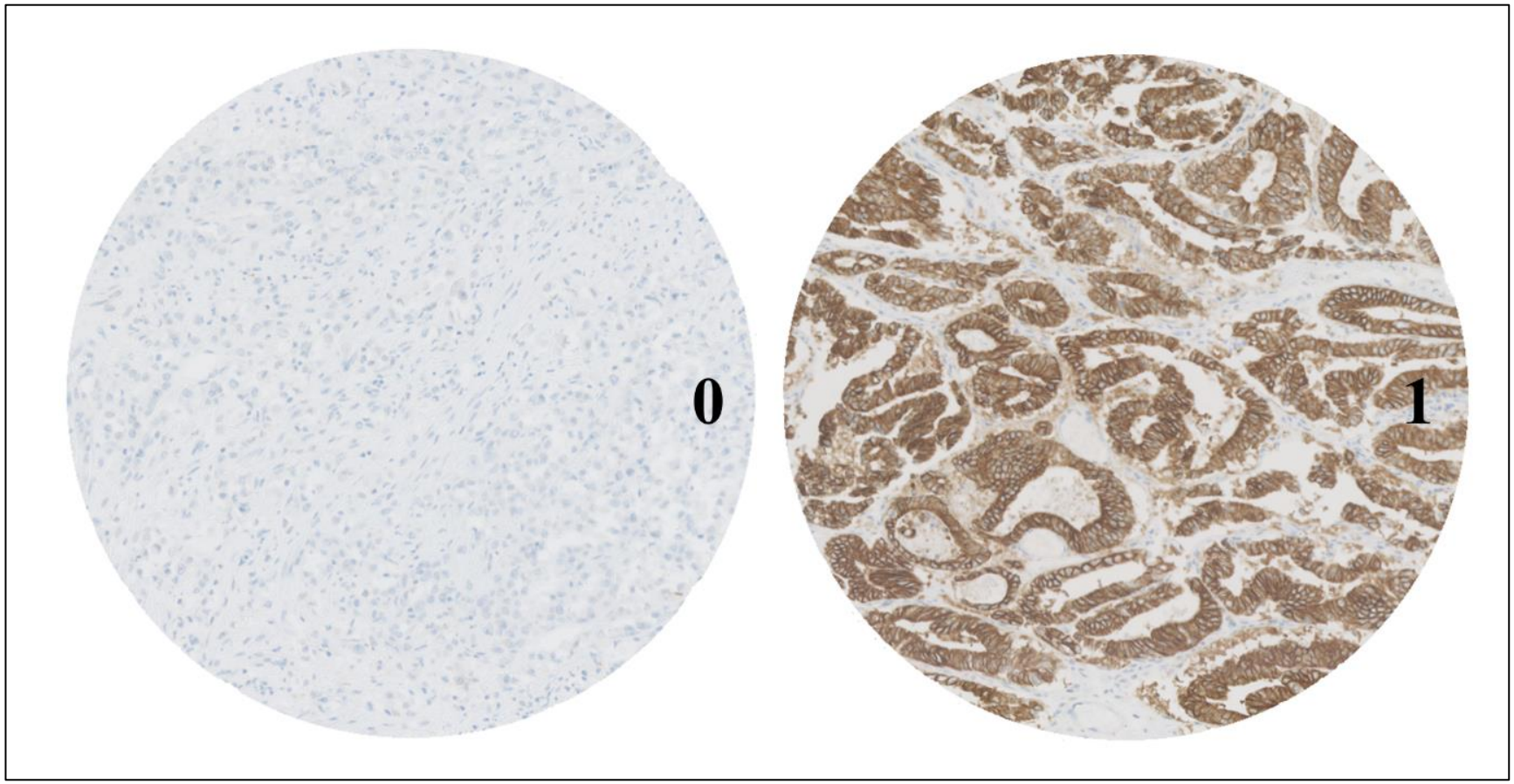


Figure 3.10 Representative images of E-cadherin staining intensity. 0 = no staining, 1 = positive staining. (40X magnification)

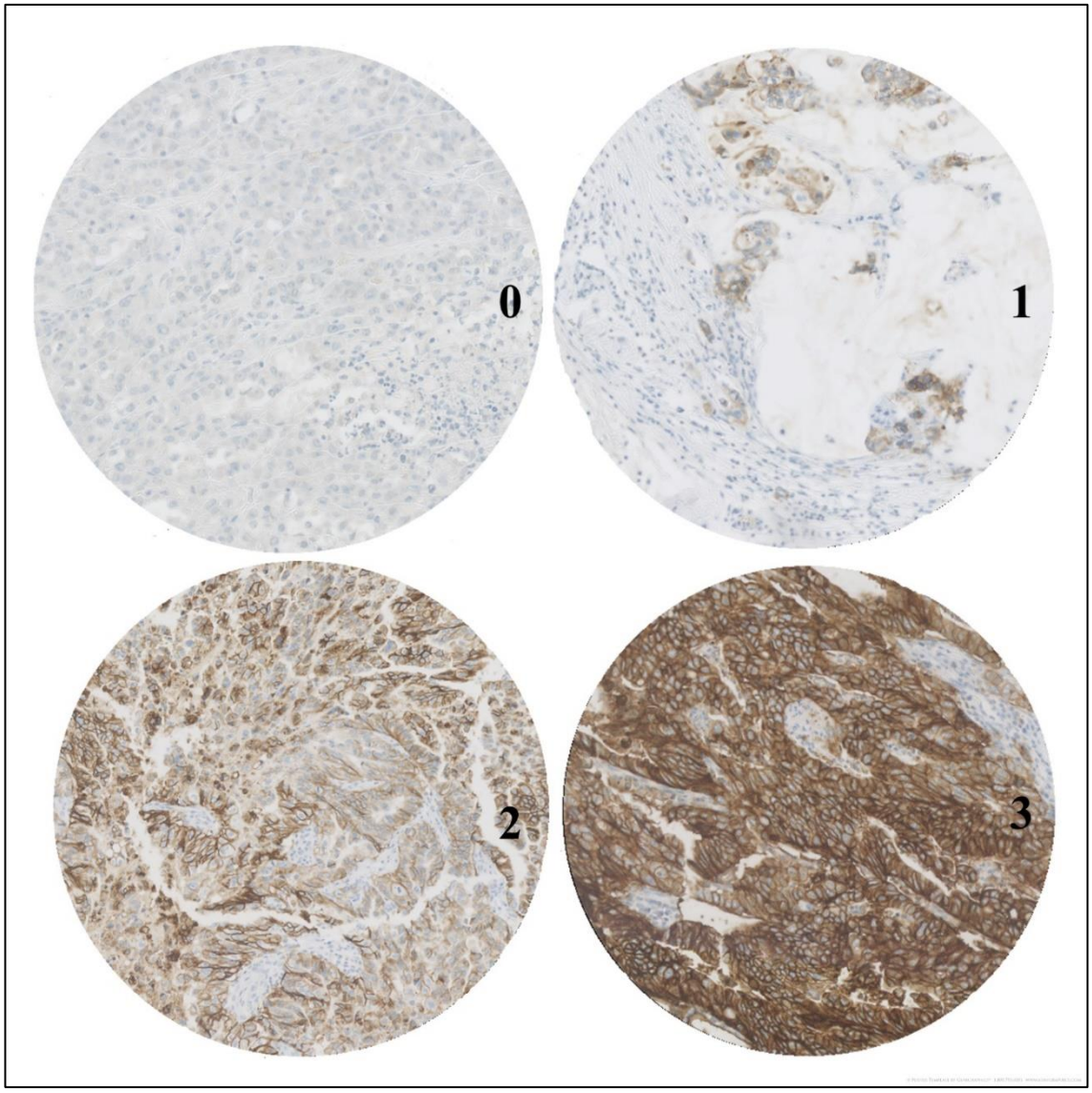


Figure 3.11 Representative images of EpCAM staining intensity. 0 = no staining, 1 = weak/incomplete staining, 2 = moderate staining, 3 = strong staining. (40X magnification)

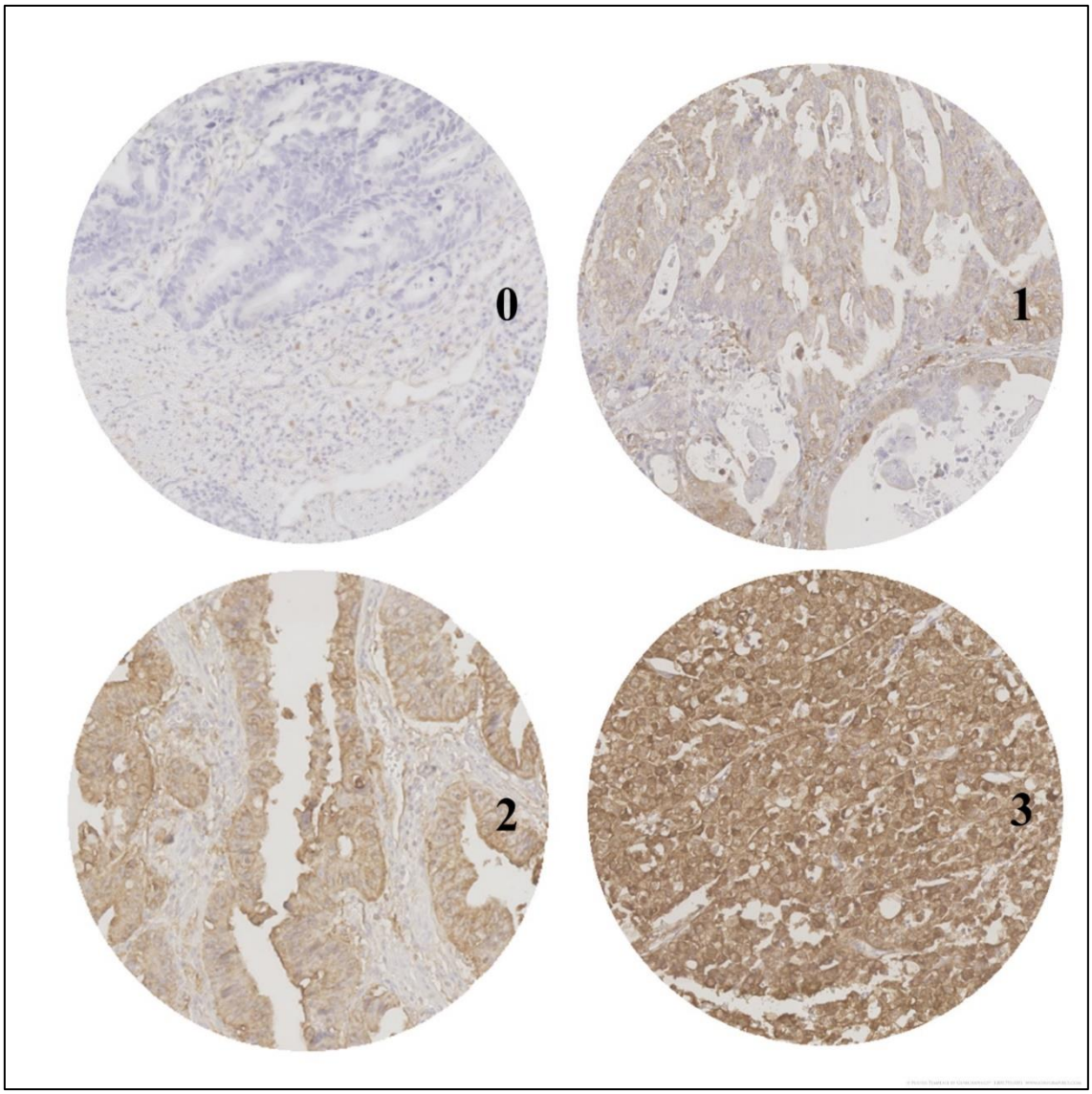


Figure 3.12 Representative images of ALDH1A1 staining intensity. 0 = no staining, 1 = weak staining, 2 = moderate staining, 3 = strong staining. (40X magnification)

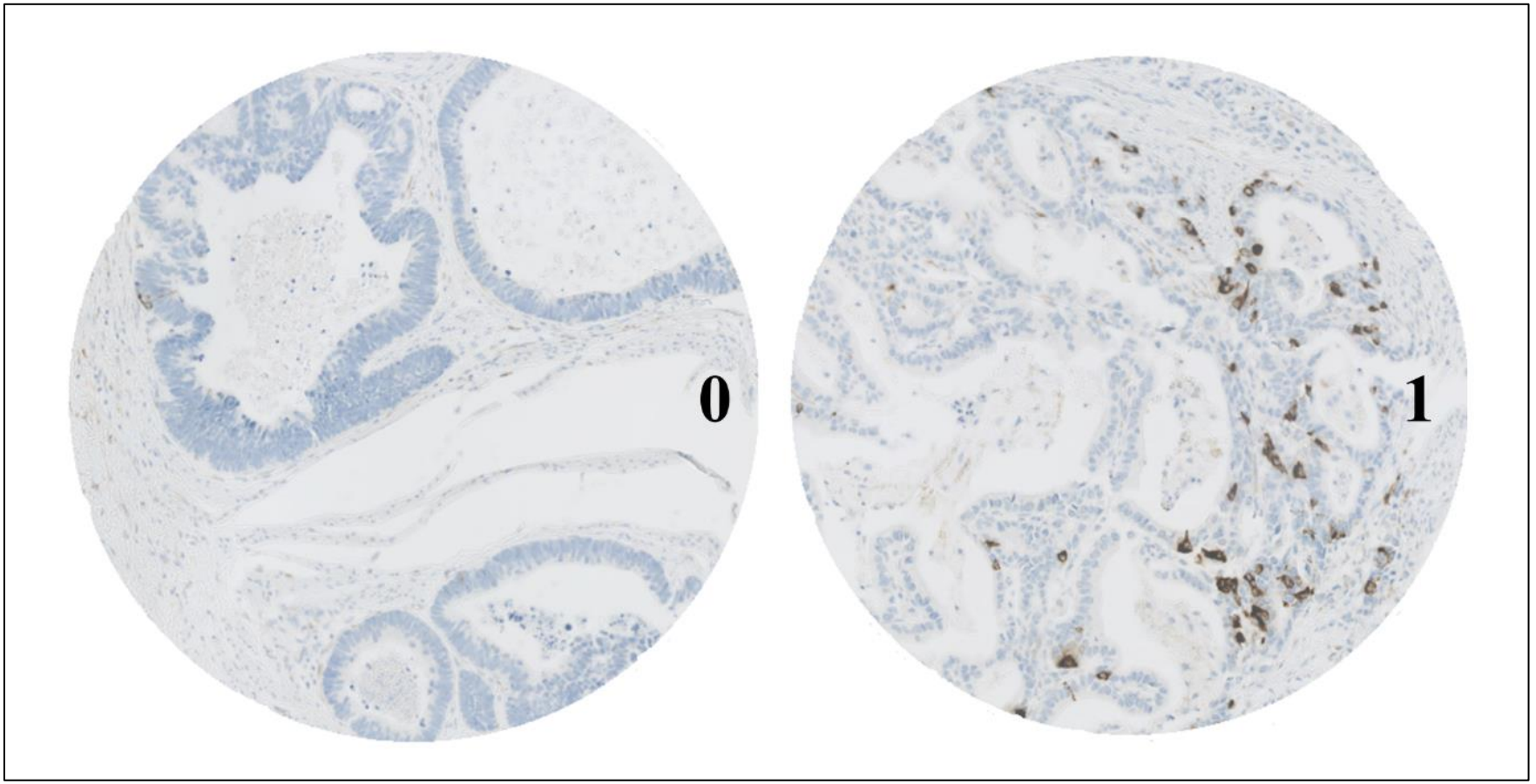


Figure 3.13 Representative images of Oct-4 intensity. 0 = no staining, 1 = positive staining. (40X magnification)

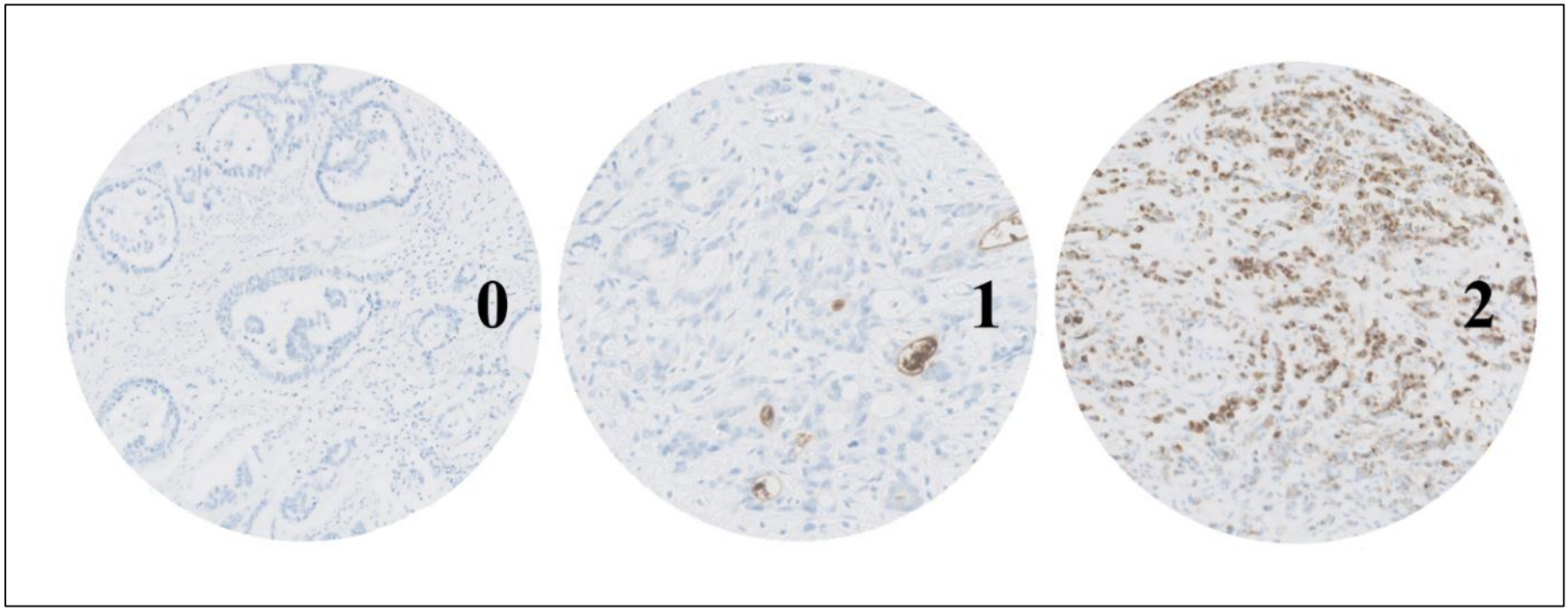


Figure 3.14 Representative images of CD133 topology. 0 = no staining, 1 = single cells, 2 = single cells and clusters (≥5 cells). (40X magnification) The central core (1) contains areas of luminal staining, interpreted as negative.

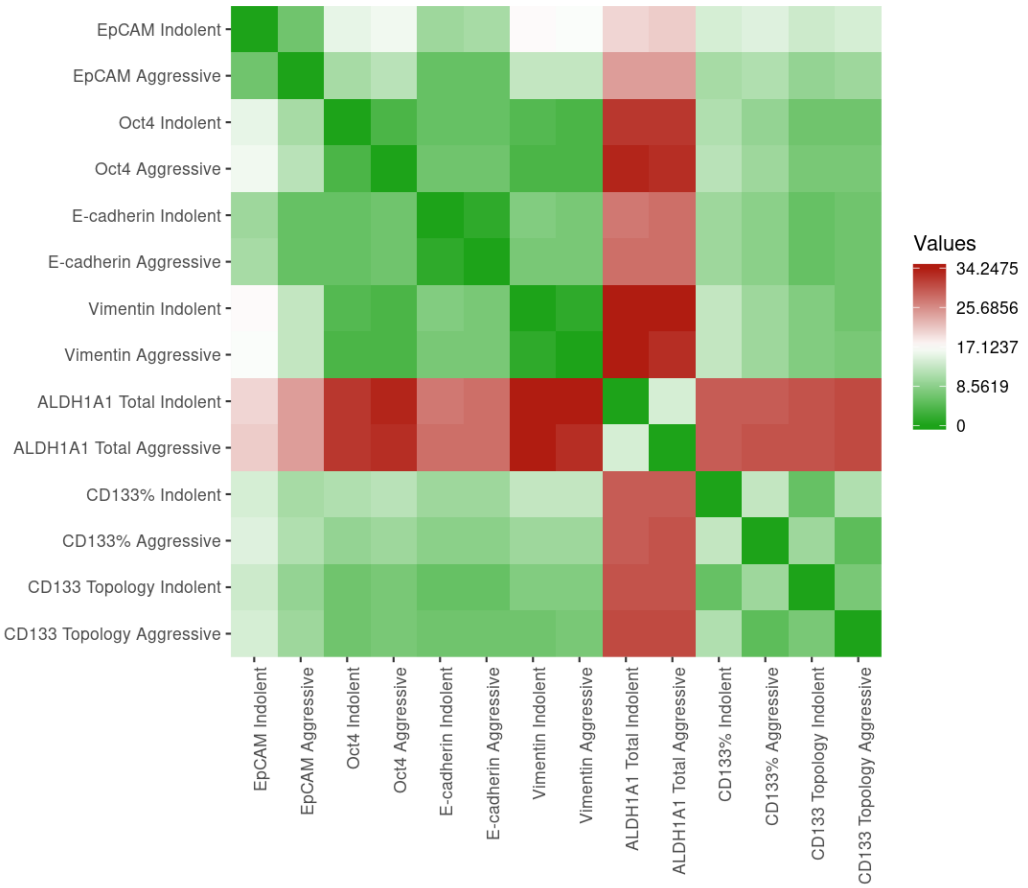
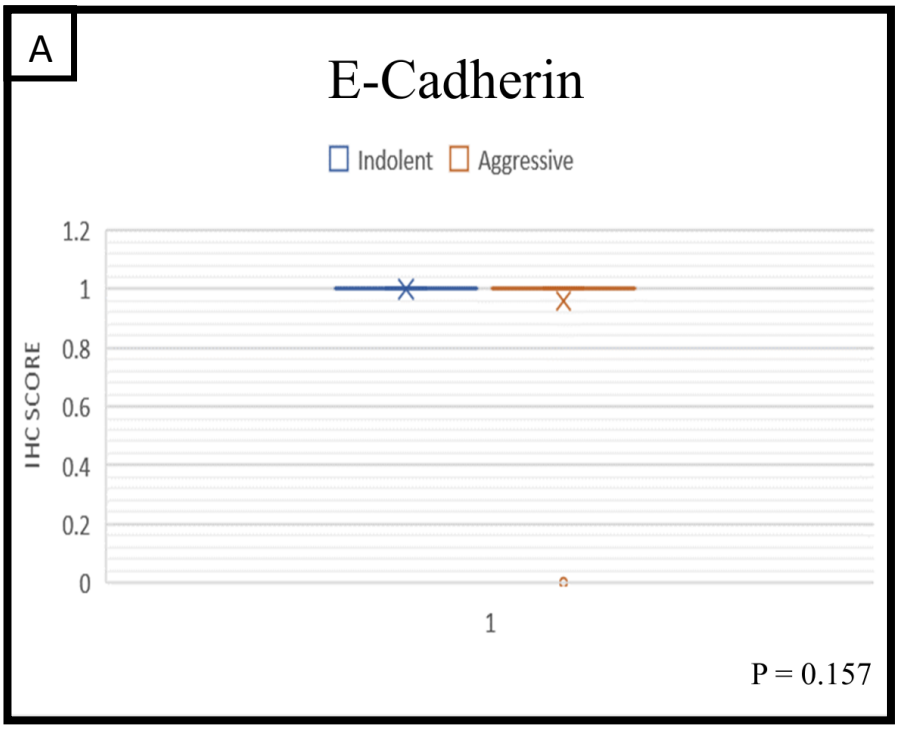
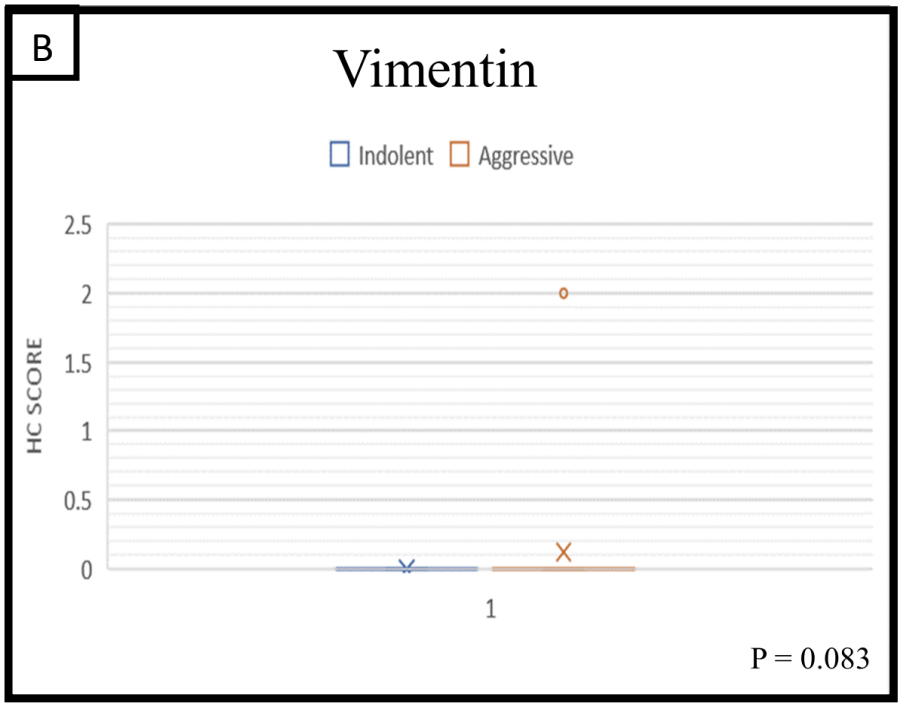
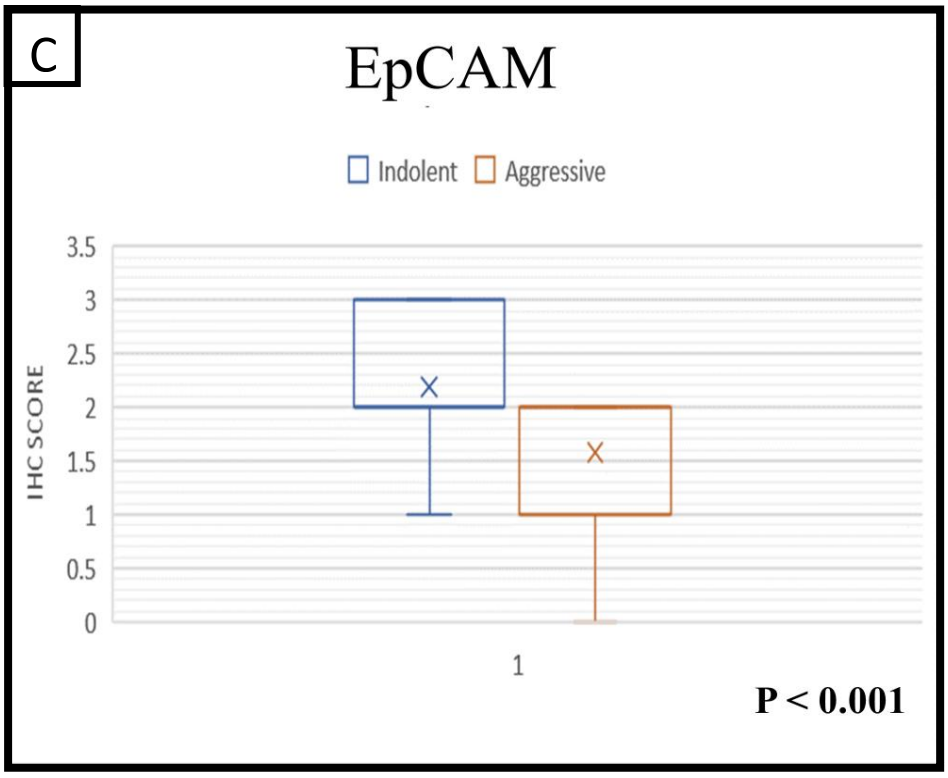
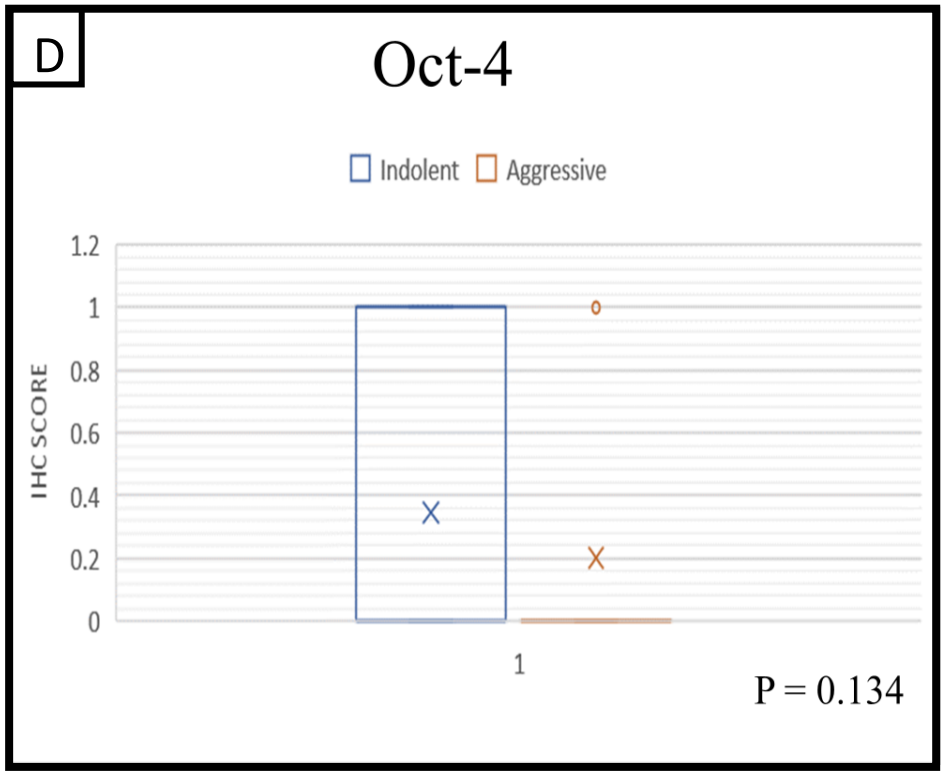


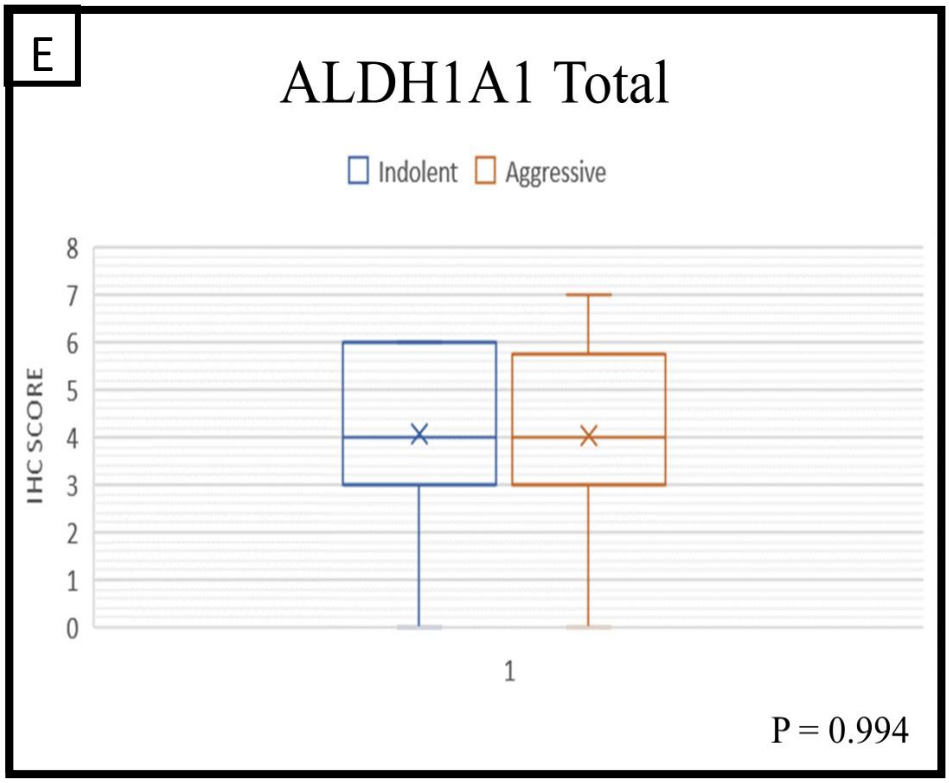
Figure 3.15 Heatmap showing the fold change in protein expression between paired 'lower grade' and 'higher grade' samples. Data is expressed as ordinal values in a distance matrix. Image generated using the online web tool 'Heatmapper'.

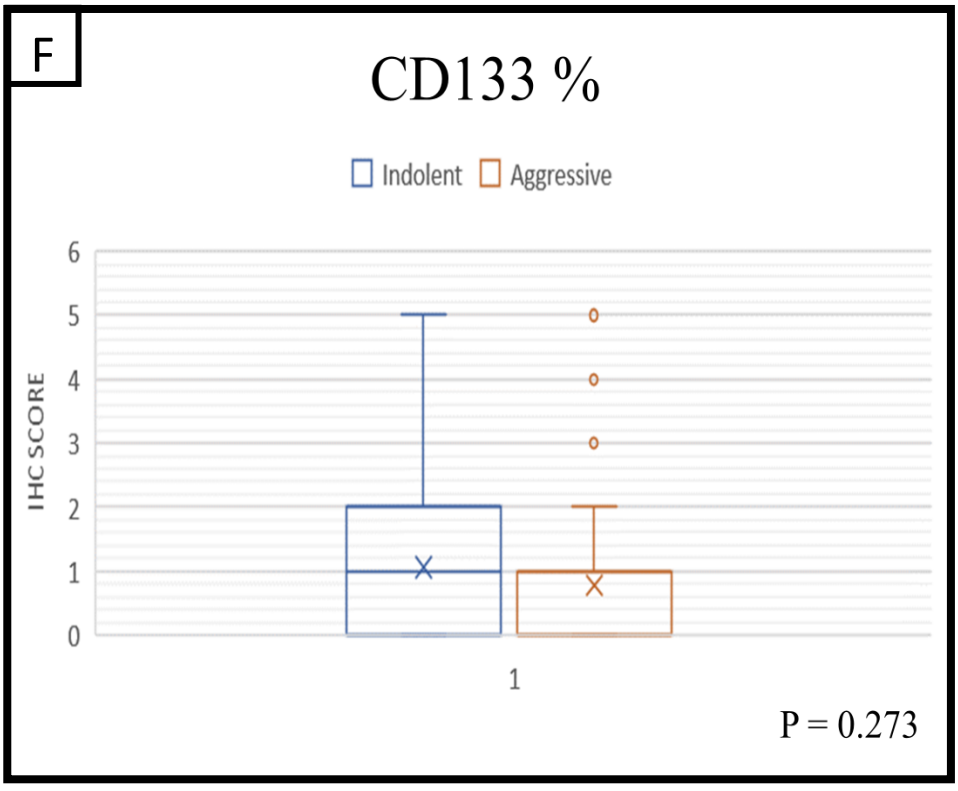

A
E-Cadherin
Indolent
Aggressive
IHC SCORE
1.2
1
0.8
0.6
0.4
0.2
0
1
P = 0.157


B
Vimentin
Indolent
Aggressive
HC SCORE
2.5
2
1.5
1
0.5
0
1
P = 0.083


C
EpCAM
Indolent
Aggressive
IHC SCORE
3.5
3
2.5
2
1.5
1
0.5
0
1
P < 0.001


D
Oct-4
Indolent
Aggressive
IHC SCORE
1.2
1
0.8
0.6
0.4
0.2
0
1
P = 0.134


E
ALDH1A1 Total
Indolent
Aggressive
IHC SCORE
8
7
6
5
4
3
2
1
0
1
P = 0.994


F
CD133 %
Indolent
Aggressive
IHC SCORE
6
5
4
3
2
1
0
1
P = 0.273

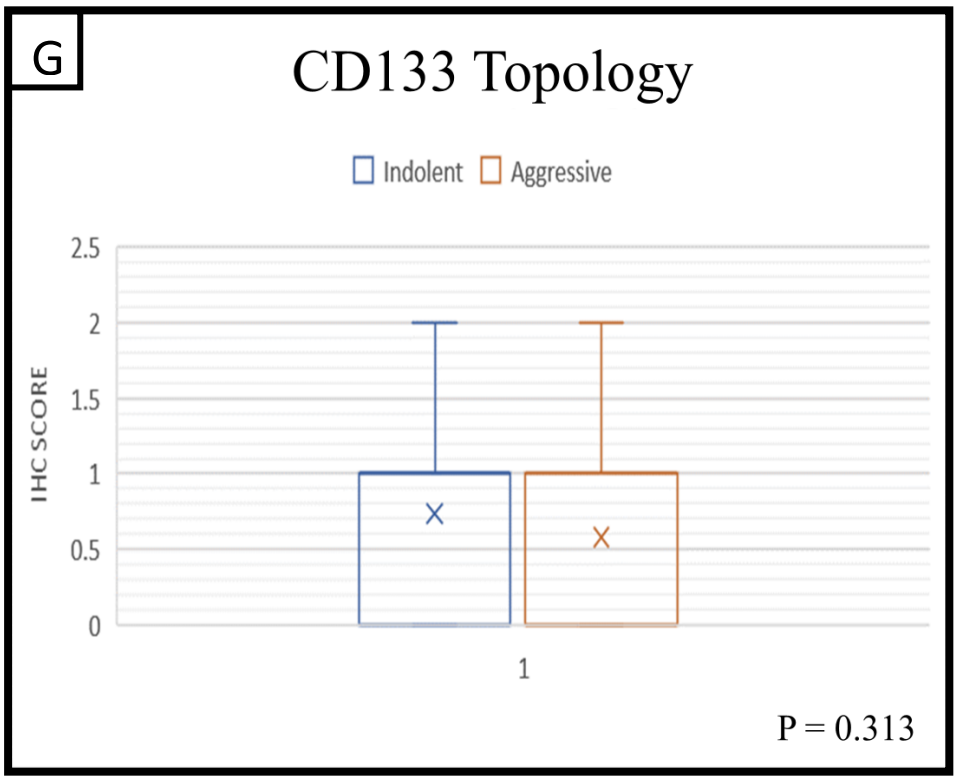


Figure 3.16 Box and whisker plots showing the difference IHC scores of protein expression between paired lower grade and higher grade areas of tumour. A = E-cadherin, B = Vimentin, C = EpCAM, D = Oct-4, E = ALDH1A1 Total, F = CD133 Percentage, G = CD133 Topology. A significant difference in expression was noted in EpCAM ($p<0.0010$). No statistically significant differential expression was observed any other protein. Statistical significance: Wilcoxon matched pairs test, $p<0.05$. Data graphed as mean, median, range and interquartile range. X represents the mean. The central horizontal line represents the median. Where present, the top line represents the first quartile and the bottom line represents the third quartile. The whiskers denote the maximum and minimum values.

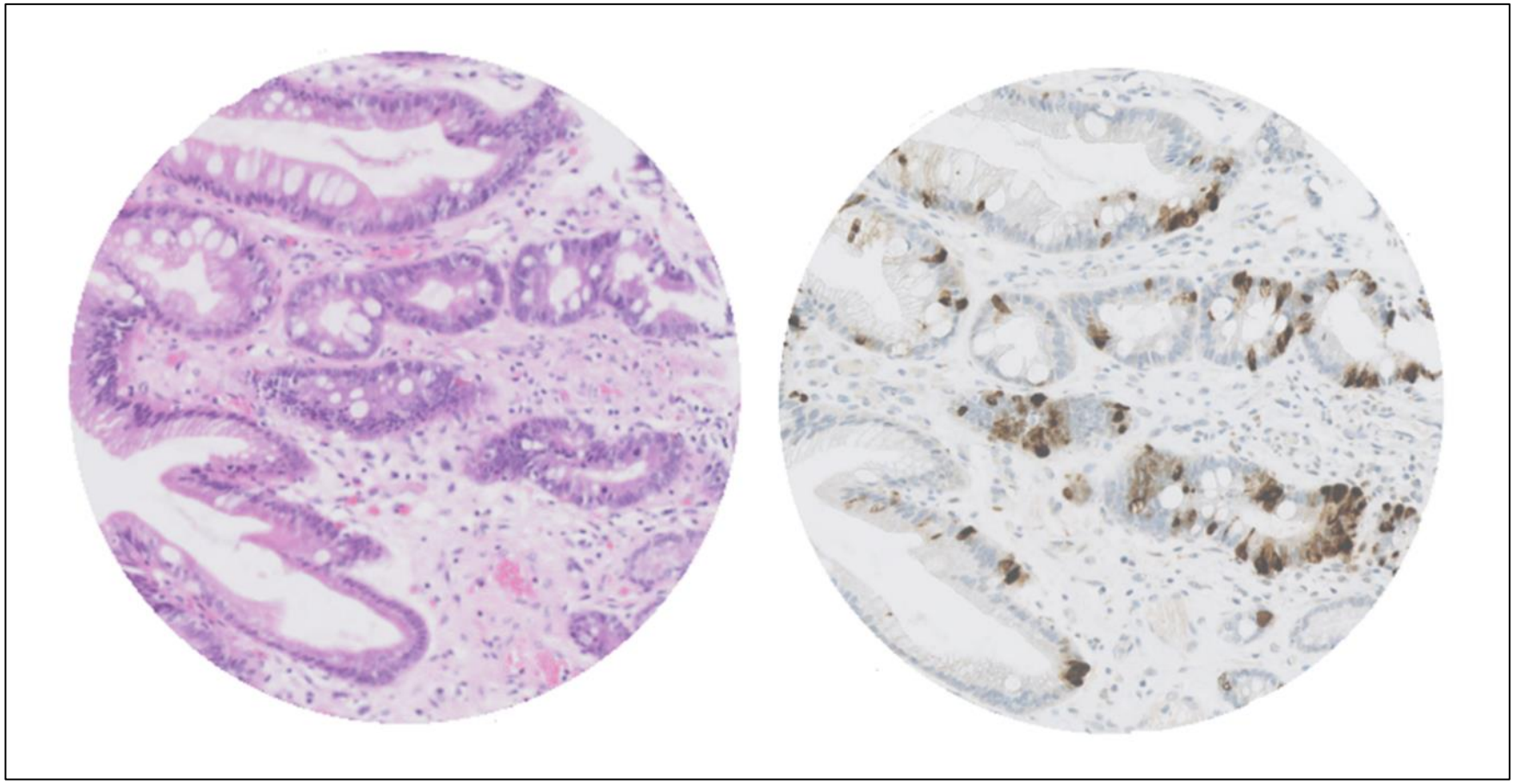


Figure 3.17 Representative images of Oct-4 staining in benign glandular tissue with intestinal metaplasia. (40X magnification)

3.4.3 Association Between Clinicopathological Characteristics and Protein Expression in 'Higher Grade' Tumour Samples

'Higher grade' samples were analysed in all patients to determine the presence of any significant correlation between protein expression and clinicopathological characteristics.

Percentage staining with CD133 was significantly associated with disease recurrence (p=0.025). No other statistically significant correlation between protein expression and clinicopathological features were identified in any protein in this panel, however a strong association was noted between serosal involvement and E-cadherin expression (p=0.075), CD133 percentage staining (p=0.088) and CD133 topology (p=0.062) (Tables 3.12 and 3.13). Age at diagnosis also showed a strong negative correlation with CD133 percentage staining (p=0.053, r= -0.233) (Table 3.14).

Table 3.12 Correlation Between Clinicopathological Characteristics and Protein Expression Associated with Stemness.

	EpCAM			Oct-4			ALDH1A1 Total			CD133%			CD133 Topology		
	Cases (*n*)	**Median (IQR)****	***p***	**Cases (*n*)**	**Median (IQR)****	***p***	**Cases (*n*)**	**Median (IQR)****	***p***	**Cases (*n*)**	**Median (IQR)****	***p***	**Cases (*n*)**	**Median (IQR)****	***p***
Disease Recurrence															
Yes	36	2 (1-2)	0.789	38	0 (0-0)	0.973	37	4 (3-6)	0.288	39	4 (0-2)	**0.025**	39	1 (0-1)	0.149
No	30	2 (1-2)		30	0 (0-0)		28	5 (3.75-6)		31	0 (0-1)		31	0 (0-1)	
TRG*															
2	3	1 (1-1.5)	0.169	3	0 (0-0.5)	0.467	2	4 (3-6)	0.688	3	0 (0-2)	0.935	3	0 (0-1)	0.796
3	11	1 (1-2)		12	0 (0-1)		13	5 (3.75-6)		14	0 (0-2.5)		14	0 (0-1)	
4	30	2 (1.25-2)		30	0 (0-1)		29	4 (3-6)		31	0 (0-1.5)		31	1 (0-1)	
5	11	2 (1-2)		12	0 (0-0)		11	5 (3.75-6)		12	1 (0-1.25)		12	1 (0-1)	
LVI															
Yes	48	2 (1-2)	0.966	50	0 (0-0)	0.623	48	4 (3-6)	0.427	51	0 (0-1.5)	0.633	51	1 (0-1)	0.408
No	18	2 (1-2)		18	0 (0-0.75)		17	5 (4-6)		19	0 (0-1.5)		19	0 (0-1)	
PNI															
Yes	11	2 (2-2)	0.139	10	0 (0-1)	0.277	12	5 (4-6)	0.622	12	0 (0-1)	0.632	12	0.5 (0-1)	0.945
No	55	2 (1-2)		56	0 (0-0)		53	5 (3-6)		58	0 (0-2)		58	0 (0-1)	
Serosal Involvement															
Yes	23	2 (1-2)	0.936	22	0 (0-0)	0.475	21	4 (3-6)	0.591	22	1 (0-2)	0.088	22	1 (0-1.75)	0.062
No	43	2 (1-2)		46	0 (0-0.75)		44	5 (3-6)		48	0 (0-1)		48	0 (0-1)	
HER2 Status															
Positive	8	1.5 (1-2)	0.374	8	0 (0-0)	0.796	7	3 (2.5-6)	0.859	8	1 (0-1.75)	0.952	8	1 (0-1.25)	0.921
Negative	25	2 (1-2)		26	0 (0-0)		25	4 (3-6)		26	1 (0-3)		26	1 (0-1.75)	
Siewert Classification															
I	21	2 (1-2)	0.752	21	0 (0-0)	0.965	19	5 (4-5.5)	0.583	21	1 (0-2)	0.628	21	1 (0-2)	0.574
II	21	2 (1-2)		22	0 (0-0)		21	5 (3-5)		23	0 (0-1)		23	1 (0-1)	
III	22	2 (1-2)		23	0 (0-0.5)		23	5 (3.5-6)		24	0 (0-2)		24	0 (0-1)	
T Stage															
T0	0	0 (0-0)	0.456	0	0 (0-0)	0.645	0	0 (0-0)	0.421	0	0 (0-0)	0.430	0	0 (0-0)	0.571
Tis, T1-2	7	1 (1-2)		8	0 (0-0.25)		7	4 (3-5)		8	0.5 (0-2)		8	1 (0-1.25)	
T3-4	59	2 (1-2)		60	0 (0-0)		58	5 (3.25-6)		62	0 (0-1.75		62	0 (0-1)	
N Stage															
N0	21	2 (1-2)	0.308	20	0 (0-1)	0.420	21	5 (4-5)	0.926	22	0 (0-1)	0.437	22	0 (0-1)	0.437
N+	45	2 (1-2)		48	0 (0-0)		44	4 (3-6)		48	0 (0-2)		48	0.5 (0-1)	

* Cases with TRG 1 were excluded from the study as they fall under the category of (y)pT0. ** Interquartile Range (Quartile 1 – Quartile 3)

Table 3.13 Correlation Between Clinicopathological Characteristics and Protein Expression Associated with EMT.

	E-Cadherin			Vimentin		
	Cases (*n*)	**Median (IQR)****	***p***	**Cases (*n*)**	**Median (IQR)****	***p***
Disease Recurrence						
Yes	38	1 (1-1)	0.431	37	0 (0-0)	0.701
No	30	1 (1-1)		29	0 (0-0)	
TRG*						
2	3	1 (1-1)	0.679	3	0 (0-0)	0.688
3	12	1 (1-1)		12	0 (0-0)	
4	31	1 (1-1)		29	0 (0-0)	
5	11	1 (1-1)		11	0 (0-0)	
LVI						
Yes	49	0 (1-1)	0.203	47	0 (0-0)	0.141
No	19	1 (1-1)		19	0 (0-0)	
PNI						
Yes	11	1 (1-1)	0.369	10	0 (0-0)	0.372
No	57	1 (1-1)		56	0 (0-0)	
Serosal Involvement						
Yes	23	1 (1-1)	0.075	22	0 (0-0)	1.000
No	45	1 (1-1)		44	0 (0-0)	
HER2 Status						
Positive	8	1 (1-1)	0.765	8	0 (0-0)	0.765
Negative	26	1 (1-1)		26	0 (0-0)	
Siewert Classification						
I	21	1 (1-1)	0.768	21	0 (0-0)	0.342
II	22	1 (1-1)		22	0 (0-0)	
III	23	1 (1-1)		21	0 (0-0)	
T Stage						
T0	0	1 (1-1)	0.525	0	0 (0-0)	0.634
Tis, T1-2	8	1 (1-1)		8	0 (0-0)	
T3-4	60	1(1-1)		58	0 (0-0)	
N Stage						
N0	21	1 (1-1)	0.171	21	0 (0-0)	0.229
N+	47	1 (1-1)		45	0 (0-0)	

* Cases with TRG 1 were excluded from the study as they fall under the category of (y)pT0. ** Interquartile Range (Quartile 1 – Quartile 3)

Table 3.14 Correlation Between Age, Tumour Size and miRNA Expression.

	EpCAM	Oct-4	E-Cadherin	Vimentin	ALDH1A1 Total	CD133 Percentage	CD133 Topology
Age at Diagnosis							
Pearson correlation (*r*)	0.043	0.002	0.199	-0.022	0.092	-0.233	-0.166
Number (*n*)	66	68	67	66	65	70	70
Significance (*p*)	0.733	0.986	0.104	0.864	0.468	0.053	0.170
Macroscopic Tumour Size							
Spearman correlation (ρ)	0.109	-0.114	-0.083	0.024	0.122	0.166	0.195
Number (*n*)	60	63	62	60	59	64	64
Significance (*p*)	0.408	0.375	0.522	0.853	0.359	0.191	0.123

3.5 Survival Outcome Analysis

3.5.1 Relationship Between Expression Data and Overall Survival

Overall survival (OS) was measured from the date of diagnosis to date of death. An exact date of death was not obtainable for 5 patients; here instead the 'date lost to follow up' was used as a surrogate end date for purposes of survival analyses. The end date for measurement of survival outcomes was August 6th 2019. EpCAM expression was unsuitable for inclusion in the analysis due to an unequal distribution of cases across all 4 categories of staining intensity. Similarly, CD133 percentage staining was excluded from this analysis as cases were unevenly distributed across 5 categories. Of note, whilst ALDH1A1 intensity was included as a parameter in survival analyses, only the total ALDH1A1 was validated for use in data interpretation (Yang et al., 2018).

Multivariate analysis using a Cox regression model demonstrated a significant correlation between OS and male gender [HR 0.233, 95%CI 0.08-0.677, p=0.007], miR-16 expression [HR 0.381, 95%CI 0.187-0.776, p=0.008], miR-221 expression [HR 2.675, 95%CI 1.098-6.516, p=0.03], Oct-4 positivity [HR 3.78, 95%CI 1.541-9.273, p=0.004] and ALDH1A1 total score [HR 2.059, 95%CI 1.349-3.143, p<0.001] (Table 3.15, Figure 3.19). Male gender and increased expression of miR-16 were associated with a decreased risk of death from GEJA. By comparison, increased miR-221 expression, Oct-4 positivity and a high ALDH1A1 total were associated with an increased risk of death from GEJA. Figure 3.18 demonstrates that disease recurrence was associated with a worse probability of survival (p=0.0018).

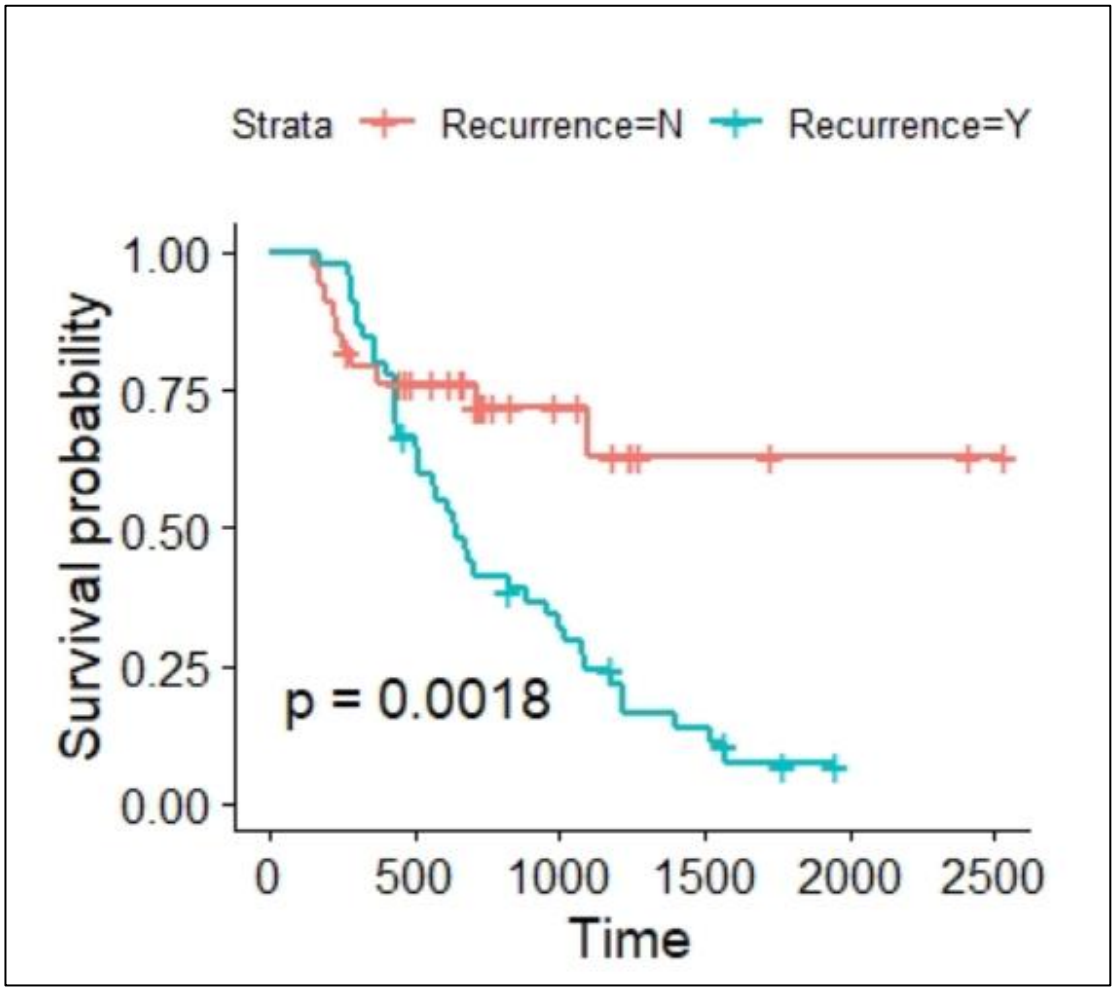


Figure 3.18 Kaplan Meier survival chart for OS with disease recurrence in 79 patients. Patients with disease recurrence have a significantly degreased overall survival (p=0.0018). Time is graphed in days.

Table 3.15 Multivariate Hazard Ratios for Overall Survival.

	Number (*n*)	HR (95% CI)	*p* value
Sex			
Female	10		
Male	69	0.233 (0.08-0.677)	**0.007**
Age at Diagnosis*		1.018 (0.977-1.06)	0.401
NANOG*		0.507 (0.211-1.214)	0.127
POU5F1*		1.903 (0.645-5.614)	0.244
VIM*		1.623 (0.579-5.544)	0.357
CDH1*		0.063 (0.004-1.014)	0.051
miR-16*		0.381 (0.187-0.776)	**0.008**
miR-17*		2.324 (0.858-6.294)	0.097
miR-21*		0.976 (0.608-1.568)	0.92
miR-221*		2.675 (1.098-6.516)	**0.03**
miR-10b*		0.834 (0.217-3.211)	0.792
miR-223*		0.022 (0.0001-3.491)	0.14
miR-200a*		0.739 (0.366-2.691)	0.398
miR-141*		0.956 (0.339-2.691)	0.932
miR-224*		0.368 (0.122-1.11)	0.076
Oct-4#			
0	63		
1	16	3.78 (1.541-9.273)	**0.004**
E-cadherin			
0	4		
1	75	0.37 (0.053-2.595)	0.317
Vimentin			
0	76		
1	3	6.306 (0.678-58.679)	0.106
ALDH1A1 Intensity Score			
0	8		
1	47	0.035 (0.003-0.37)	**0.005**
2	22	0.007 (0.0003-0.142)	**0.001**
3	2	0.004 (2.63e-05-0.492)	**0.025**
ALDH1A1 Total	79	2.059 (1.349-3.143)	**<0.001**
CD133 Topology			
0	45		
1	22	1.593 (0.603-4.211)	0.348
2	12	1.814 (0.587-5.608)	0.301

* All 79 patients were included in this analysis. # Labelled as 'POU5F1.1' in Figure 3.18. *HR = hazard ratio; 95% CI = 95% confidence interval.*

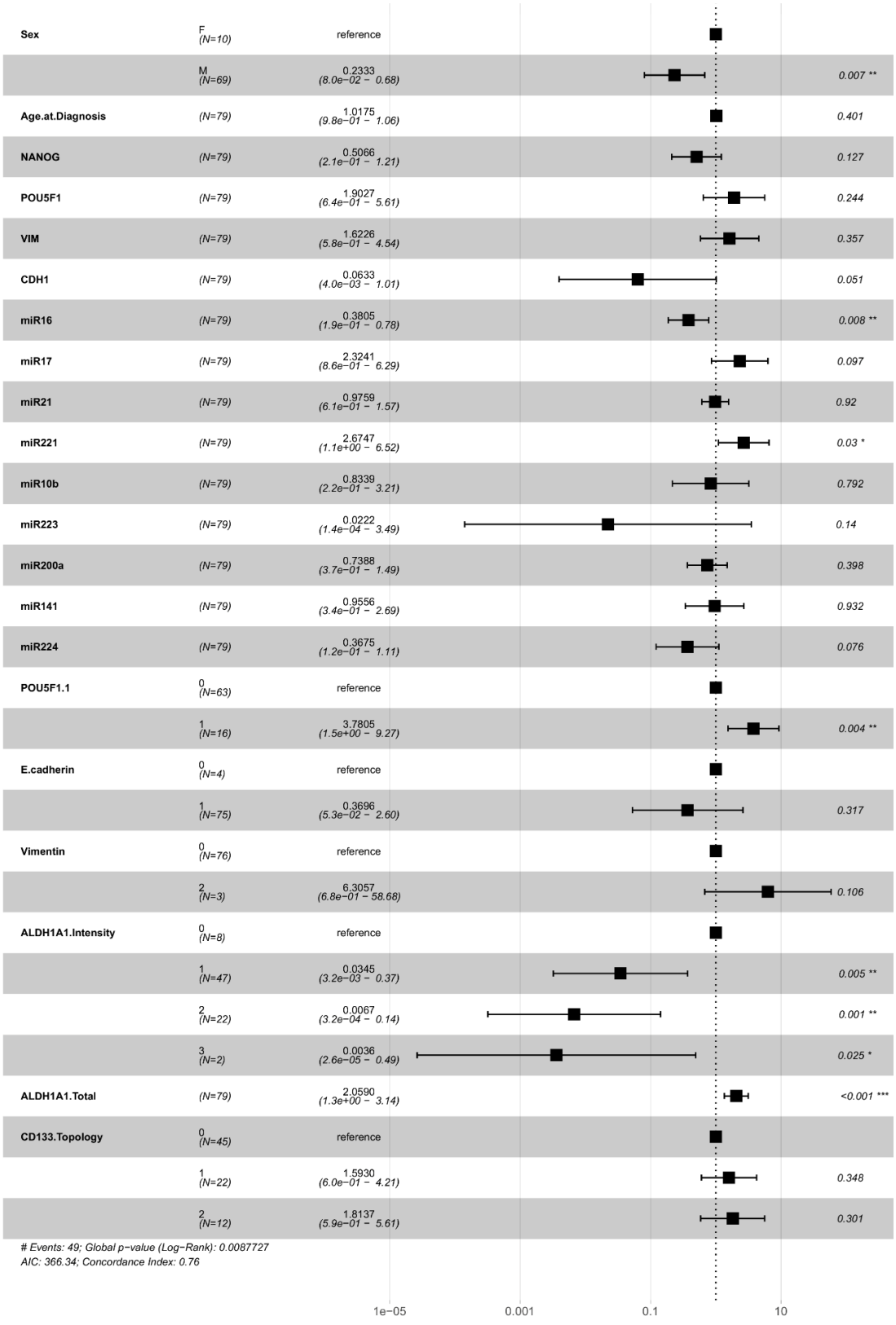


Figure 3.19 Forest plot based on multivariate hazard ratios from Cox regression analysis for OS. Statistical significance: $p<0.05$.

3.5.2 Relationship Between Expression Data and Disease Free Survival

Disease free survival (DFS) was measured from the date of surgery to date of disease relapse, where present. The end date for measurement of survival outcomes was August 6th 2019. EpCAM expression and CD133 percentage staining were unsuitable for inclusion in the analysis due to an unequal distribution of cases across all categories. As discussed in Section 3.5.1, ALDH1A1 total score was considered in data interpretation, not the ALDH1A1 intensity score.

Multivariate analysis using a Cox regression model demonstrated a significant correlation between DFS and male gender [HR 0.194, 95%CI 0.066-0.573, p=0.003], increased CDH1 expression [HR 0.005, 95%CI 0.0002-0.114, p<0.001], increased Vimentin mRNA expression [HR 3.367, 95%CI 1.002-11.318, p=0.05], increased miR-221 expression [4.185, 95%CI 1.577-11.104, p=0.004], Oct-4 positivity [HR 3.992, 95%CI 1.511-10.547, p=0.005], ALDH1A1 total score [HR 2.016, 95%CI 1.353-3.005, p<0.001] and CD133 staining in clusters of neoplastic cells [HR 3.066, 95%CI 1.002-9.381, p=0.05] (Table 3.16, Figure 3.20). Male gender and CDH1 expression were associated with a decreased risk of disease recurrence in GEJA, whilst Vimentin mRNA expression, miR-221 expression, Oct-4 positivity, total ALDH1A1 score and CD133 positivity in clusters were associated with an increased risk of disease recurrence in GEJA.

Table 3.16 Multivariate Hazard Ratios for Disease Free Survival.

	Number (*n*)	HR (95% CI)	P Value
Sex			
Female	10		
Male	69	0.194 (0.066-0.573)	**0.003**
Age at Diagnosis*		0.997 (0.956-1.039)	0.879
NANOG*		0.616 (0.265-1.429)	0.259
POU5F1*		2.085 (0.718-6.058)	0.177
VIM*		3.367 (1.002-11.318)	**0.05**
CDH1*		0.005 (0.0002-0.114)	**<0.001**
miR-16*		0.291 (0.12-0.705)	0.006
miR-17*		3.155 (0.986-10.095)	0.053
miR-21*		0.943 (0.598-1.487)	0.802
miR-221*		4.185 (1.577-11.104)	**0.004**
miR-10b*		1.369 (0.354-5.433)	0.655
miR-223*		0.018 (0.0001-2.936)	0.123
miR-200a*		0.788 (0.389-1.595)	0.508
miR-141*		0.807 (0.293-2.223)	0.678
miR-224*		0.355 (0.113-1.111)	0.075
Oct-4#			
0	63		
1	16	3.992 (1.511-10.547)	**0.005**
E-cadherin			
0	4		
1	75	0.666 (0.108-4.093)	0.661
Vimentin			
0	76		
1	3	7.531 (0.862-65.781)	0.068
ALDH1A1 Intensity Score	8		
0	47	0.012 (0.001-0.139)	**<0.001**
1	22	0.003 (0.0001-0.065)	**<0.001**
2	2	0.029 (0.0002-3.362)	0.144
3			
ALDH1A1 Total	79	2.016 (1.353-3.005)	**<0.001**
CD133 Topology			
0	45		
1	22	2.982 (0.856-6.173)	0.099
2	12	3.066 (1.002-9.381)	**0.05**

* All 79 patients were included in this analysis. # Labelled as 'POU5F1.1' in Figure 3.19. *HR = hazard ratio; 95% CI = 95% confidence interval.*

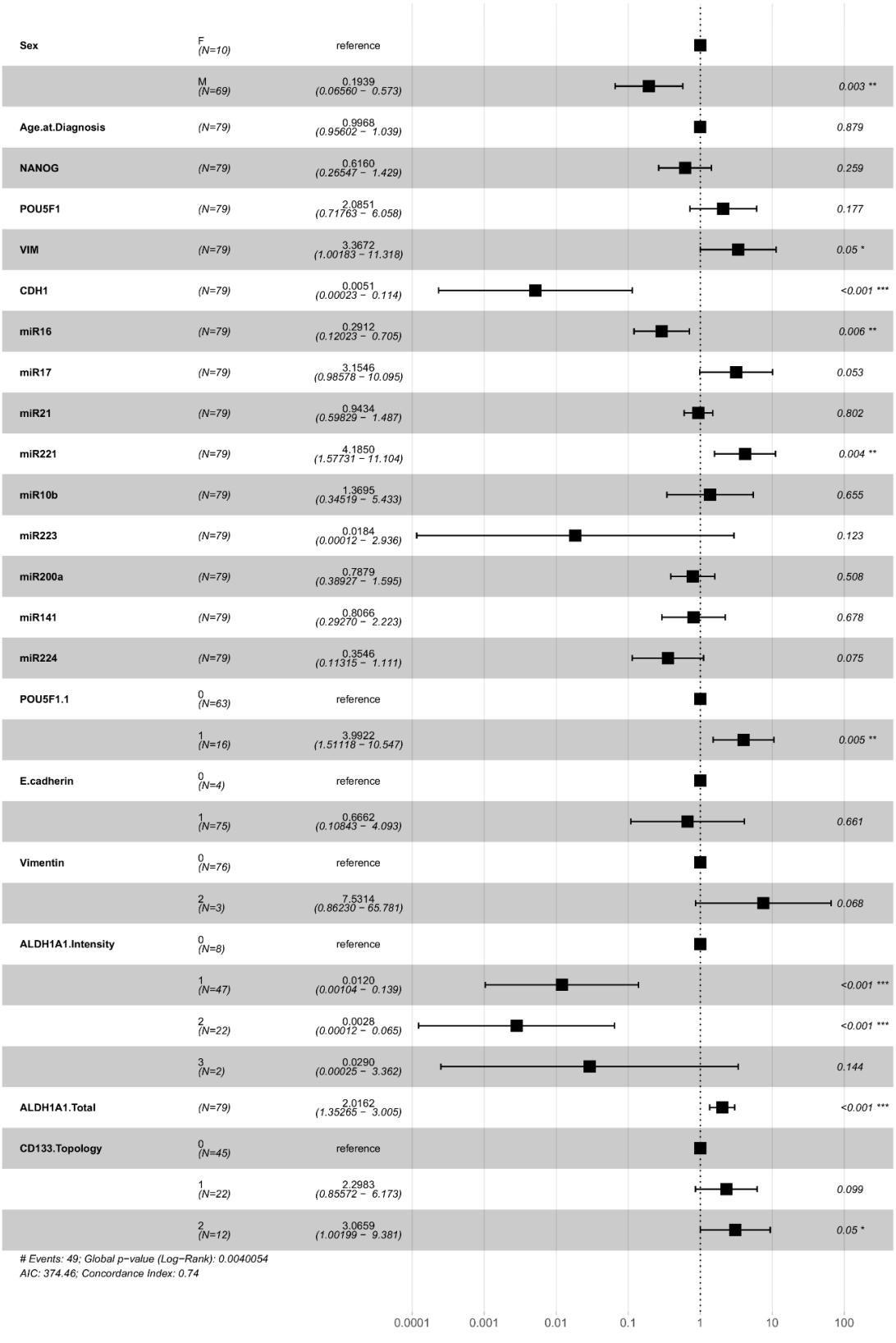


Figure 3.20 Forest plot based on multivariate hazard ratios from Cox regression analysis for DFS. Statistical significance: $p<0.05$.

3.6 Comparison of Biological Expression Profiles in Tumours with and without Neoadjuvant Treatment

Sixty nine of 79 patients included in in this study received NAT before surgery. As detailed in Table 3.1, most patients received either MAGIC (26/69) or CROSS (18/69) neoadjuvant regimens. Paired data was analysed using a Mann-Whitney U test, with no significant findings (Table 3.17, Appendix 2.1). However, due to the unequal ratio of patients with and without NAT, in addition to the variability of NAT regimens received, no conclusions can be drawn from this data.

Table 3.17 Correlation Between NAT Status and Clinicopathological Characteristics.

	Number (*n*)	Median (IQR)*	*p*
NANOG			
Neoadjuvant	69	1.255 (0.403-2.871)	0.312
No Neoadjuvant	10	1.861 (1.072-3.03)	
Total	79		
POU5F1			
Neoadjuvant	69	1.861 (0.521-4.273)	0.696
No Neoadjuvant	10	1.639 (1.042-4.399)	
Total	79		
Vimentin			
Neoadjuvant	69	0 (0-0.053)	0.808
No Neoadjuvant	10	0 (0-0.734)	
Total	79		
CDH1			
Neoadjuvant	69	0 (0-0)	0.412
No Neoadjuvant	10	0 (0-0)	
Total	79		
Serpine1			
Neoadjuvant	69	0 (0-0)	0.626
No Neoadjuvant	10	0 (0-0)	
Total	79		
miR-16			
Neoadjuvant	69	0 (0-0.405)	0.508
No Neoadjuvant	10	0 (0-0)	
Total	79		
miR-17			
Neoadjuvant	69	0 (0-0)	0.517
No Neoadjuvant	10	0 (0-0)	
Total	79		
miR-21			
Neoadjuvant	69	0.052 (0-0.311)	0.925
No Neoadjuvant	10	0.111 (0-0.34)	
Total	79		
miR-221			
Neoadjuvant	69	0 (0-0.502)	0.074
No Neoadjuvant	10	0 (0-0)	
Total	79		
miR-10b			
Neoadjuvant	69	0 (0-0)	0.371
No Neoadjuvant	10	0 (0-0)	
Total	79		
miR-223			
Neoadjuvant	69	0 (0-0)	0.780
No Neoadjuvant	10	0 (0-0)	
Total	79		
miR-200a			
Neoadjuvant	69	0.291 (0-1)	0.919
No Neoadjuvant	10	0.519 (0.006-0.751)	
Total	79		

Table 3.17 Correlation Between NAT Status and Clinicopathological Characteristics.

miR-141 Neoadjuvant No Neoadjuvant Total	 69 10 79	 0.402 (0-2.677) 1.350 (0-3.345)	0.574
miR-224 Neoadjuvant No Neoadjuvant Total	 69 10 79	 0 (0-0.990) 0.054 (0-5.260)	0.438
EpCAM Neoadjuvant No Neoadjuvant Total	 69 10 79	 2 (1-2) 2 (1-2)	0.698
Oct-4 Neoadjuvant No Neoadjuvant Total	 69 10 79	 0 (0-1) 0 (0-0)	0.076
E-cadherin Neoadjuvant No Neoadjuvant Total	 69 10 79	 1 (1-1) 1 (1-1)	0.424
Vimentin Neoadjuvant No Neoadjuvant Total	 69 10 79	 0 (0-0) 0 (0-0)	0.313
ALDH1A1 Total Neoadjuvant No Neoadjuvant Total	 69 10 79	 4.5 (3-6) 5 (4-6)	0.443
CD133 % Neoadjuvant No Neoadjuvant Total	 69 10 79	 0 (0-2) 0 (0-0)	0.114
CD133 Topology Neoadjuvant No Neoadjuvant Total	 69 10 79	 1 (0-1) 0 (0-0)	0.067

Chapter Four: Utility of Combined mRNA, miRNA and Protein Expression Data as a Predictive Tool

4.1 Ability of Combined Expression Data to Predict Tumour Classification

A random forest plot model was constructed using all mRNA, miRNA and protein data to determine if specific clinical outcomes could be predicted. All results were cross validated to ensure that they were not overfit and that they were generalizable to other patient cohorts. miR-103a, miR-203a, miR-133b and CD34 were excluded from this analysis as no sample displayed any expression.

The ability of mRNA, miRNA and protein expression data, taken in combination, to predict if a tumour sample is ' lower grade' or 'higher grade' was first assessed. The model had a classification accuracy of 66.42% and a Receiver Operating Characteristic Curve (ROC) of 0.61 (Figure 4.1). Table 4.1 shows the breakdown of the true versus predicted labels from the model for 'lower grade' versus 'higher grade' samples. The model is more sensitive at predicting 'higher grade' samples, with an 83.5% accuracy rate. This is compared to a 43.1% accuracy rate when predicting 'lower grade' samples.

Table 4.1 Breakdown of True versus Predicted Labels for 'Lower Grade' and 'Higher Grade' Samples.

		Predicted Label	
True Label		Lower Grade	Higher Grade
	Lower Grade	66	13
	Higher Grade	33	25

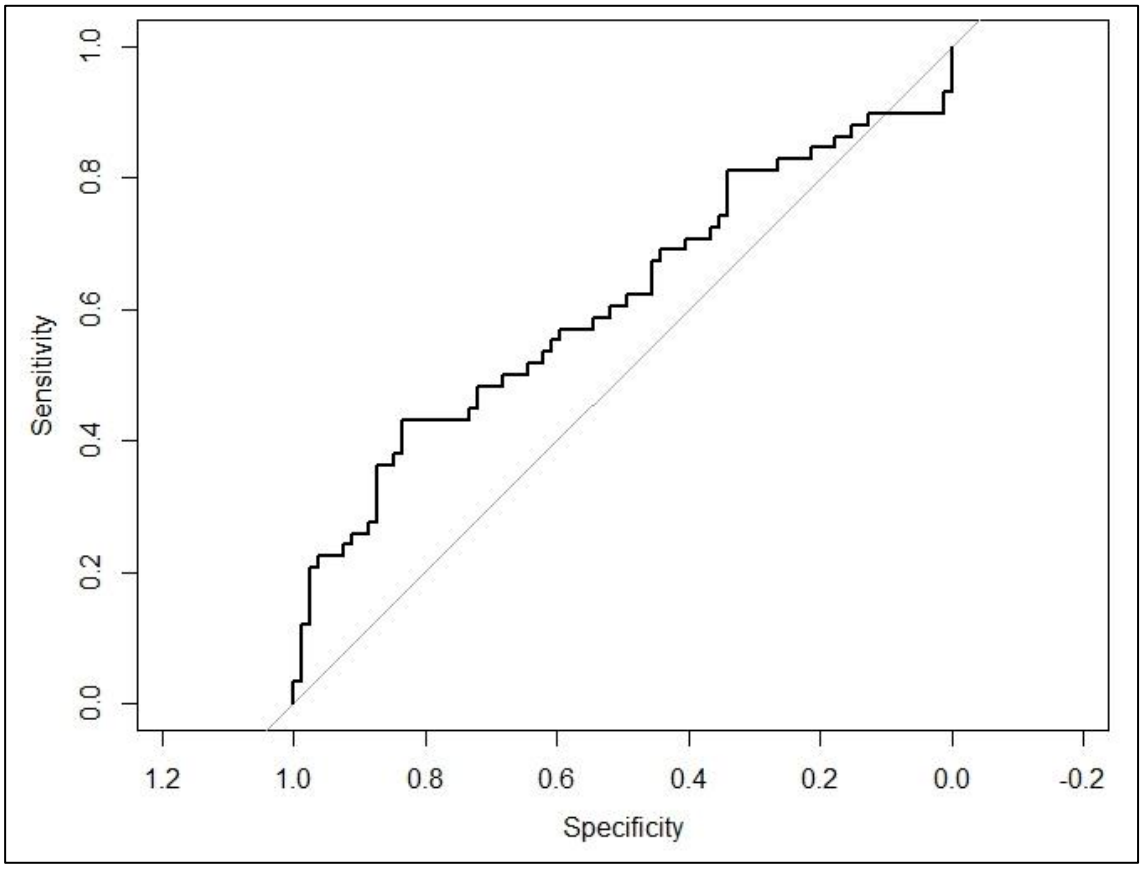


Figure 4.1 ROC curve of the ability of the combined expression model to predict 'lower grade' versus 'higher grade' tumour samples. The ROC is 0.61.

Figure 4.2 shows the breakdown of variable importance scores for each of the predictor variables included in the model. The variables at the top of the model are ranked in order of their importance in making the prediction, and vice versa with variables at the bottom. EpCAM was the most important predictor variable in this model.

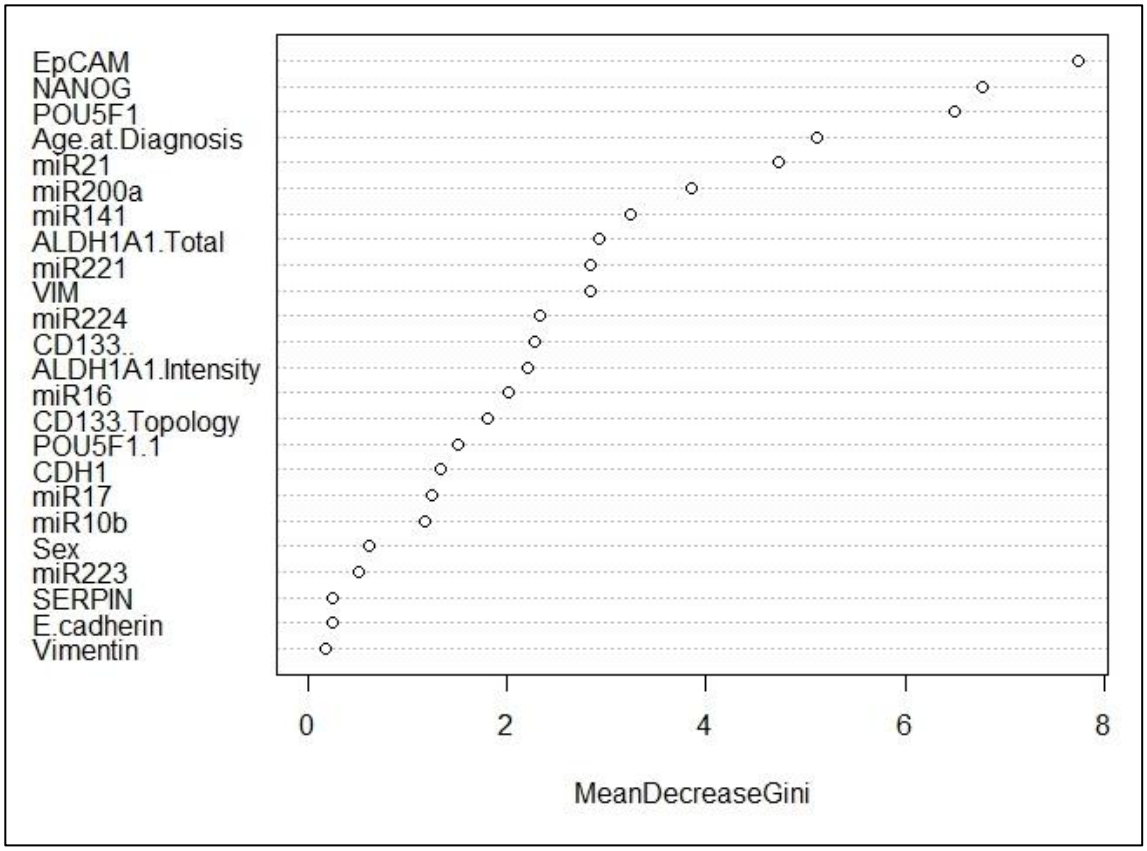


Figure 4.2. Variable importance scores for the combined expression data model in predicting 'lower grade' versus 'higher grade' tumour samples. EpCAM, NANOG expression and POU5F1 expression are the most important parameters in achieving predictive accuracy. By contrast, Serpine1, E-cadherin and Vimentin protein expression contribute the least to this prediction model. VIM = vimentin mRNA; POU5F1.1 = Oct-4 protein; CD133.. = CD133 percentage staining score.

4.2 Ability of Combined Expression Data to Predict Disease Recurrence

The same model was used as in Section 4.1 to predict disease recurrence. The model had a classification accuracy of 69.34% and a ROC of 0.73 (Figure 4.3). Table 4.2 shows the breakdown of the true versus predicted labels from the model for disease recurrence. As was the case with predicting 'lower grade' versus 'higher grade' samples, the classification accuracy of the model is superior when predicting cases with disease recurrence (81.01%) than those without (53.5%).

Table 4.2 Breakdown of True versus Predicted Labels for Disease Recurrence.

		Predicted Label	
True Label		No	Yes
	No	31	27
	Yes	15	64

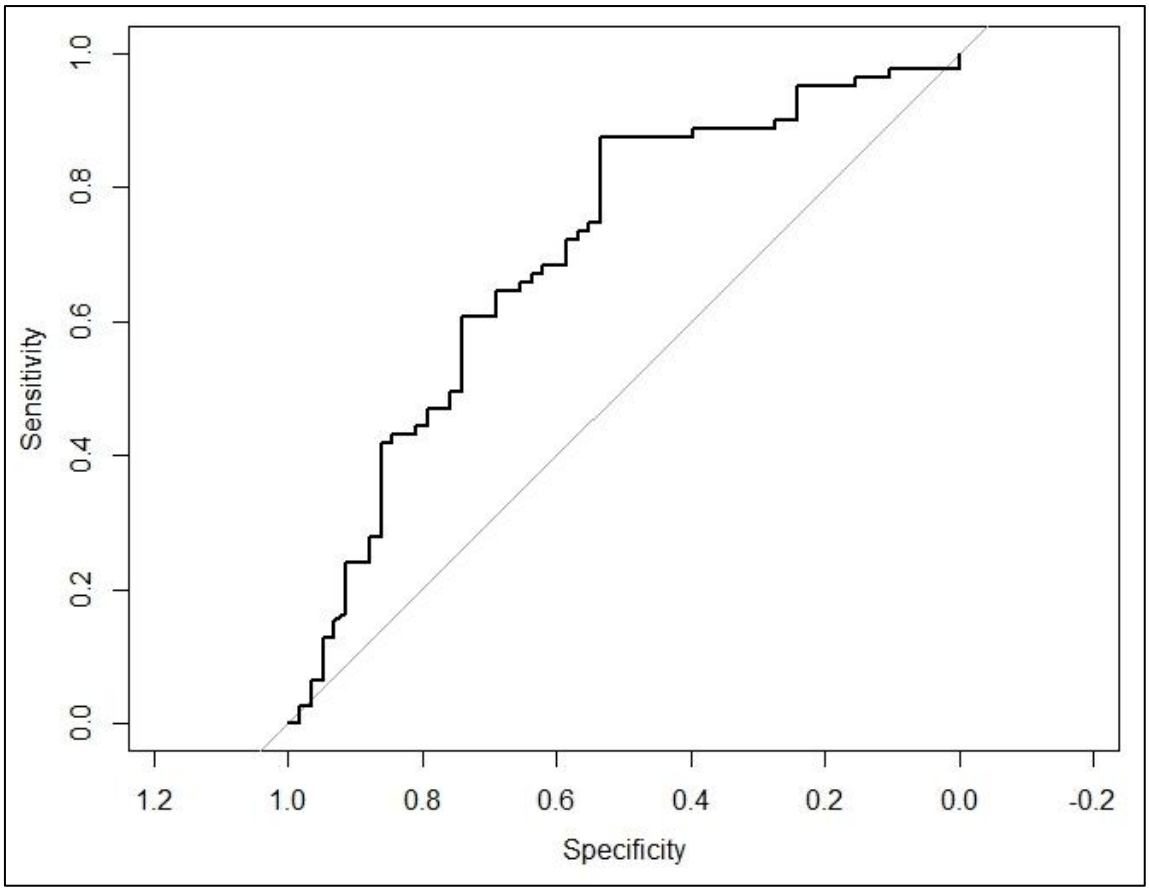


Figure 4.3 ROC curve of the ability of the combined expression model to predict disease recurrence. The ROC is 0.73.

Figure 4.4 shows the breakdown of variable importance scores for each of the predictor variables included in the model. Age at diagnosis was the most important predictor variable in the model, yet NANOG was the most important predictive biomarker.

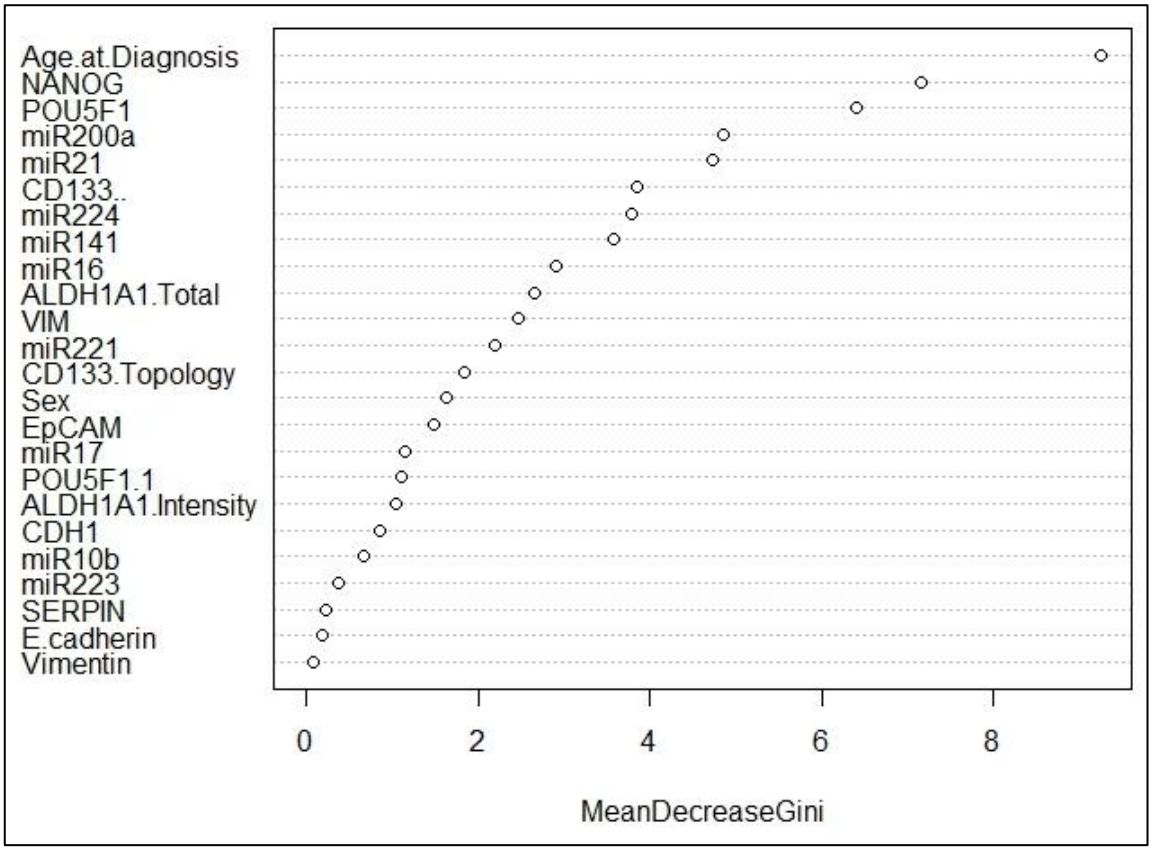


Figure 4.4. Variable importance scores for the combined expression data model in predicting disease recurrence. Age at diagnosis, NANOG expression and POU5F1 expression are the most important parameters in achieving predictive accuracy. By contrast, Serpine1, E-cadherin and Vimentin protein expression contribute the least to this prediction model. VIM = vimentin mRNA; POU5F1.1 = Oct-4 protein; CD133.. = CD133 percentage staining score.

4.3 Ability of Combined Expression Data to Predict Tumour Regression Grade (TRG)

The same model was used again to predict TRG. The model was not accurate in predicting TRG classification. This is due to the presence of four separate categories, of which TRG2 and TRG3 are not well populated. The predictive accuracy of the model is 50.44%. The breakdown of the true versus predicted labels is shown in Table 4.3. The model predicted that the respondent was group 4 in 87% of cases, thus highlighting its predictive inaccuracy. The ROC curve and variable importance scores were not reported because the model was not accurate.

Table 4.3 Breakdown of True versus Predicted Labels for Tumour Regression Grade.

	Predicted Label				
True Label		TRG2	TRG3	TRG4	TRG5
	TRG1	0	1	2	0
	TRG2	0	3	20	2
	TRG3	0	7	48	5
	TRG4	0	2	17	6

4.4 Bioinformatic Interrogation of Signalling Pathways

DAVID (Database for Annotation, Visualization and Integrated Discovery) is an online open access functional annotation tool which can be used to generate signalling pathways from a list of gene targets. The programme requires the genes to be input according to their Entrez or Ensembl code. From this a KEGG (Kyoto Encyclopaedia of Genes and Genomes) pathway is generated, displayed as a schematic diagram of the molecular targets within a signalling pathway.

miRWalk 3.0 is an online open source bioinformatic computational database that is used to generate both predicted and validated downstream miRNA targets in humans, mice, rats, dogs and cows. This tool was used to generate a list of experimentally validated gene targets for our panel of 12 miRNAs (Appendix 3.1). A correlation matrix was generated to show the main overlapping miRNA:target interactions within our panel (Table 4.4). The 6 most common mRNA targets were the Activin A receptors (ACVR), multiple types; Bone Morphogenetic Protein Receptor Type 1B (BMPR1B), Wnt family members, SMAD family members, the KRAS proto-oncogene and Polycomb group ring finger (PGFR), multiple types.

Table 4.4 Correlation Matrix of Overlapping miRNA : Target Interactions.

	ACVR	BMPR1B	WNT	SMAD	KRAS	PCGF
miR-16	X	X	X		X	
miR-17	X	X	X	X		X
miR-21						X
miR-221	X		X	X		
miR-223			X	X		X
miR-224	X		X	X		X
miR-10b	X	X	X	X		
miR-203a		X	X	X	X	X
miR-200a	X		X	X	X	X
miR-141	X		X	X		

‘Target mining’ was first performed, whereby the list of 12 candidate miRNAs, as identified by their accession number, was input into the miRWalk programme. The target genes for all miRNAs shown to be of significance in GEJA were each individually input into DAVID to determine their relevant KEGG pathways. The results were interrogated to seek any overlap between clinicopathological outcomes, gene targets and signalling pathways between the different miRNAs. Overlapping miRNA:target interactions within a signalling pathway were manually input into the same KEGG pathway. Signalling pathways regulated by miR-21, miR-221, miR-141, miR-224 and miR-16 were individually investigated.

All 5 miRNAs were implicated in regulation of the ‘signalling pathway regulating pluripotency of stem cells’ (Figure 4.5). miR-141, miR-16, miR-224 and miR-221 were additionally involved in regulation of the TGF-β signalling pathway (Figure 4.6). miR-141, miR-224 and miR-16 also played a role in regulating the Hippo signalling pathway (Figure 4.7). miR-224 was implicated in the Wnt signalling pathway (Figure 4.8).

The genes included in the mRNA expression analysis study were involved in the ‘signalling pathway regulating pluripotency of stem cells’ (NANOG and POU5F1) and hippo signalling pathway (E-cadherin and Serpine-1 (PAI-1)). miR-221 is involved in regulating expression of POU5F1 (Figure 4.5). No other miRNAs in our cohort were directly related to the control of expression of any of these genes. Despite little direct interaction between the genes and miRNAs in our study panels however, their molecular functions within these signalling pathways nevertheless overlap greatly.

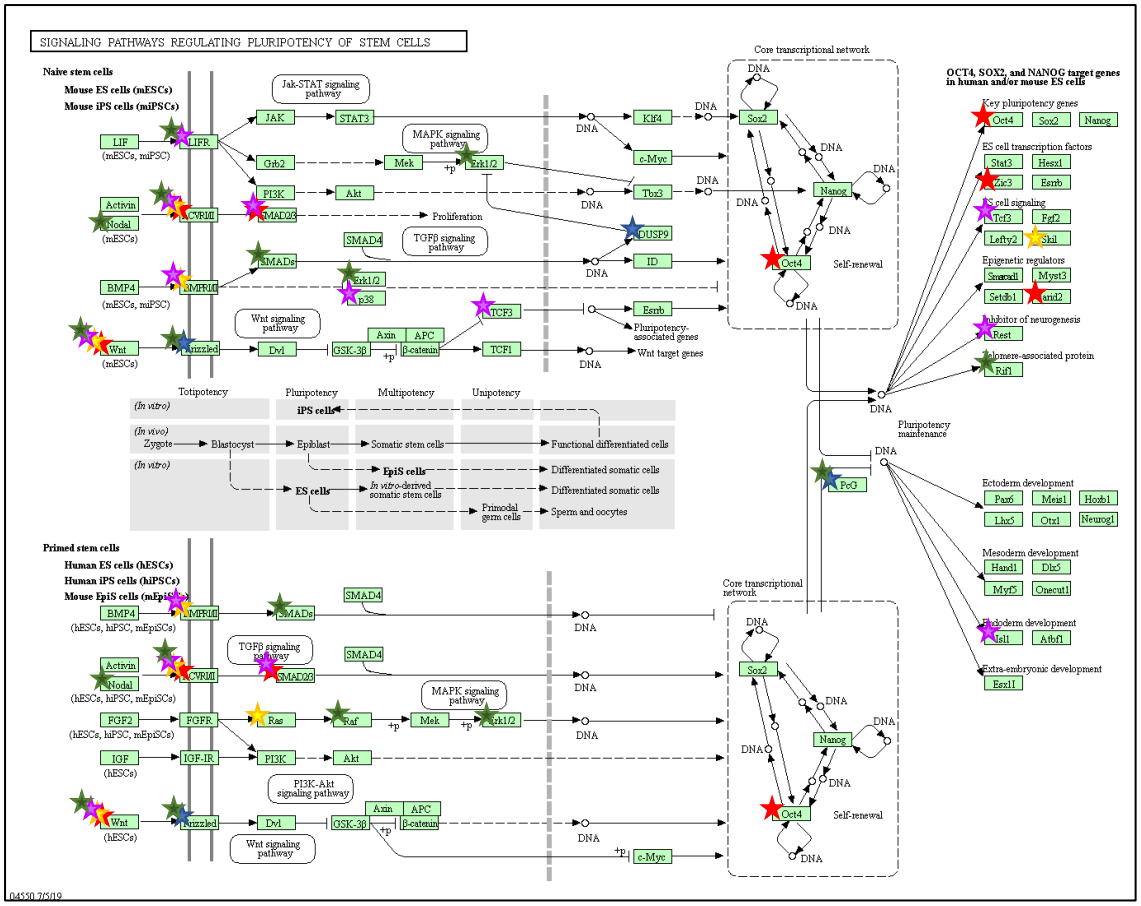


Figure 4.5 Signalling pathways for regulating pluripotency of stem cells. This image provides a visual depiction of the molecular pathways regulated by miR-21 (blue stars), miR-221 (red stars), miR-16 (yellow stars), miR-141 (purple stars) and miR-224 (green stars). These miRNAs were all found to target genes within the same commonly dysregulated pathways involved in the acquisition and regulation of stemness properties.

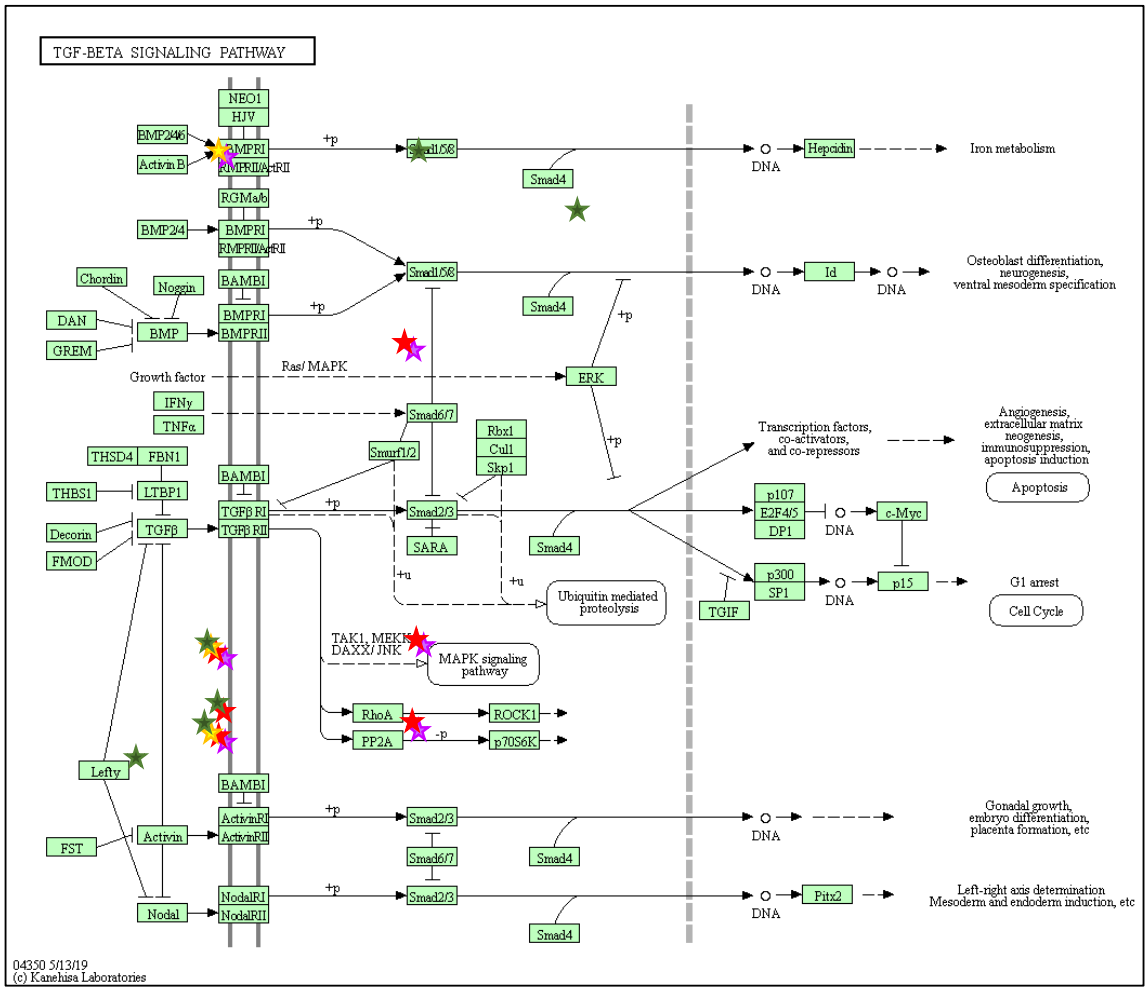


Figure 4.6 TGFβ signalling pathway. This image provides a visual depiction of the molecular pathways regulated by miR-221 (red stars), miR-16 (yellow stars), miR-141 (purple stars) and miR-224 (green stars). These miRNAs were all found to target genes within the same pathways involved in EMT and stemness.

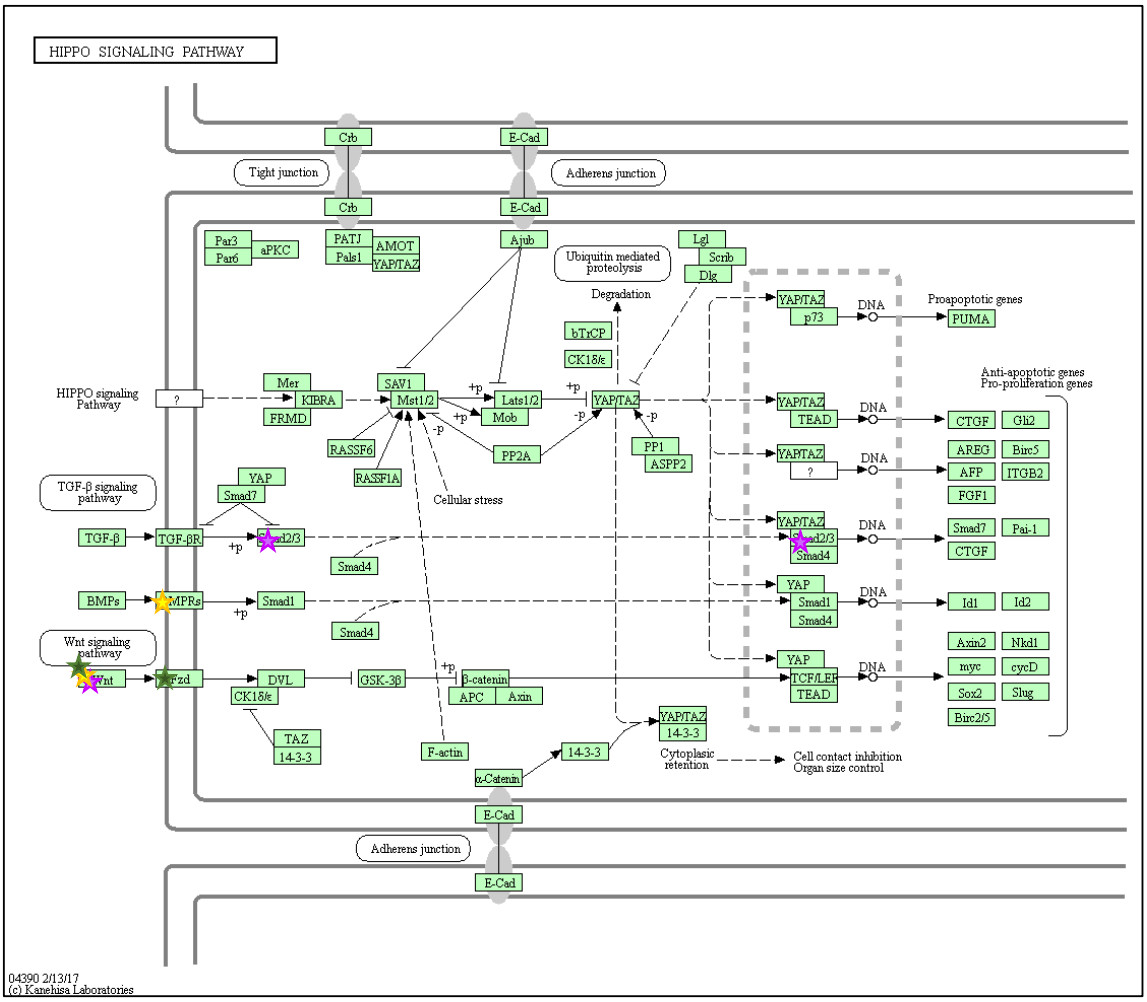


Figure 4.7 Hippo signalling pathway. This image provides a visual representation of the molecular pathways regulated by miR-16 (yellow stars), miR-141 (purple stars) and miR-224 (green stars). These miRNAs were all found to target genes within the same pathways involved in human cancer.

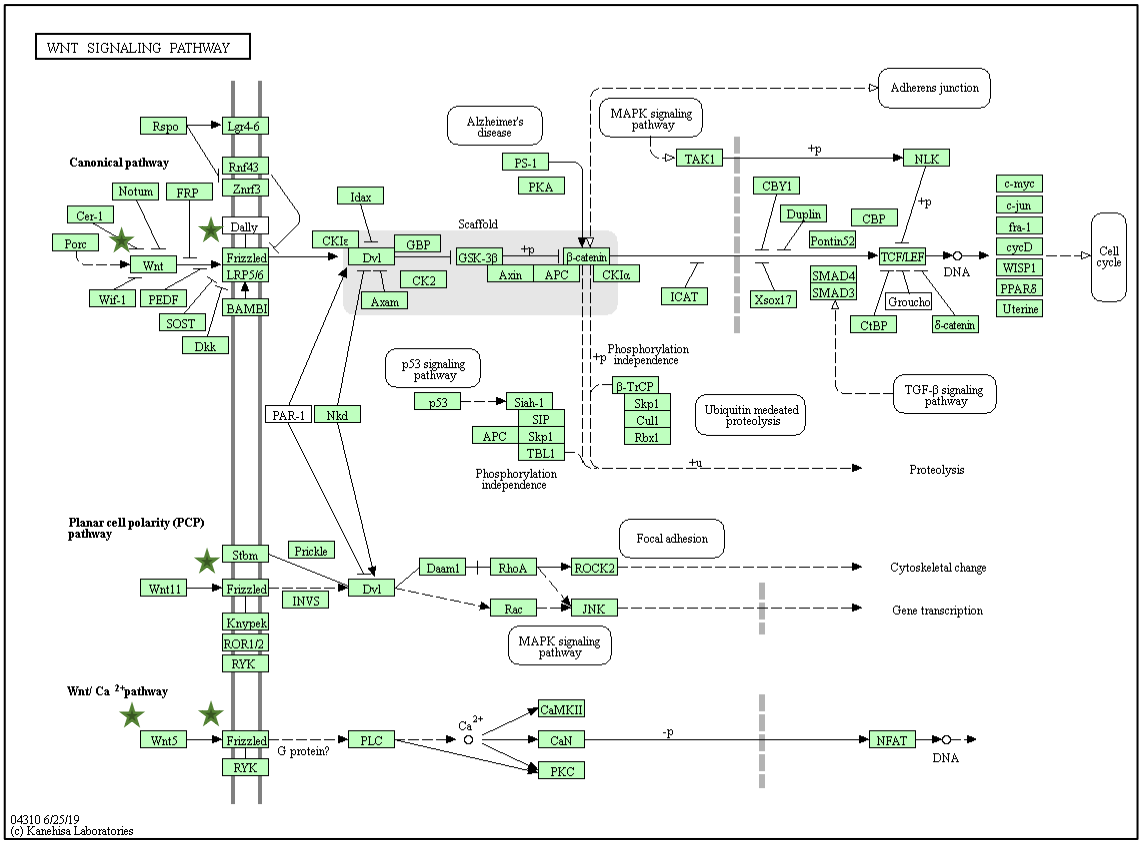


Figure 4.8 Wnt signalling pathway. This network provides a visual depiction of the mRNA targets regulated by miR-224 (green stars) in multiple signalling pathways.

Chapter Five: Discussion

The incidence of GEJA has increased worldwide; available treatments for this malignancy are limited, rates of drug resistance are high and survival outcomes are poor (Brungs et al., 2019; NCRI, 2019; SEER, 2019). Whilst a large body of knowledge exists surrounding the molecular biology of both OAC and gastric adenocarcinoma, it is important that the molecular profile of GEJA is assessed separately due to the growing opinion that these tumours are best regarded as a distinct entity (Hayakawa et al., 2016). This study sought to confirm and define the role of EMT and CSCs in GEJA, given the roles they play role in tumorigenesis, metastasis and drug resistance in a number of solid organ malignancies (T. Chen et al., 2017; Eun et al., 2017; S. Li & Li, 2014; Mani et al., 2008; Medema, 2013; Nunes et al., 2018).

A multi-omics approach was used to refine the current knowledge regarding the specific CSC and EMT markers expressed in GEJA, thorough interrogation of multiple mRNA, miRNA and protein expressions. The hypothesis that is central to this thesis states that a specific expression pattern of mRNA, miRNA and protein markers exists which characterises ‘aggressive’ tumours and can thus be used to aid in the identification of advanced disease in the clinical setting. Taking into consideration the well documented heterogeneity present within these tumours (Pectasides et al., 2018), this study additionally investigated the expression patterns of separate ‘higher grade’ and ‘lower grade’ areas within each tumour, as identified based on their histological appearances. It was proposed that poorly differentiated tumour cells present at the infiltrative tumour edge are most likely to be CSCs and will thus express a stem-like molecular profile and a mesenchymal phenotype, whilst better differentiated and less infiltrative areas will not contain CSCs and will therefore express an epithelial phenotype and lower levels of stemness markers.

The patient cohort included in this study was comprised predominantly of males, with a median age of 67 years. Interestingly, male gender was associated with a better OS and DFS, as compared with females. Despite this the clinical outcomes across all patients were poor, as is in keeping with the known aggressive clinical course of GEJA. Fifty-seven percent of patients had a relapse of their disease within the timeframe of the study, with a median DFS of 216 days (range 45-1347). OS was very poor, with a median of 510 days (range 150-1575). Of patients who received their initial diagnosis over 5 years before the study end date, only 19% (4/21) survived for 5 years. This figure is in keeping with the national and international 5 year survival rates of 22.6% and 19.9% respectively, and reflects the aggressive clinical course of these malignancies (NCRI, 2019; SEER, 2019).

The majority of patients included in this study received NAT prior to surgical excision of their tumour. This trend is consistent with the current literature: at least 14 separate studies to date have compared the impact of pre-operative chemotherapy followed by surgery to surgery alone in GEJA, with overall results supporting the use of NAT (Altorki & Harrison, 2017). Of all NAT regimens administered in this study, MAGIC was the most common, followed by CROSS. EOX and FLOT together accounted for a small minority of cases. Twenty patients classified as 'other' either received a different treatment regimen not listed above, or else the details of their regimen were not identified upon review of the clinical records. 36.7% of patients received adjuvant therapy following surgical excision, which may have had an impact on their survival outcomes. However, the details of these adjuvant therapies were not collected.

Response to treatment was variable within the post-NAT patient cohort, which is in keeping studies which previously described response to NAT in GEJA (Al-Batran et al., 2016; Cunningham et al., 2006; van Hagen et al., 2012). As previously stated, CROSS

and MAGIC were the two main regimens included in this study. Whilst this study did not examine the response to therapy for each individual regimen separately, previous studies have reported a CPR in 29% of patients who received the CROSS regimen, whilst pathologic findings from the MAGIC trial included a reduction in average tumour size from 5 cm in patients who underwent surgery alone to 3 cm in patients who additionally received NAT (Cunningham et al., 2006; van Hagen et al., 2012). These findings indicate an inconsistent response to NAT in GEJA.

This present study assessed the pathological response to NAT using the Mandard TRG system: a TRG score of 1/2 is considered to be a 'major' response, TRG 3/4 is a 'minor' response and TRG 5 is no response (Depypere et al., 2016). 72.5% of patients who received NAT had a minor response to treatment, whilst 20.2% had no response. Only 4.3% of the cases were classified as TRG 2. No case had a CPR (TRG 1) because all tumours included in this study by definition had residual tumour cells for the purposes of performing expression analyses. This data is therefore not entirely representative of the true response to NAT in GEJA as the dataset contains an over-representation of TRG 3-5 tumours.

There was no significant association between TRG and the differentiation status of the resected tumour by predominant area, an observation which is likely to be attributable to the distribution bias in the study cohort. Most tumours were of an advanced stage (TNM classification, 8^{th} edition): 84.8% had T3-4 stage disease and 70.9% had nodal metastasis. Only 3.8% of patients had histologically confirmed M1 disease, however this is not an accurate reflection of the true prevalence of distant metastases within this study population.

5.1 Gene Expression in GEJA

The study hypothesis states that ‘lower grade’ areas of tumour should express higher levels of genes associated with an epithelial phenotype, including CDH1, and have lower expression of genes associated with stemness, such as NANOG and POU5F1. By comparison, ‘higher grade’ areas should show greater expression of genes associated with a mesenchymal phenotype, such as Vimentin, and higher levels of genes related to stemness. Serpine-1 was included in the analysis due to its proposed role in metastasis. It was anticipated that the expression of mesenchymal and stemness markers in ‘higher grade’ samples would correlate with adverse clinical outcomes. Appendix 4.1 lists the significant associations identified with increased expression of each individual gene, as discussed below.

Vimentin

Vimentin was expressed in a minority of paired cases, of which only 38.1% showed the expected increase in expression from ‘lower grade’ to ‘higher grade’ paired tumour samples. Analysis of all ‘higher grade’ samples revealed that increased Vimentin gene expression was associated with a shorter DFS, yet it showed no association with OS. The link between Vimentin and poor DFS is in keeping with the known relationship between a mesenchymal phenotype and aggressive tumour characteristics, as has been consistently demonstrated across a range of malignancies (Bhardwaj et al., 2019; Tanaka et al., 2015; J. Y. Wang et al., 2018; S. Wu, Du, Beckford, & Alachkar, 2018).

Contrary to the study hypothesis, increased levels of Vimentin were shown to be significantly associated with a low TRG, and therefore with a good response to treatment.

From this it can be inferred that decreased Vimentin is associated with a high TRG and thus enhanced drug resistance. This finding was mirrored in a study by Huo *et al* which looked at the role of Vimentin in ovarian cancer. Here the authors demonstrated that silencing of Vimentin led to increased drug resistance, upregulation of CSC markers and acquisition of CSC traits (Huo et al., 2016). The authors proposed that Vimentin enhanced drug resistance via a prolonged G2 arrest, which provided more time for the tumour cells to repair their damaged DNA. However, a separate study which looked at the role of Vimentin in drug resistance in NSCLC showed the opposite result, demonstrating instead a correlation between increased Vimentin expression and sensitivity to the EGFR inhibitor Gefitinib (Hu et al., 2017). Taken together, these studies suggest that the role of Vimentin in promoting or preventing resistance to anti-cancer drugs varies between different malignancies.

Conventional chemotherapeutic regimens classically target more differentiated epithelial cells, which do not express Vimentin (Yao, Dai, & Peng, 2011). Taking this into consideration, the association between decreased Vimentin expression and high TRG grades identified in this study suggests that drug resistance in GEJA is regulated by mechanisms unrelated to EMT. A range of alternative mechanisms of drug resistance exist, including drug inactivation, DNA damage repair, cell death inhibition and epigenetic modifications (Housman et al., 2014). It is interesting to speculate that Vimentin induces a prolonged G2 phase in GEJA as in ovarian cancer cells, however further studies are required to determine the precise mechanism of drug resistance in Vimentin downregulated GEJA cancer cells.

CDH1

CDH1 was similarly expressed in only a small number of paired samples, of which the majority were downregulated as expected in 'higher grade' samples as compared to 'lower grade'. Increased CDH1 expression was significantly associated with a longer DFS, and strongly associated with a longer OS. This finding is supported by other studies in the literature which link decreased CDH1 expression with adverse clinical outcomes across a number of different malignancies, including both oesophageal and gastric cancers (Ishiguro et al., 2016; Z. Li & Guo, 2019; J. Liu et al., 2016).

NANOG

NANOG – named after the Land of Youth 'Tír na nÓg' in Celtic mythology (I. Chambers et al., 2003) – demonstrated the expected increase in expression from 'lower grade' to 'higher grade' paired samples in only a small majority of cases, which was of no significance. NANOG expression showed a strong association with both lymph node metastases and serosal involvement, yet there were no significant associations between its expression and any clinicopathological parameters or survival outcomes.

These findings are at odds with both the study hypothesis and with previous published studies on this subject. A recent meta-analysis of 33 studies demonstrated that NANOG is a useful prognostic tumour biomarker across a range of malignancies due to its association with poor survival outcomes and adverse clinicopathological features, however these studies used IHC to assess NANOG expression and are therefore not directly comparable to the findings of this study (L. Zhao et al., 2018). However, studies which investigated NANOG mRNA expression in head and neck and cervical cancers indicated that NANOG as a suitable CSC marker (Ding et al., 2016; Habu et al., 2015).

A discrepancy exists between these observations and the findings of this thesis. This is likely attributable to the methods of RNA extraction used: whilst this present study extracted RNA from FFPE, both other studies used RNA extracted from cell cultures. FFPE tissue is known to have a greater variation in quality as compared to other sample types, which may account for the lack of significance in our results (Gaffney, Riegman, Grizzle, & Watson, 2018). Confirmation of the significance of NANOG in GEJA thus requires further analysis using alternative means of RNA extraction.

POU5F1

Similar to NANOG, expression of POU5F1 was increased in only a small majority of paired cases which showed differential expression and is therefore of no significance. Its expression in higher grade samples was not associated with any clinicopathological parameters, nor with any survival outcomes. These findings were unexpected as multiple previous studies have identified an association between adverse clinical outcomes and increased gene expression of POU5F1 in other malignancies, including colorectal cancer, gastric cancer and OAC (Z. Chen et al., 2009; Miyoshi et al., 2018; X. Wang et al., 2014). Whilst the sample size of each of these studies was comparable to that of this thesis (range 55-95 patients), the tissue from which RNA was extracted differed. As discussed above, this present study extracted RNA from FFPE, whilst each other study mentioned used fresh tumour samples snap frozen in liquid nitrogen, which is a superior means of preservation of DNA and RNA (Gaffney et al., 2018). These differences in RNA extraction methods may account for the discrepancy in significance of results between the studies. Further analysis using alternative methods of RNA extraction may thus reveal a significant expression pattern of POU5F1 in GEJA.

Serpine-1

Expression of Serpine-1 – also known as PAI-1 – showed no association with any clinical outcomes, nor did it demonstrate a significant change in expression between paired tumour samples. Indeed, the majority of tumour samples analysed did not express Serpine-1.

These results are at odds with studies in the published literature. Brungs *et al* performed a meta-analysis of 22 studies, amounting to 1966 patients, which described an association between PAI-1 and both high risk disease and poor survival outcomes in GEJA (Brungs et al., 2017). Whilst the definition of GEJA in this study incorporated OAC, GEJA and gastric adenocarcinomas, only one study actually analysed true GEJA samples, amounting to 0.01% (26/1966) of all tumours analysed. This indicates that the results of Brungs' meta-analysis are not directly comparable to the findings presented in this thesis. Additionally, the analysis methods varied amongst the studies, ranging from ELISA to IHC to RT-PCR, thus complicating accurate comparison of studies. Future interrogation of GEJA in isolation, as defined clinically as Siewert I-III, may be warranted to determine the precise role of Serpine-1 in GEJA.

5.2 miRNA Expression in GEJA

As per the study hypothesis, it was anticipated that 'higher grade' tumour samples would show a miRNA expression profile more in keeping with EMT and/or stemness properties, as compared to their matched 'lower grade' samples. In 'higher grade' samples, 5 miRNAs were expected to be upregulated (miR-10b, miR-21, miR-221 miR-223 and

miR-224) and 6 miRNAs were expected to be downregulated (miR-16, miR-17, miR-203a, miR-200a, miR-133b and miR-141) (Chai et al., 2019; Z. Chen et al., 2013; L. L. Fang et al., 2017; Knoll et al., 2014; B. Liu et al., 2018; Lynam-Lennon et al., 2017; Ma et al., 2015; Sheedy & Medarova, 2018; Taube et al., 2013; T. Wang et al., 2017; Y. Wang et al., 2016). The opposite pattern of expression was predicted in 'lower grade' tumour samples. miRNA expression profiles associated with 'higher grade' tumour samples were anticipated to correlate with adverse clinical outcomes. Appendix 4.2 lists the significant associations identified with increased expression of each individual miRNA, as discussed below.

As discussed in Section 3.3.1, a suitable miRNA normalizer was not identified for use in GEJA. Global normalization was chosen as the most appropriate means by which to normalize the miRNA expression data. Whilst ideally used in in studies which analyse large numbers of miRNAs, previous studies have also used this method in the absence of a stable reference gene (Faraldi et al., 2019).

miR-221

miR-221 is an oncogenic miRNA whose over-expression has been shown to play a role in enhancing the tumorigenicity of CSCs and predicting poorer clinical outcomes across a range of solid organ malignancies including breast, pancreatic and colorectal cancers (Cheng et al., 2018; B. Li et al., 2016; B. Liu et al., 2018; Mukohyama et al., 2019). Its expression has previously been described in gastric cancer, and has been shown to enhance resistance to 5-FU and alter the expression of EMT-associated genes in oesophageal cancer (Shajari & Mollasalehi, 2020; Y. Wang et al., 2016; Zhou, Peng, & Xu, 2019).

Whilst this current data is novel in describing the patterns of miR-221 expression in GEJA, the findings are in keeping with previous studies pertaining to its association with adverse clinical outcomes in malignancies at different sites. This present study showed that miR-221 expression is significantly increased in 'higher grade' tumour samples, as compared to their 'lower grade' pairs. Increased miR-221 expression was also shown to be significantly associated with adverse clinical outcomes including disease recurrence, poor DFS and poor OS.

Interrogation of the signalling pathways regulated by miR-221 revealed its role in pathways involved in the regulation of stem cells, through which it was shown to directly regulate POU5F1 gene expression. It was also implicated in the TGF-β pathway, through which it plays an important role in EMT induction. Taken in combination, these findings clearly demonstrate an association between miR-221 expression and EMT, CSC-like cells and poor survival outcomes in GEJA. As such, miR-221 holds great promise as a prognostic biomarker or a novel therapeutic candidate for targeted treatments in aggressive GEJA.

miR-21

miR-21 is another oncogenic miRNA which has been previously shown to target a number of tumour suppressor genes and play an important role in tumorigenesis in a wide variety of malignancies (Shajari & Mollasalehi, 2020). Circulating miR-21 has been proposed to be a potential diagnostic biomarker in both gastric and oesophageal cancers (Komatsu et al., 2011; Tsujiura et al., 2010). This present data is in keeping with this trend, describing a significant association between miR-21 over-expression and both poor treatment response (TRG 4/5) and disease recurrence.

An interesting study by Odenthal *et al* investigated serum miR-21 levels in GEJA before and after NAT, concluding that NAT had no significant influence over expression levels (Odenthal et al., 2015). These findings, taken in tandem with this study's observation that miR-21 upregulation is associated with a poor response to NAT, suggest that miR-21 holds potential as a biomarker with which to predict treatment response at the point of initial diagnosis. However, Odenthal's study analysed miR-21 expression in serum, whilst this analysis was performed on tissue, thus the results are not directly comparable and would require further interrogation in order to determine their clinical applicability.

miR-21 was implicated in signalling pathways involved in regulating stem cell pluripotency, whereby it was shown to modify a number of gene targets known to be involved in the promotion of EMT and CSCs, including Dual-specificity phosphatase 9 (DUSP9) and Polycomb group ring finger 2 (PCGF2) (H. Lu et al., 2018; X. F. Wang et al., 2016). These findings indicate a critical role of miR-21 in regulation of both EMT and putative CSCs in aggressive GEJA and point towards its potential utility as a prognostic biomarker.

miR-16

miR-16 acts as a tumour suppressor across a wide range of malignancies; its downregulation is associated with EMT and the generation of CSCs (Farace et al., 2020; H. Zhang & Li, 2019). Whilst its inhibitory effect in gastric cancer has previously been described, as of yet no studies have investigated its role in GEJA (Jiang & Wang, 2018).

A significant association was identified between miR-16 upregulation and a good OS. This is very much in keeping with the role of miR-16 in other malignancies, whereby its over-expression is associated with a decrease in tumorigenicity (W. L. Wu, Wang, Yao,

& Li, 2015). Bioinformatic interrogation of signalling pathways showed that miR-16 is involved in the regulation of multiple gene targets within a range of signalling pathways, including pathways which regulate the pluripotency of stem cells; the TGF-β pathway, which is implicated in EMT; and the Hippo pathway, which is involved in tumorigenesis and generation of CSCs (Mo, Park, & Guan, 2014). These findings once again support the role of miR-16 in the regulation of putative CSCs and EMT in aggressive GEJA.

miR-224

miR-224 upregulation was shown to be significantly associated with disease recurrence and show a strong correlation with higher TRG groups and therefore poor response to NAT. This data is novel as no other studies have investigated the role of miR-224 in GEJA. Upregulation of miR-224 has been previously demonstrated in other malignancies including gastric and colorectal carcinoma however, and has shown an association with poor survival outcomes, as is in keeping with this data (G. J. Zhang, Zhou, Xiao, Li, & Zhou, 2013; Y. Zhang, Li, Ma, Ding, & Zhang, 2016).

Bioinformatic interrogation of the downstream targets of miR-224 pointed towards its role in the regulation of four main signalling pathways which are critical in the control of EMT and CSCs: stem cell pluripotency, TGF-β, Hippo and Wnt pathways. This data suggests a role of miR-224 in the regulation of EMT and CSC-like cells in clinically aggressive GEJA.

miR-141

A significant association between miR-141 and HER2 positivity was demonstrated in this study. HER2 over-expression shows marked intra-tumour heterogeneity in GEJA (D. Zhao, Klempner, & Chao, 2019). Accordingly, miR-141 may hold great promise as an adjunct biomarker for use in directing HER2-targeted therapies and monitoring the development of acquired resistance to these treatments. Statistical analysis to determine the outcomes of patients with HER2 positive tumours and high miR-141 levels was not performed in this study due to incomplete data collection and an unequal distribution of cases between each HER2 expression category.

An association between downregulation of miR-141 and the promotion of EMT and CSCs has been demonstrated in a number of solid organ malignancies, yet no data exists regarding its role in GEJA (Al-Khalaf & Aboussekhra, 2017; J. Ye et al., 2017). This study did not identify any significant correlation between miR-141 expression and adverse clinical outcomes, indicating that it does not play a role in the identification of putative CSCs in GEJA. However, analysis of a larger number of tumour samples would be required to confirm or refute this finding.

miR-17

Based on the findings of Lynam-Lennon *et al*, who previously showed that miR-17 downregulation is associated with CSC-like cells and radio-resistance in OAC (Lynam-Lennon et al., 2017), it was anticipated that downregulation of miR-17 would also be seen in GEJA; however, this was not the case. miR-17 is part of the oncogenic miR-17-92 cluster, which is highly expressed across a range of malignancies including lung, pancreatic and prostate cancer (L. L. Fang et al., 2017). Interestingly, this cluster may

also express tumour suppressor characteristics in OAC, as demonstrated by Lynam-Lennon. Whilst no significant differential expression was observed between paired 'lower grade' and 'higher grade' samples in this present study, it was noted that over half of the pairs analysed did not express miR-17 in either sample. Taking this into consideration, it is interesting to speculate that miR-17 may indeed act as a tumour suppressor in advanced GEJA, irrespective of histological features, however further large scale validation studies are required to validate this assumption.

miR-200a

miR-200a was expected to be downregulated in 'higher grade' samples, as compared to their 'lower grade' pairs. A marked downregulation of miR-200a has previously been demonstrated in cells which have undergone EMT in response to TGF-β (Gregory et al., 2008), however this present study did not identify this pattern in GEJA. The discrepancy between the anticipated and recorded results in this study may be explained by the varied functions of miR-200a, which has been shown to play both an oncogenic and a tumour suppressive role in different settings (Becker, Takwi, Lu, & Li, 2015). Chen *et al* previously described an upregulation of miR-200a in OAC and a downregulation in gastric adenocarcinoma (Z. Chen et al., 2013), whilst Saad *et al* showed that miR-200a upregulation is more pronounced in early stage OAC, suggesting that it plays a role in tumour development rather than tumour progression (Saad et al., 2013). Taking into consideration these opposing functions, further work is required in order to accurately define the role of miR-200a in GEJA.

miR-223 and miR-10b

It was proposed that upregulation of miR-223 and miR-10b would be significantly associated with an aggressive tumour phenotype and clinical course, yet this was not identified. Contrary to the negative findings of this study, Fassan *et al* described the utility of miR-223 as a biomarker for identification of early GEJA (Fassan et al., 2017). The discrepancy between these results may be explained by the observation that the resection specimens included in this study were predominantly advanced cancers (T3/4) and thus their expression patterns may differ from that of the early GEJAs described by Fassan.

Whilst no studies have examined the role of miR-10b in GEJA, its oncogenic role has been described across a range of malignancies including pancreatic cancer, melanoma and OAC (Ouyang, Gore, Deitz, & Korc, 2014; Sheedy & Medarova, 2018; Tian et al., 2010). By comparison, there is evidence to suggest that miR-10b has a tumour suppressor role in some malignancies including clear cell renal cell carcinoma and bladder cancer (Sheedy & Medarova, 2018). In light of the significant number of paired samples which did not express miR-10b in this present study, it is interesting to postulate that miR-10b may exert a tumour suppressor effect in GEJA, however further studies are required to substantiate this claim.

miR-133b, miR-203a and miR-103a

miR-133b, miR-203a and miR-103a were not expressed in any tumour sample, thus it is not possible to infer anything about their role in GEJA from this data. Interestingly however, we expected miR-203a to be downregulated in GEJA due to its known function in suppressing EMT and CSC properties (Taube et al., 2013). Similarly, miR-133b downregulation was anticipated in GEJA as it has been previously demonstrated in OAC

and gastric adenocarcinoma (Z. Chen et al., 2013). It is possible that the lack of expression in the tumour samples may still reflect a downregulation of miR-133b and miR-203a as compared to normal GEJA, however this was not investigated as part of this study and therefore requires further investigation to determine the significance of these miRNAs in GEJA. miR-103a was included in the study as a possible internal control (Z. Chen et al., 2013), thus its lack of expression indicates that it is not useful as a miRNA normalizer in GEJA.

5.3 Protein Expression in GEJA

The study hypothesis proposes that higher expression of proteins associated with EMT, including Vimentin, will be observed in 'higher grade' tumour samples and will be associated with worse clinical outcomes. By comparison, protein markers associated with the epithelial phenotype, such as E-cadherin, are expected to show higher expression in 'lower grade' areas of tumour and correlate with better clinical outcomes. Increased expression of each stemness marker (CD34, Oct-4, EpCAM, ALDH1A1 and CD133) was proposed to show higher levels of expression in 'higher grade' tumour samples as compared with their matched 'lower grade' samples, with higher expression correlating with worse clinical outcomes. Appendix 4.3 lists the significant associations with increased expression of each individual protein marker, as discussed below.

EpCAM

EpCAM – a putative CSC marker – showed a significant differential expression pattern between 'lower grade' and 'higher grade' tumour samples. Contrary to the study

hypothesis, 92.9% of these cases demonstrated a decrease in expression from ‘lower grade’ to ‘higher grade’ samples. Interestingly, this trend was previously identified by Driemel *et al*, who described a context driven expression of EpCAM in oesophageal cancer (Driemel et al., 2014). Driemel proposed a correlation between EpCAM (high) phenotypes with proliferative tumour stages, and between EpCAM (low/negative) phenotype with migration, invasion and dissemination. Applying this interpretation of EpCAM staining to GEJA infers that ‘higher grade’ tumour samples may be associated with a more aggressive disease phase than ‘lower grade’ samples, as is in keeping with the study hypothesis. These findings additionally highlight the intra-tumour heterogeneity present in GEJA.

Despite the clear delineation of staining intensities between paired tumour samples, EpCAM expression in ‘higher grade’ samples was not associated with any adverse clinicopathological features. Survival outcomes were not interrogated because the distribution of cases across the IHC scoring classifications was unequal and thus unsuitable for statistical analysis. Whilst the results of this study suggest that EpCAM expression may be of use in disease stratification for therapeutic approaches, determining its use as a prognostic and putative CSC biomarker in GEJA requires further analysis with a larger study population in order to formally assess its relationship with survival outcomes.

E-cadherin and Vimentin

Expression of the cell surface proteins E-cadherin and Vimentin did not show any significant association with any clinicopathological feature, nor with any survival outcome, although E-cadherin did show a strong association with serosal involvement.

This lack of significance is at odds with expression of their gene counterparts CDH1 and Vimentin, which both showed a significant association with DFS. A differential expression pattern was not identified between paired tumour samples for either protein. Indeed, the majority of tumour samples expressed an epithelial phenotype with E-cadherin positivity and negative Vimentin expression, regardless of their histological differentiation status. This expression pattern is considered to be most in keeping with aggressive disease.

These findings suggest that by contrast to their mRNA counterparts, protein markers of EMT are not of use in delineating cancer cells going through EMT. It is only possible to postulate as to the reason for this alteration in significance at the protein level. This is because a wide range of regulatory processes that may occur following mRNA production, including post-transcriptional, translational and protein degradation regulation, as modulated by miRNAs or other mechanisms (Vogel & Marcotte, 2012). However, it is also conceivable that the discrepancy in these results may be attributable to intra-tumour heterogeneity: LCM and TMA construction may have inadvertently sampled tumour cells from slightly different areas within the tumour, leading to differences in expression patterns. Further analysis is required in order to confirm the validity of these results.

CD34

CD34 is a cell surface antigen that has been reported to select small populations of putative CSCs (T. Chen et al., 2017). It was not expressed in any tumour sample in this analysis. This observation is particularly interesting in light of a previous study which demonstrated CD34 positivity in gastric carcinomas without serosal involvement (Cabuk

et al., 2007). This difference in expression pattern between gastric carcinoma and GEJA supports the opinion that GEJAs are best regarded as a distinct entity.

A separate study which examined the role of the tumour microenvironment in OAC described higher CD34 expression in peritumoral adipose tissue, as compared to distal adipose, with a decrease in CD34 expression observed in peri-tumoural samples in the setting of NAT (Carraro et al., 2017). Whilst the authors did not investigate the expression of CD34 in cancer cells, they did however describe how treatment of adipose derived stem cells with a conditioned medium derived from tumoral and adipose tissue of patients with OAC resulted in increased expression of CD34. These findings suggest a role of CD34 as a stemness marker and a marker of treatment resistance in OAC.

The expression patterns here demonstrated in GEJA are contrary to that of gastric carcinoma and OAC, indicating that GEJAs represent a biologically distinct subset of malignancies. Furthermore, due to a diffuse negative staining pattern, CD34 is of no use in identification of putative CSCs in GEJA.

Oct-4

Oct-4 is a protein which is frequently used as a stem cell marker. Unexpectedly, it was upregulated from ‘higher grade’ to ‘lower grade’ samples in only 16.7% of pairs with differential expression. Oct-4 expression in ‘higher grade’ tumour samples from each patient showed no association with any adverse clinicopathological characteristics, however it was significantly associated with a poor OS and DFS.

This data suggests that Oct-4 protein expression is of prognostic use in GEJA. Interestingly, miR-221 is known to be involved in the regulation of Oct-4 through

signalling pathways governing the regulation of pluripotency: this relationship is supported by the fact that miR-221 is itself associated with a poor OS and DFS. Regulation of Oct-4 by miR-221 has been previously described in breast cancer, whereby increased miR-221 was shown to be associated with an increase in both Oct-4 gene and protein expression as a result of a decrease in DNA promotor methylation (Roscigno et al., 2016). As previously discussed in Section 5.1, whilst this study did not identify any significant pattern in POU5F1 expression, it is proposed that gene expression analysis using frozen GEJA samples is likely to yield a significant association between POU5F1 and adverse survival outcomes.

Of interest, this study additionally observed that Oct-4 was expressed in both benign and dysplastic junctional glandular tissue, with and without IM. This is in keeping with prior studies which describe the role of Oct-4 in the progression from BO to OAC (Shen et al., 2016; X. Wang et al., 2014). Wang *et al* additionally demonstrated a loss of stemness traits in experimentally isolated stem cells in response to downregulation of Oct-4 by siRNAs (X. Wang et al., 2014).

Overall, these findings support the role of Oct-4 as a marker of putative CSCs in GEJA. Whilst Oct-4 positivity was noted in pre-malignant conditions within the GEJ, these expression patterns were not formally quantified and thus it is not possible to comment on their significance. The role of miR-221 in regulation of POU5F1 in GEJA also merits further interrogation at both the mRNA and protein level.

ALDH1A1

ALDH1A1 is another cell surface marker used to identify CSC-like cells in tumours including GEJA (Brungs et al., 2019). Comparison of total ALDH1A1 staining patterns

between paired samples revealed no significant trends. ALDH1A1 expression in all 'higher grade' samples showed no association with any adverse clinicopathological parameters, however its expression was significantly associated with a poor OS and DFS. These findings indicate that ALDH1A1 may be of use in the clinical setting as a prognostic tool for survival outcomes.

ALDH1A1 expression has been consistently associated with poor survival outcomes across a range of malignancies (van der Waals, Borel Rinkes, & Kranenburg, 2018; Xing, Luo, Long, Zeng, & Li, 2014; Y. Ye, Zhang, Chen, Wang, & Wang, 2018). Brungs *et al* concurred with the findings of this study by demonstrating an association between poor OS and ALDH1A1 staining in metastatic deposits of GEJA (Brungs et al., 2019). Increased expression has previously been shown to reflect resistance to anti-cancer treatments across a range of malignancies (Ajani et al., 2014; Roy, Connor, Al-Niaimi, Rose, & Mahajan, 2018), however this trend was not identified in this present study. Nevertheless, the association between ALDH1A1 and poor survival outcomes is in keeping with its known role as a CSC marker, thus pointing towards its utility in the identification of CSC-like cells in GEJA.

CD133

CD133 has also been used as a putative CSC marker in GEJA, in addition to a range of other malignancies including liver, brain, colorectal and pancreatic cancers (Brungs et al., 2019; K. Chen et al., 2013). The prognostic role of CD133 in GEJA was investigated in this study through analysis of both percentage staining and staining topology within tumour samples. A higher percentage of positive neoplastic cells and positive staining of

clusters of cancer cells were both expected to correlate with worse clinical outcomes (Zeppernick et al., 2008).

A significant association was identified between CD133 staining patterns and adverse clinical outcomes. CD133 percentage staining was significantly associated with disease recurrence, whilst a strong association was identified between serosal involvement and both CD133 percentage staining and topology. Whilst CD133 percentage staining was excluded from the survival analyses, CD133 staining of clusters of cancer cells was significantly associated with a poor DFS.

Our findings suggest that CD133 is of potential use in prognostication for survival and disease recurrence. These findings are very much in keeping with the current literature, which links CD133 expression with poor survival outcomes and resistance to treatment across a range of malignancies (C. Fang et al., 2017; Hou, Chao, Tung, Wang, & Shan, 2014; Zhong, Shi, Zhou, & Wang, 2018). Additionally, as with Oct-4, CD133 expression has also previously been linked to the BO – early OAC – advanced OAC sequence (Mokrowiecka et al., 2017): a trend which merits investigation in GEJA.

It is important to take caution in comparing studies which investigate CD133 staining due to a variability in staining patterns between different clones (Hermansen et al., 2011). Nevertheless, taken in the context of consistent reports of CD133 as a CSC marker, these findings support its role as putative CSC marker and prognostic aid in GEJA.

5.4 Significance of mRNA, miRNA and Protein Expression Data

Evidence to Support the Role of EMT in GEJA

Overall, the gene expression analyses identified a significant relationship between EMT genes and clinical outcomes. The association between poor DFS with upregulation of Vimentin and downregulation of CDH1 expression supports the role of EMT in promoting aggressive tumour behaviour in GEJA.

Whilst the link between increased Vimentin expression and poor DFS seems incongruous with the described association between decreased expression and enhanced drug resistance, both observations are nevertheless considered to be valid. It is certainly possible that Vimentin exerts its effect on disease recurrence and treatment response through different molecular pathways. Additionally, the impact of adjuvant therapy upon disease recurrence has not been taken into consideration and may play a role in enhancing the DFS of patients with low Vimentin expression. However, it is important to confirm the validity of these findings through analysis of Vimentin gene expression in a larger patient cohort.

Contrary to the results of the gene expression analysis, the expression patterns of E-cadherin and Vimentin proteins were not in keeping with the study hypothesis that 'higher grade' and 'lower grade' tumour samples would express mesenchymal and epithelial phenotypes respectively. It was observed that the majority of samples, regardless of histological delineation, expressed a mesenchymal phenotype, which is classically associated with aggressive tumour behaviour. This observation is notable in the context of a study population, which contains a majority of advanced GEJAs (T3/4), yet despite

this the observed expression patterns were not significantly associated with any adverse outcomes.

Evidence to Support the Role and Characterisation of CSCs in GEJA

Whilst the expression patterns of genes associated with EMT were shown to be significantly associated with adverse clinical outcomes, no stemness genes showed significant expression patterns. It was anticipated that expression of Serpine-1 and stemness markers NANOG, POU5F1 would also correlate with adverse clinical outcomes, however the lack of significant findings has been attributed to the use of FFPE for RNA extraction in this study. It is also important to remember that intra-tumour heterogeneity and the inherent plasticity of CSCs both complicate analysis of these expression patterns.

Our protein expression analysis supports the use of EpCAM, Oct-4, ALDH1A1 and CD133 as markers of putative CSCs in GEJA. As previously discussed in Section 1.3.1, there are many limitations to accurate identification of CSCs, including CSC plasticity, heterogenous expression throughout the tumour and indeed co-expression patterns in differentiated cell types (Prasetyanti & Medema, 2017; Y. Wu & Wu, 2009). Further studies looking at the expression of these markers at the cellular level are required to validate their role as CSCs in GEJA.

Evidence to Support the Overlapping Roles of EMT and CSCs in GEJA

The results of the gene and protein expression analyses were discordant, however miRNA expression patterns indicate an overlapping role of EMT and CSCs in GEJA. Because the

activities of miRNAs frequently overlap, thus their regulatory actions must not be interpreted as independent of each other. Indeed, the miRNAs shown to have a significant association with EMT, CSC-like cells and adverse clinical outcomes in this study (miR-21, miR-221, miR-16, miR-141 and miR-224) overlapped greatly in terms of the signalling pathways and gene targets that they regulate (Appendix 3). For example, each of these 5 miRNAs was implicated in the regulation of the signalling pathways involved in regulating stem cell pluripotency, with many shared gene targets including Wnt and SMAD family members. Furthermore, the regulation of the TGF-β pathway involved in EMT induction is shared by miR-221, miR-224, miR-16 and miR-141, again with frequently overlapping gene targets.

These observations highlight the role of these miRNAs in the regulation of EMT and putative CSCs in GEJA. They support the use of a miRNA signature panel in the clinical setting as a prognostic biomarker for aggressive tumour behaviour, and they have identified a number of potential overlapping gene targets for potential use in the development of novel anti-cancer treatments. Additionally, miRNA expression patterns may help to explain the discordance noted between the significance of mRNA and protein expression patterns related to EMT and CSCs in this study due to their influence on a wide range of regulatory processes that occur via a range of different signalling pathways post mRNA production. However, this is only speculatory and further investigation would be required to confirm or refute this theory.

5.5 Prognostic Utility of a Molecular Expression Panel in GEJA

Whilst the expression patterns of individual genes, miRNAs and proteins showed some significant associations with adverse clinical outcomes, their utility as a collective model was additionally investigated to determine their combined predictive ability.

The predictive model used in this study included each mRNA, miRNA and protein marker, in addition to sex and age at diagnosis. It was shown to be of use in predicting 'higher grade' tumour samples, with an accuracy rate of 83.5%, although its accuracy decreased to 43.1% when trying to identify 'lower grade' samples. A similar pattern was seen in prediction of disease recurrence, whereby recurrence was predicted with 81% accuracy, as compared to 53.5% for prediction of no recurrence.

Of great interest, NANOG and POU5F1 were amongst the most important datapoints used by the model to make these predictions. This indicates that despite their limited prognostic utility in univariate analyses, they may still hold potential as part of a multivariate predictive model. It is also interesting to note that the most significant markers involved in these predictive models are markers of stemness and putative CSCs, whilst EMT markers held significantly less weight. No other studies were identified in the current literature which described this trend.

This is to date the only multivariate analysis of this size looking at the predictive value of a cohort of expression markers in GEJA. Whilst the degree of accuracy of this model is not yet sufficient to support its use in the clinical setting, it nevertheless warrants further interrogation with a refinement of certain datapoints and a greater sample size to determine its potential role in prognostication for patients with GEJA.

5.6 Study Applications

This study has identified a range of molecular markers associated with CSC-like cells and EMT which may be of use in the clinical setting to identify patients with clinically aggressive GEJA. The potential prognostic utility of these markers, used in isolation or in combination as part of a predictive model, has been clearly demonstrated. Furthermore, characterization of the molecular markers of CSC-like cells and the signalling pathways involved in their regulation presents a number of molecular targets which may be used in the development of novel anti-cancer drug therapies. Such targeted therapies may help to improve treatment response and thus survival outcomes in patients with GEJA.

5.7 Study Limitations

This study is not without its limitations. It is important to remember that the clinical outcomes discussed in this study are somewhat limited by the small sample size, which may have increased the variability of the data collected. This issue is further exacerbated by incomplete data collection, as was the case in conducting the survival analyses. Here, because this is a retrospective cohort study, all patients who were not formally recorded in the clinical notes as deceased were considered to be alive on the final date of data collection. Additionally, the 'lost to follow up' date was used as a surrogate date of death for five patients who had no recorded clinical episodes for a number of years but for whom no definite date of death was identifiable. The survival analyses were thus limited by the assumptions of survival outcomes in patients without a recorded date of death in the clinical notes.

LCM sampling of 'lower grade' and 'higher grade' areas of tumour was limited by inter-observer variability and intra-tumour heterogeneity, in addition to technical difficulties in extracting pure tumour cells without simultaneously capturing small volumes of adjacent connective tissue. Should this study be replicated in the future, differences in histological interpretation of tumour differentiation and interpretation of what depth of tumour constitutes the 'tumour edge' may decrease the reproducibility of the results. Furthermore, generous versus prudent tumour sampling by the surgical pathologist upon initial receipt of the resection specimen may have had a bearing on whether the areas sampled were truly reflective of the biological nature of the entire tumour. This point is emphasised in a study by Agoston *et al*, which described the need to submit the entire tumour bed in OAC specimens without macroscopic evidence of residual tumour: they stated that variable outcomes for patients with reported CPR was explained by the adequacy of histologic examination of the tumour bed (Agoston et al., 2015).

This study was also somewhat limited by difficulties in quantification of small volumes of RNA. It was difficult to ascertain the number of cells sampled during LCM, and thus was reflected in the difficulties in quantifying small volumes of RNA. Because many of the cases had only tiny volumes of residual tumour post-NAT, formal RNA quantification was not performed following extraction due to the decreased sensitivity in measurement of such small volumes. Pre-amplification was instead performed on all samples, thus the inherent sampling bias remained equal across the entire cohort.

Due to an unequal distribution of NAT amongst the study population, it was not possible to determine if there was a significant difference in the molecular expression patterns between the two groups. The expression data must therefore be interpreted with caution as it cannot be assumed that NAT does not confer a lasting alteration in molecular patterns upon a tumour. Furthermore, patients within the study population were not subdivided as

per individual NAT regimens received as the resulting small numbers would have made analyses unreliable. The impact of adjuvant therapy was not explored in this study, but it is important to remember its potential influence upon survival outcomes that will not be reflected in the molecular expression patterns of the excision specimen.

5.8 Conclusions and Recommendations for Further Research

An extensive expression analysis on an Irish cohort of surgical resection specimens for GEJA has been performed in this thesis in order to determine the presence or absence of a specific molecular expression pattern associated with aggressive disease. As molecular medicine continues to grow in importance in the clinical setting, it is imperative that the knowledge pertaining to GEJA continues to increase in order to provide the best clinical care for patients. Analysis of the expression data in this study points towards a role of certain markers, used both individually and as part of a combined predictive model, as prognostic biomarkers of aggressive GEJA.

The ability of this study to conclusively identify CSCs, as distinct from non-CSCs, was limited by intra-tumour heterogeneity, tumour cell plasticity and a paucity of published literature pertaining to this subject with which to compare the expression patterns observed in GEJA. Nevertheless, the analyses have identified significant expression patterns of known CSC markers within aggressive areas of GEJA. Additionally, bioinformatic interrogation of miRNA expression patterns demonstrated that each miRNA of significance was involved in the regulation of signalling pathways associated with pluripotency of stem cells. This data merits further interrogation at the cellular level

in order to confirm the role of these markers in aggressive GEJA and thus support their role as a prognostic tool in the clinical setting.

Further large scale validation studies using frozen tissue or cell cultures are required in order to investigate the role of stemness genes NANOG and POU5F1 in GEJA. A larger comprehensive study of patients with GEJA would be of additional benefit in order to determine if the expression markers which were identified as significant hold equal weight both pre- and post-NAT. Should this be the case, they would hold great benefit as prognostic markers at the point of initial diagnosis, and as a potential targets for novel first line drug therapies.

It is hoped that this research will contribute to the growing body of knowledge regarding disease progression and the role of CSCs in GEJA. This study is believed to be the first to investigate such a wide panel of expression markers in GEJA, and it is hoped that it will direct future studies and hopefully help to unravel the mystery of the roles and regulation of CSCs in GEJA.

Bibliography

Abdi, E., Latifi-Navid, S., Zahri, S., Yazdanbod, A., & Pourfarzi, F. (2019). Risk factors predisposing to cardia gastric adenocarcinoma: Insights and new perspectives. *Cancer Med, 8*(13), 6114-6126. doi:10.1002/cam4.2497

Agoston, A. T., Zheng, Y., Bueno, R., Lauwers, G. Y., Odze, R. D., & Srivastava, A. (2015). Predictors of Disease Recurrence and Survival in Esophageal Adenocarcinomas With Complete Response to Neoadjuvant Therapy. *Am J Surg Pathol, 39*(8), 1085-1092. doi:10.1097/pas.0000000000000420

Aires, A., Ocampo, S. M., Simoes, B. M., Josefa Rodriguez, M., Cadenas, J. F., Couleaud, P., . . . Cortajarena, A. L. (2016). Multifunctionalized iron oxide nanoparticles for selective drug delivery to CD44-positive cancer cells. *Nanotechnology, 27*(6), 065103. doi:10.1088/0957-4484/27/6/065103

Ajani, J. A., Wang, X., Song, S., Suzuki, A., Taketa, T., Sudo, K., . . . Berry, D. (2014). ALDH-1 expression levels predict response or resistance to preoperative chemoradiation in resectable esophageal cancer patients. *Mol Oncol, 8*(1), 142-149. doi:10.1016/j.molonc.2013.10.007

Al-Batran, S. E., Hofheinz, R. D., Pauligk, C., Kopp, H. G., Haag, G. M., Luley, K. B., . . . Tannapfel, A. (2016). Histopathological regression after neoadjuvant docetaxel, oxaliplatin, fluorouracil, and leucovorin versus epirubicin, cisplatin, and fluorouracil or capecitabine in patients with resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4-AIO): results from the phase 2 part of a multicentre, open-label, randomised phase 2/3 trial. *Lancet Oncol, 17*(12), 1697-1708. doi:10.1016/s1470-2045(16)30531-9

Al-Batran, S. E., Homann, N., Pauligk, C., Goetze, T. O., Meiler, J., Kasper, S., . . . Hofheinz, R. D. (2019). Perioperative chemotherapy with fluorouracil plus leucovorin, oxaliplatin, and docetaxel versus fluorouracil or capecitabine plus

cisplatin and epirubicin for locally advanced, resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4): a randomised, phase 2/3 trial. *Lancet, 393*(10184), 1948-1957. doi:10.1016/s0140-6736(18)32557-1

Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J., & Clarke, M. F. (2003). Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A, 100*(7), 3983-3988. doi:10.1073/pnas.0530291100

Al-Khalaf, H. H., & Aboussekhra, A. (2017). p16(INK4A) induces senescence and inhibits EMT through microRNA-141/microRNA-146b-5p-dependent repression of AUF1. *Mol Carcinog, 56*(3), 985-999. doi:10.1002/mc.22564

Alakhova, D. Y., Zhao, Y., Li, S., & Kabanov, A. V. (2013). Effect of doxorubicin/pluronic SP1049C on tumorigenicity, aggressiveness, DNA methylation and stem cell markers in murine leukemia. *PLoS One, 8*(8), e72238. doi:10.1371/journal.pone.0072238

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., . . . Stratton, M. R. (2013). Signatures of mutational processes in human cancer. *Nature, 500*(7463), 415-421. doi:10.1038/nature12477

Altorki, N., & Harrison, S. (2017). What is the role of neoadjuvant chemotherapy, radiation, and adjuvant treatment in resectable esophageal cancer? *Ann Cardiothorac Surg, 6*(2), 167-174. doi:10.21037/acs.2017.03.16

Amenabar, A., Hoppo, T., & Jobe, B. A. (2013). Surgical management of gastroesophageal junction tumors. *Semin Radiat Oncol, 23*(1), 16-23. doi:10.1016/j.semradonc.2012.09.002

Arnold, M., Soerjomataram, I., Ferlay, J., & Forman, D. (2015). Global incidence of oesophageal cancer by histological subtype in 2012. *Gut, 64*(3), 381-387. doi:10.1136/gutjnl-2014-308124

Bang, Y. J., Van Cutsem, E., Fuchs, C. S., Ohtsu, A., Tabernero, J., Ilson, D. H., . . . Shitara, K. (2019). KEYNOTE-585: Phase III study of perioperative chemotherapy with or without pembrolizumab for gastric cancer. *Future Oncol, 15*(9), 943-952. doi:10.2217/fon-2018-0581

Bass, A., Thorsson, V., Shmulevich, I. et al. (2014). Comprehensive molecular characterization of gastric adenocarcinoma. *Nature, 513*(7517), 202-209. doi:10.1038/nature13480

Basu, S., Cheriyamundath, S., & Ben-Ze'ev, A. (2018). Cell-cell adhesion: linking Wnt/beta-catenin signaling with partial EMT and stemness traits in tumorigenesis. *F1000Res, 7*. doi:10.12688/f1000research.15782.1

Batlle, E., & Clevers, H. (2017). Cancer stem cells revisited. *Nat Med, 23*(10), 1124-1134. doi:10.1038/nm.4409

Batool, S., Khan, M., Akbar, S. A., & Ashraf, I. (2019). Risk factors and patterns of recurrence after curative resection in Gastroesophageal Junction Adenocarcinoma. *Pak J Med Sci, 35*(5), 1276-1283. doi:10.12669/pjms.35.5.963

Battaglin, F., Naseem, M., Puccini, A., & Lenz, H. J. (2018). Molecular biomarkers in gastro-esophageal cancer: recent developments, current trends and future directions. *Cancer Cell Int, 18*, 99. doi:10.1186/s12935-018-0594-z

Beck, B., Lapouge, G., Rorive, S., Drogat, B., Desaedelaere, K., Delafaille, S., . . . Blanpain, C. (2015). Different levels of Twist1 regulate skin tumor initiation, stemness, and progression. *Cell Stem Cell, 16*(1), 67-79. doi:10.1016/j.stem.2014.12.002

Becker, L. E., Takwi, A. A., Lu, Z., & Li, Y. (2015). The role of miR-200a in mammalian epithelial cell transformation. *Carcinogenesis, 36*(1), 2-12. doi:10.1093/carcin/bgu202

Beerling, E., Seinstra, D., de Wit, E., Kester, L., van der Velden, D., Maynard, C., . . . van Rheenen, J. (2016). Plasticity between Epithelial and Mesenchymal States Unlinks EMT from Metastasis-Enhancing Stem Cell Capacity. *Cell Rep, 14*(10), 2281-2288. doi:10.1016/j.celrep.2016.02.034

Bhardwaj, M., Sen, S., Chosdol, K., Bakhshi, S., Pushker, N., Sharma, A., . . . Singh, V. K. (2019). Vimentin overexpression as a novel poor prognostic biomarker in eyelid sebaceous gland carcinoma. *Br J Ophthalmol*. doi:10.1136/bjophthalmol-2018-313285

Blank, S., Schmidt, T., Heger, P., Strowitzki, M. J., Sisic, L., Heger, U., . . . Ulrich, A. (2018). Surgical strategies in true adenocarcinoma of the esophagogastric junction (AEG II): thoracoabdominal or abdominal approach? *Gastric Cancer, 21*(2), 303-314. doi:10.1007/s10120-017-0746-1

Bonnet, D., & Dick, J. E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med, 3*(7), 730-737. doi:10.1038/nm0797-730

Bonnomet, A., Syne, L., Brysse, A., Feyereisen, E., Thompson, E. W., Noel, A., . . . Gilles, C. (2012). A dynamic in vivo model of epithelial-to-mesenchymal transitions in circulating tumor cells and metastases of breast cancer. *Oncogene, 31*(33), 3741-3753. doi:10.1038/onc.2011.540

Brabletz, T., Jung, A., Hermann, K., Gunther, K., Hohenberger, W., & Kirchner, T. (1998). Nuclear overexpression of the oncoprotein beta-catenin in colorectal

cancer is localized predominantly at the invasion front. *Pathol Res Pract, 194*(10), 701-704. doi:10.1016/s0344-0338(98)80129-5

Brabletz, T., Jung, A., Reu, S., Porzner, M., Hlubek, F., Kunz-Schughart, L. A., . . . Kirchner, T. (2001). Variable beta-catenin expression in colorectal cancers indicates tumor progression driven by the tumor environment. *Proc Natl Acad Sci U S A, 98*(18), 10356-10361. doi:10.1073/pnas.171610498

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin, 68*(6), 394-424. doi:10.3322/caac.21492

Brooks, M. D., Burness, M. L., & Wicha, M. S. (2015). Therapeutic Implications of Cellular Heterogeneity and Plasticity in Breast Cancer. *Cell Stem Cell, 17*(3), 260-271. doi:10.1016/j.stem.2015.08.014

Brown, R. L., Reinke, L. M., Damerow, M. S., Perez, D., Chodosh, L. A., Yang, J., & Cheng, C. (2011). CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. *J Clin Invest, 121*(3), 1064-1074. doi:10.1172/jci44540

Brungs, D., Aghmesheh, M., Vine, K. L., Becker, T. M., Carolan, M. G., & Ranson, M. (2016). Gastric cancer stem cells: evidence, potential markers, and clinical implications. *J Gastroenterol, 51*(4), 313-326. doi:10.1007/s00535-015-1125-5

Brungs, D., Chen, J., Aghmesheh, M., Vine, K. L., Becker, T. M., Carolan, M. G., & Ranson, M. (2017). The urokinase plasminogen activation system in gastroesophageal cancer: A systematic review and meta-analysis. *Oncotarget, 8*(14), 23099-23109. doi:10.18632/oncotarget.15485

Brungs, D., Lochhead, A., Iyer, A., Illemann, M., Colligan, P., Hirst, N. G., . . . Ranson, M. (2019). Expression of cancer stem cell markers is prognostic in metastatic gastroesophageal adenocarcinoma. *Pathology, 51*(5), 474-480. doi:10.1016/j.pathol.2019.03.009

Buas, M. F., & Vaughan, T. L. (2013). Epidemiology and risk factors for gastroesophageal junction tumors: understanding the rising incidence of this disease. *Semin Radiat Oncol, 23*(1), 3-9. doi:10.1016/j.semradonc.2012.09.008

Byers, L. A., Diao, L., Wang, J., Saintigny, P., Girard, L., Peyton, M., . . . Heymach, J. V. (2013). An epithelial-mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance. *Clin Cancer Res, 19*(1), 279-290. doi:10.1158/1078-0432.Ccr-12-1558

Cabrera, M. C., Hollingsworth, R. E., & Hurt, E. M. (2015). Cancer stem cell plasticity and tumor hierarchy. *World J Stem Cells, 7*(1), 27-36. doi:10.4252/wjsc.v7.i1.27

Cabuk, D., Basaran, G., Celikel, C., Dane, F., Yumuk, P. F., Iyikesici, M. S., . . . Turhal, N. S. (2007). Vascular endothelial growth factor, hypoxia-inducible factor 1 alpha and CD34 expressions in early-stage gastric tumors: relationship with pathological factors and prognostic impact on survival. *Oncology, 72*(1-2), 111-117. doi:10.1159/000111118

Cannito, S., Novo, E., Compagnone, A., Valfre di Bonzo, L., Busletta, C., Zamara, E., . . . Parola, M. (2008). Redox mechanisms switch on hypoxia-dependent epithelial-mesenchymal transition in cancer cells. *Carcinogenesis, 29*(12), 2267-2278. doi:10.1093/carcin/bgn216

Carraro, A., Trevellin, E., Fassan, M., Kotsafti, A., Lunardi, F., Porzionato, A., . . . Scarpa, M. (2017). Esophageal adenocarcinoma microenvironment: Peritumoral

adipose tissue effects associated with chemoresistance. *Cancer Sci, 108*(12), 2393-2404. doi:10.1111/cas.13415

Celia-Terrassa, T., Meca-Cortes, O., Mateo, F., Martinez de Paz, A., Rubio, N., Arnal-Estape, A., . . . Thomson, T. M. (2012). Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells. *J Clin Invest, 122*(5), 1849-1868. doi:10.1172/jci59218

Chai, B., Guo, Y., Cui, X., Liu, J., Suo, Y., Dou, Z., & Li, N. (2019). MiR-223-3p promotes the proliferation, invasion and migration of colon cancer cells by negative regulating PRDM1. *Am J Transl Res, 11*(7), 4516-4523.

Chakraborty, C., Chin, K. Y., & Das, S. (2016). miRNA-regulated cancer stem cells: understanding the property and the role of miRNA in carcinogenesis. *Tumour Biol, 37*(10), 13039-13048. doi:10.1007/s13277-016-5156-1

Chambers, A. F., Groom, A. C., & MacDonald, I. C. (2002). Dissemination and growth of cancer cells in metastatic sites. *Nat Rev Cancer, 2*(8), 563-572. doi:10.1038/nrc865

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., & Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell, 113*(5), 643-655. doi:10.1016/s0092-8674(03)00392-1

Chen, C., Zhao, S., Karnad, A., & Freeman, J. W. (2018). The biology and role of CD44 in cancer progression: therapeutic implications. *J Hematol Oncol, 11*(1), 64. doi:10.1186/s13045-018-0605-5

Chen, K., Huang, Y. H., & Chen, J. L. (2013). Understanding and targeting cancer stem cells: therapeutic implications and challenges. *Acta Pharmacol Sin, 34*(6), 732-740. doi:10.1038/aps.2013.27

Chen, S., Hou, J. H., Feng, X. Y., Zhang, X. S., Zhou, Z. W., Yun, J. P., . . . Cai, M. Y. (2013). Clinicopathologic significance of putative stem cell marker, CD44 and CD133, in human gastric carcinoma. *J Surg Oncol, 107*(8), 799-806. doi:10.1002/jso.23337

Chen, T., Yang, K., Yu, J., Meng, W., Yuan, D., Bi, F., . . . Mo, X. (2012). Identification and expansion of cancer stem cells in tumor tissues and peripheral blood derived from gastric adenocarcinoma patients. *Cell Res, 22*(1), 248-258. doi:10.1038/cr.2011.109

Chen, T., You, Y., Jiang, H., & Wang, Z. Z. (2017). Epithelial-mesenchymal transition (EMT): A biological process in the development, stem cell differentiation, and tumorigenesis. *J Cell Physiol, 232*(12), 3261-3272. doi:10.1002/jcp.25797

Chen, Z., Saad, R., Jia, P., Peng, D., Zhu, S., Washington, M. K., . . . El-Rifai, W. (2013). Gastric adenocarcinoma has a unique microRNA signature not present in esophageal adenocarcinoma. *Cancer, 119*(11), 1985-1993. doi:10.1002/cncr.28002

Chen, Z., Xu, W. R., Qian, H., Zhu, W., Bu, X. F., Wang, S., . . . Xu, X. J. (2009). Oct4, a novel marker for human gastric cancer. *J Surg Oncol, 99*(7), 414-419. doi:10.1002/jso.21270

Cheng, C. W., Yu, J. C., Hsieh, Y. H., Liao, W. L., Shieh, J. C., Yao, C. C., . . . Shen, C. Y. (2018). Increased Cellular Levels of MicroRNA-9 and MicroRNA-221 Correlate with Cancer Stemness and Predict Poor Outcome in Human Breast Cancer. *Cell Physiol Biochem, 48*(5), 2205-2218. doi:10.1159/000492561

Chevallay, M., Bollschweiler, E., Chandramohan, S. M., Schmidt, T., Koch, O., Demanzoni, G., . . . Allum, W. (2018). Cancer of the gastroesophageal junction:

a diagnosis, classification, and management review. *Ann N Y Acad Sci, 1434*(1), 132-138. doi:10.1111/nyas.13954

Chow, W. H., Finkle, W. D., McLaughlin, J. K., Frankl, H., Ziel, H. K., & Fraumeni, J. F., Jr. (1995). The relation of gastroesophageal reflux disease and its treatment to adenocarcinomas of the esophagus and gastric cardia. *Jama, 274*(6), 474-477.

Cochrane, D. R., Howe, E. N., Spoelstra, N. S., & Richer, J. K. (2010). Loss of miR-200c: A Marker of Aggressiveness and Chemoresistance in Female Reproductive Cancers. *J Oncol, 2010*, 821717. doi:10.1155/2010/821717

Codd, A. S., Kanaseki, T., Torigo, T., & Tabi, Z. (2018). Cancer stem cells as targets for immunotherapy. *Immunology, 153*(3), 304-314. doi:10.1111/imm.12866

Cristescu, R., Lee, J., Nebozhyn, M., Kim, K. M., Ting, J. C., Wong, S. S., . . . Aggarwal, A. (2015). Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med, 21*(5), 449-456. doi:10.1038/nm.3850

Cunningham, D., Allum, W. H., Stenning, S. P., Thompson, J. N., Van de Velde, C. J., Nicolson, M., . . . Participants, M. T. (2006). Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. *N Engl J Med, 355*(1), 11-20. doi:10.1056/NEJMoa055531

Curtis, N. J., Noble, F., Bailey, I. S., Kelly, J. J., Byrne, J. P., & Underwood, T. J. (2014). The relevance of the Siewert classification in the era of multimodal therapy for adenocarcinoma of the gastro-oesophageal junction. *J Surg Oncol, 109*(3), 202-207. doi:10.1002/jso.23484

Dallas, N. A., Xia, L., Fan, F., Gray, M. J., Gaur, P., van Buren, G., 2nd, . . . Ellis, L. M. (2009). Chemoresistant colorectal cancer cells, the cancer stem cell

phenotype, and increased sensitivity to insulin-like growth factor-I receptor inhibition. *Cancer Res, 69*(5), 1951-1957. doi:10.1158/0008-5472.Can-08-2023

Dean, M., Fojo, T., & Bates, S. (2005). Tumour stem cells and drug resistance. *Nat Rev Cancer, 5*(4), 275-284. doi:10.1038/nrc1590

Denning, K. M., Smyth, P. C., Cahill, S. F., Finn, S. P., Conlon, E., Li, J., . . . Sheils, O. M. (2007). A molecular expression signature distinguishing follicular lesions in thyroid carcinoma using preamplification RT-PCR in archival samples. *Mod Pathol, 20*(10), 1095-1102. doi:10.1038/modpathol.3800943

Depypere, L., Moons, J., Lerut, T., De Hertogh, G., Sagaert, X., Coosemans, W., . . . Nafteux, P. (2016). Neoadjuvant chemoradiation treatment followed by surgery for esophageal cancer: there is much more than the mandard tumor regression score. *Acta Chir Belg, 116*(3), 149-155. doi:10.1080/00015458.2016.1212500

Diepenbruck, M., & Christofori, G. (2016). Epithelial-mesenchymal transition (EMT) and metastasis: yes, no, maybe? *Curr Opin Cell Biol, 43*, 7-13. doi:10.1016/j.ceb.2016.06.002

Ding, Y., Yu, A. Q., Wang, X. L., Guo, X. R., Yuan, Y. H., & Li, D. S. (2016). Forced expression of Nanog with mRNA synthesized in vitro to evaluate the malignancy of HeLa cells through acquiring cancer stem cell phenotypes. *Oncol Rep, 35*(5), 2643-2650. doi:10.3892/or.2016.4639

Dongre, A., & Weinberg, R. A. (2019). New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat Rev Mol Cell Biol, 20*(2), 69-84. doi:10.1038/s41580-018-0080-4

Driemel, C., Kremling, H., Schumacher, S., Will, D., Wolters, J., Lindenlauf, N., . . . Gires, O. (2014). Context-dependent adaption of EpCAM expression in early

systemic esophageal cancer. *Oncogene, 33*(41), 4904-4915. doi:10.1038/onc.2013.441

Duan, L., Shen, H., Zhao, G., Yang, R., Cai, X., Zhang, L., . . . Huang, Y. (2014). Inhibitory effect of Disulfiram/copper complex on non-small cell lung cancer cells. *Biochem Biophys Res Commun, 446*(4), 1010-1016. doi:10.1016/j.bbrc.2014.03.047

Dulak, A. M., Stojanov, P., Peng, S., Lawrence, M. S., Fox, C., Stewart, C., . . . Bass, A. J. (2013). Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat Genet, 45*(5), 478-486. doi:10.1038/ng.2591

Dykxhoorn, D. M., Wu, Y., Xie, H., Yu, F., Lal, A., Petrocca, F., . . . Lieberman, J. (2009). miR-200 enhances mouse breast cancer cell colonization to form distant metastases. *PLoS One, 4*(9), e7181. doi:10.1371/journal.pone.0007181

Esteller, M. (2008). Epigenetics in cancer. *N Engl J Med, 358*(11), 1148-1159. doi:10.1056/NEJMra072067

Eun, K., Ham, S. W., & Kim, H. (2017). Cancer stem cell heterogeneity: origin and new perspectives on CSC targeting. *BMB Rep, 50*(3), 117-125. doi:10.5483/bmbrep.2017.50.3.222

Eyler, C. E., & Rich, J. N. (2008). Survival of the fittest: cancer stem cells in therapeutic resistance and angiogenesis. *J Clin Oncol, 26*(17), 2839-2845. doi:10.1200/jco.2007.15.1829

Fan, D., Ren, B., Yang, X., Liu, J., & Zhang, Z. (2016). Upregulation of miR-501-5p activates the wnt/beta-catenin signaling pathway and enhances stem cell-like phenotype in gastric cancer. *J Exp Clin Cancer Res, 35*(1), 177. doi:10.1186/s13046-016-0432-x

Fang, C., Fan, C., Wang, C., Huang, Q., Meng, W., Yu, Y., . . . Zhou, Z. (2017). Prognostic value of CD133(+) CD54(+) CD44(+) circulating tumor cells in colorectal cancer with liver metastasis. *Cancer Med, 6*(12), 2850-2857. doi:10.1002/cam4.1241

Fang, L. L., Wang, X. H., Sun, B. F., Zhang, X. D., Zhu, X. H., Yu, Z. J., & Luo, H. (2017). Expression, regulation and mechanism of action of the miR-17-92 cluster in tumor cells (Review). *Int J Mol Med, 40*(6), 1624-1630. doi:10.3892/ijmm.2017.3164

Farace, C., Pisano, A., Grinan-Lison, C., Solinas, G., Jimenez, G., Serra, M., . . . Madeddu, R. (2020). Deregulation of cancer-stem-cell-associated miRNAs in tissues and sera of colorectal cancer patients. *Oncotarget, 11*(2), 116-130. doi:10.18632/oncotarget.27411

Faraldi, M., Gomarasca, M., Sansoni, V., Perego, S., Banfi, G., & Lombardi, G. (2019). Normalization strategies differently affect circulating miRNA profile associated with the training status. *Sci Rep, 9*(1), 1584. doi:10.1038/s41598-019-38505-x

Farmer, P., Bonnefoi, H., Anderle, P., Cameron, D., Wirapati, P., Becette, V., . . . Delorenzi, M. (2009). A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer. *Nat Med, 15*(1), 68-74. doi:10.1038/nm.1908

Fassan, M., Saraggi, D., Balsamo, L., Realdon, S., Scarpa, M., Castoro, C., . . . Rugge, M. (2017). Early miR-223 Upregulation in Gastroesophageal Carcinogenesis. *Am J Clin Pathol, 147*(3), 301-308. doi:10.1093/ajcp/aqx004

Fedchenko, N., & Reifenrath, J. (2014). Different approaches for interpretation and reporting of immunohistochemistry analysis results in the bone tissue - a review. *Diagn Pathol, 9*, 221. doi:10.1186/s13000-014-0221-9

Feith, M., Stein, H. J., & Siewert, J. R. (2006). Adenocarcinoma of the esophagogastric junction: surgical therapy based on 1602 consecutive resected patients. *Surg Oncol Clin N Am, 15*(4), 751-764. doi:10.1016/j.soc.2006.07.015

Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Pineros, M., . . . Bray, F. (2019). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer, 144*(8), 1941-1953. doi:10.1002/ijc.31937

Fitzgerald, T. L., & McCubrey, J. A. (2014). Pancreatic cancer stem cells: association with cell surface markers, prognosis, resistance, metastasis and treatment. *Adv Biol Regul, 56*, 45-50. doi:10.1016/j.jbior.2014.05.001

Fouad, Y. A., & Aanei, C. (2017). Revisiting the hallmarks of cancer. *Am J Cancer Res, 7*(5), 1016-1036.

Frankell, A. M., Jammula, S., Li, X., Contino, G., Killcoyne, S., Abbas, S., . . . Fitzgerald, R. C. (2019). The landscape of selection in 551 esophageal adenocarcinomas defines genomic biomarkers for the clinic. *Nat Genet, 51*(3), 506-516. doi:10.1038/s41588-018-0331-5

Fuchs, C. S., Doi, T., Jang, R. W., Muro, K., Satoh, T., Machado, M., . . . Yoon, H. H. (2018). Safety and Efficacy of Pembrolizumab Monotherapy in Patients With Previously Treated Advanced Gastric and Gastroesophageal Junction Cancer: Phase 2 Clinical KEYNOTE-059 Trial. *JAMA Oncol, 4*(5), e180013. doi:10.1001/jamaoncol.2018.0013

Fuchs, C. S., Tomasek, J., Yong, C. J., Dumitru, F., Passalacqua, R., Goswami, C., . . . Tabernero, J. (2014). Ramucirumab monotherapy for previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (REGARD):

an international, randomised, multicentre, placebo-controlled, phase 3 trial. *Lancet, 383*(9911), 31-39. doi:10.1016/s0140-6736(13)61719-5

Gaffney, E. F., Riegman, P. H., Grizzle, W. E., & Watson, P. H. (2018). Factors that drive the increasing use of FFPE tissue in basic and translational cancer research. *Biotech Histochem, 93*(5), 373-386. doi:10.1080/10520295.2018.1446101

Galvan, J. A., Zlobec, I., Wartenberg, M., Lugli, A., Gloor, B., Perren, A., & Karamitopoulou, E. (2015). Expression of E-cadherin repressors SNAIL, ZEB1 and ZEB2 by tumour and stromal cells influences tumour-budding phenotype and suggests heterogeneity of stromal cells in pancreatic cancer. *Br J Cancer, 112*(12), 1944-1950. doi:10.1038/bjc.2015.177

Gerlinger, M., Rowan, A. J., Horswell, S., Math, M., Larkin, J., Endesfelder, D., . . . Swanton, C. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med, 366*(10), 883-892. doi:10.1056/NEJMoa1113205

Gilder, A. S., Natali, L., Van Dyk, D. M., Zalfa, C., Banki, M. A., Pizzo, D. P., . . . Gonias, S. L. (2018). The Urokinase Receptor Induces a Mesenchymal Gene Expression Signature in Glioblastoma Cells and Promotes Tumor Cell Survival in Neurospheres. *Sci Rep, 8*(1), 2982. doi:10.1038/s41598-018-21358-1

Gonzalez, D. M., & Medici, D. (2014). Signaling mechanisms of the epithelial-mesenchymal transition. *Sci Signal, 7*(344), re8. doi:10.1126/scisignal.2005189

Goossens, S., Vandamme, N., Van Vlierberghe, P., & Berx, G. (2017). EMT transcription factors in cancer development re-evaluated: Beyond EMT and MET. *Biochim Biophys Acta Rev Cancer, 1868*(2), 584-591. doi:10.1016/j.bbcan.2017.06.006

Gort, E. H., Groot, A. J., van der Wall, E., van Diest, P. J., & Vooijs, M. A. (2008). Hypoxic regulation of metastasis via hypoxia-inducible factors. *Curr Mol Med, 8*(1), 60-67. doi:10.2174/156652408783565568

Graham, S. M., Jorgensen, H. G., Allan, E., Pearson, C., Alcorn, M. J., Richmond, L., & Holyoake, T. L. (2002). Primitive, quiescent, Philadelphia-positive stem cells from patients with chronic myeloid leukemia are insensitive to STI571 in vitro. *Blood, 99*(1), 319-325. doi:10.1182/blood.v99.1.319

Greally, M., Chou, J. F., Chatila, W. K., Margolis, M., Capanu, M., Hechtman, J. F., . . . Ku, G. Y. (2019). Clinical and Molecular Predictors of Response to Immune Checkpoint Inhibitors in Patients with Advanced Esophagogastric Cancer. *Clin Cancer Res, 25*(20), 6160-6169. doi:10.1158/1078-0432.Ccr-18-3603

Greenburg, G., & Hay, E. D. (1982). Epithelia suspended in collagen gels can lose polarity and express characteristics of migrating mesenchymal cells. *J Cell Biol, 95*(1), 333-339. doi:10.1083/jcb.95.1.333

Gregory, P. A., Bert, A. G., Paterson, E. L., Barry, S. C., Tsykin, A., Farshid, G., . . . Goodall, G. J. (2008). The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat Cell Biol, 10*(5), 593-601. doi:10.1038/ncb1722

Gubin, M. M., Zhang, X., Schuster, H., Caron, E., Ward, J. P., Noguchi, T., . . . Schreiber, R. D. (2014). Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature, 515*(7528), 577-581. doi:10.1038/nature13988

Habu, N., Imanishi, Y., Kameyama, K., Shimoda, M., Tokumaru, Y., Sakamoto, K., . . . Ogawa, K. (2015). Expression of Oct3/4 and Nanog in the head and neck squamous carcinoma cells and its clinical implications for delayed neck

metastasis in stage I/II oral tongue squamous cell carcinoma. *BMC Cancer, 15*, 730. doi:10.1186/s12885-015-1732-9

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell, 144*(5), 646-674. doi:10.1016/j.cell.2011.02.013

Hao, Y., Baker, D., & Ten Dijke, P. (2019). TGF-beta-Mediated Epithelial-Mesenchymal Transition and Cancer Metastasis. *Int J Mol Sci, 20*(11). doi:10.3390/ijms20112767

Haverkamp, L., Ruurda, J. P., van Leeuwen, M. S., Siersema, P. D., & van Hillegersberg, R. (2014). Systematic review of the surgical strategies of adenocarcinomas of the gastroesophageal junction. *Surg Oncol, 23*(4), 222-228. doi:10.1016/j.suronc.2014.10.004

Hayakawa, Y., Sethi, N., Sepulveda, A. R., Bass, A. J., & Wang, T. C. (2016). Oesophageal adenocarcinoma and gastric cancer: should we mind the gap? *Nat Rev Cancer, 16*(5), 305-318. doi:10.1038/nrc.2016.24

Hayeck, T. J., Kong, C. Y., Spechler, S. J., Gazelle, G. S., & Hur, C. (2010). The prevalence of Barrett's esophagus in the US: estimates from a simulation model confirmed by SEER data. *Dis Esophagus, 23*(6), 451-457. doi:10.1111/j.1442-2050.2010.01054.x

Hermanek, P., & Wittekind, C. (1994). The pathologist and the residual tumor (R) classification. *Pathol Res Pract, 190*(2), 115-123. doi:10.1016/s0344-0338(11)80700-4

Hermansen, S. K., Christensen, K. G., Jensen, S. S., & Kristensen, B. W. (2011). Inconsistent immunohistochemical expression patterns of four different CD133 antibody clones in glioblastoma. *J Histochem Cytochem, 59*(4), 391-407. doi:10.1369/0022155411400867

Hezova, R., Kovarikova, A., Srovnal, J., Zemanova, M., Harustiak, T., Ehrmann, J., . . . Slaby, O. (2016). MiR-205 functions as a tumor suppressor in adenocarcinoma and an oncogene in squamous cell carcinoma of esophagus. *Tumour Biol, 37*(6), 8007-8018. doi:10.1007/s13277-015-4656-8

Honing, J., Pavlov, K. V., Meijer, C., Smit, J. K., Boersma-van Ek, W., Karrenbeld, A., . . . Plukker, J. T. (2014). Loss of CD44 and SOX2 expression is correlated with a poor prognosis in esophageal adenocarcinoma patients. *Ann Surg Oncol, 21 Suppl 4*, S657-664. doi:10.1245/s10434-014-3763-x

Hou, Y. C., Chao, Y. J., Tung, H. L., Wang, H. C., & Shan, Y. S. (2014). Coexpression of CD44-positive/CD133-positive cancer stem cells and CD204-positive tumor-associated macrophages is a predictor of survival in pancreatic ductal adenocarcinoma. *Cancer, 120*(17), 2766-2777. doi:10.1002/cncr.28774

Housman, G., Byler, S., Heerboth, S., Lapinska, K., Longacre, M., Snyder, N., & Sarkar, S. (2014). Drug resistance in cancer: an overview. *Cancers (Basel), 6*(3), 1769-1792. doi:10.3390/cancers6031769

Hu, Y., Zang, J., Qin, X., Yan, D., Cao, H., Zhou, L., . . . Feng, J. F. (2017). Epithelial-to-mesenchymal transition correlates with gefitinib resistance in NSCLC cells and the liver X receptor ligand GW3965 reverses gefitinib resistance through inhibition of vimentin. *Onco Targets Ther, 10*, 2341-2348. doi:10.2147/ott.S124757

Huo, Y., Zheng, Z., Chen, Y., Wang, Q., Zhang, Z., & Deng, H. (2016). Downregulation of vimentin expression increased drug resistance in ovarian cancer cells. *Oncotarget, 7*(29), 45876-45888. doi:10.18632/oncotarget.9970

Imano, M., Itoh, T., Satou, T., Yasuda, A., Nishiki, K., Kato, H., . . . Shiozaki, H. (2013). High expression of epithelial cellular adhesion molecule in peritoneal

metastasis of gastric cancer. *Target Oncol, 8*(4), 231-235. doi:10.1007/s11523-012-0239-4

Ishiguro, H., Wakasugi, T., Terashita, Y., Sakamoto, N., Tanaka, T., Mizoguchi, K., . . . Takeyama, H. (2016). Decreased expression of CDH1 or CTNNB1 affects poor prognosis of patients with esophageal cancer. *World J Surg Oncol, 14*(1), 240. doi:10.1186/s12957-016-0956-8

Islam, F., Gopalan, V., & Lam, A. K. (2018). Identification of Cancer Stem Cells in Esophageal Adenocarcinoma. *Methods Mol Biol, 1756*, 165-176. doi:10.1007/978-1-4939-7734-5_15

Islami, F., & Kamangar, F. (2008). Helicobacter pylori and esophageal cancer risk: a meta-analysis. *Cancer Prev Res (Phila), 1*(5), 329-338. doi:10.1158/1940-6207.Capr-08-0109

Janjigian, Y. Y., Bendell, J., Calvo, E., Kim, J. W., Ascierto, P. A., Sharma, P., . . . Le, D. T. (2018). CheckMate-032 Study: Efficacy and Safety of Nivolumab and Nivolumab Plus Ipilimumab in Patients With Metastatic Esophagogastric Cancer. *J Clin Oncol, 36*(28), 2836-2844. doi:10.1200/jco.2017.76.6212

Janjigian, Y. Y., Chou, J. F., Simmons, M., Momtaz, P., Sanchez-Vega, F., Shcherba, M., . . . Hechtman, J. F. (2019). First-line pembrolizumab (P), trastuzumab (T), capecitabine (C) and oxaliplatin (O) in HER2-positive metastatic esophagogastric adenocarcinoma (mEGA). *Journal of Clinical Oncology, 37*(4_suppl), 62-62. doi:10.1200/JCO.2019.37.4_suppl.62

Janjigian, Y. Y., Sanchez-Vega, F., Jonsson, P., Chatila, W. K., Hechtman, J. F., Ku, G. Y., . . . Schultz, N. (2018). Genetic Predictors of Response to Systemic Therapy in Esophagogastric Cancer. *Cancer Discov, 8*(1), 49-58. doi:10.1158/2159-8290.Cd-17-0787

Jiang, X., & Wang, Z. (2018). miR-16 targets SALL4 to repress the proliferation and migration of gastric cancer. *Oncol Lett, 16*(3), 3005-3012. doi:10.3892/ol.2018.8997

Joshi, S. S., Maron, S. B., & Catenacci, D. V. (2018). Pembrolizumab for treatment of advanced gastric and gastroesophageal junction adenocarcinoma. *Future Oncol, 14*(5), 417-430. doi:10.2217/fon-2017-0436

Kalluri, R., & Weinberg, R. A. (2009). The basics of epithelial-mesenchymal transition. *J Clin Invest, 119*(6), 1420-1428. doi:10.1172/jci39104

Kamangar, F., Dawsey, S. M., Blaser, M. J., Perez-Perez, G. I., Pietinen, P., Newschaffer, C. J., . . . Taylor, P. R. (2006). Opposing risks of gastric cardia and noncardia gastric adenocarcinomas associated with Helicobacter pylori seropositivity. *J Natl Cancer Inst, 98*(20), 1445-1452. doi:10.1093/jnci/djj393

Kang, Y. K., Boku, N., Satoh, T., Ryu, M. H., Chao, Y., Kato, K., . . . Chen, L. T. (2017). Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet, 390*(10111), 2461-2471. doi:10.1016/s0140-6736(17)31827-5

Kao, S. H., Wang, W. L., Chen, C. Y., Chang, Y. L., Wu, Y. Y., Wang, Y. T., . . . Yang, P. C. (2014). GSK3beta controls epithelial-mesenchymal transition and tumor metastasis by CHIP-mediated degradation of Slug. *Oncogene, 33*(24), 3172-3182. doi:10.1038/onc.2013.279

Katsuno, Y., Ehata, S., Yashiro, M., Yanagihara, K., Hirakawa, K., & Miyazono, K. (2012). Coordinated expression of REG4 and aldehyde dehydrogenase 1 regulating tumourigenic capacity of diffuse-type gastric carcinoma-initiating

cells is inhibited by TGF-beta. *J Pathol*, *228*(3), 391-404. doi:10.1002/path.4020

Kaufmann, M., Heider, K. H., Sinn, H. P., von Minckwitz, G., Ponta, H., & Herrlich, P. (1995). CD44 variant exon epitopes in primary breast cancer and length of survival. *Lancet, 345*(8950), 615-619. doi:10.1016/s0140-6736(95)90521-9

Kelly, R. J., Lockhart, A. C., Jonker, D. J., Melichar, B., Andre, T., Chau, I., . . . Moehler, M. (2017). CheckMate 577: A randomized, double-blind, phase 3 study of nivolumab (Nivo) or placebo in patients (Pts) with resected lower esophageal (E) or gastroesophageal junction (GEJ) cancer. *Journal of Clinical Oncology, 35*(4_suppl), TPS212-TPS212. doi:10.1200/JCO.2017.35.4_suppl.TPS212

Khan, A. Q., Ahmed, E. I., Elareer, N. R., Junejo, K., Steinhoff, M., & Uddin, S. (2019). Role of miRNA-Regulated Cancer Stem Cells in the Pathogenesis of Human Malignancies. *Cells, 8*(8). doi:10.3390/cells8080840

Kim, J., Bowlby, R., Mungall, A. et al. (2017). Integrated genomic characterization of oesophageal carcinoma. *Nature, 541*(7636), 169-175. doi:10.1038/nature20805

Kim, J. Y., Bae, B. N., Kim, K. S., Shin, E., & Park, K. (2009). Osteopontin, CD44, and NFkappaB expression in gastric adenocarcinoma. *Cancer Res Treat, 41*(1), 29-35. doi:10.4143/crt.2009.41.1.29

Knoll, S., Furst, K., Kowtharapu, B., Schmitz, U., Marquardt, S., Wolkenhauer, O., . . . Putzer, B. M. (2014). E2F1 induces miR-224/452 expression to drive EMT through TXNIP downregulation. *EMBO Rep, 15*(12), 1315-1329. doi:10.15252/embr.201439392

Komatsu, S., Ichikawa, D., Takeshita, H., Tsujiura, M., Morimura, R., Nagata, H., . . . Otsuji, E. (2011). Circulating microRNAs in plasma of patients with

oesophageal squamous cell carcinoma. *Br J Cancer, 105*(1), 104-111. doi:10.1038/bjc.2011.198

Kumamoto, T., Kurahashi, Y., Niwa, H., Nakanishi, Y., Okumura, K., Ozawa, R., . . . Shinohara, H. (2019). True esophagogastric junction adenocarcinoma: background of its definition and current surgical trends. *Surg Today*. doi:10.1007/s00595-019-01843-4

Kuroda, T., Hirohashi, Y., Torigoe, T., Yasuda, K., Takahashi, A., Asanuma, H., . . . Sato, N. (2013). ALDH1-high ovarian cancer stem-like cells can be isolated from serous and clear cell adenocarcinoma cells, and ALDH1 high expression is associated with poor prognosis. *PLoS One, 8*(6), e65158. doi:10.1371/journal.pone.0065158

Lagergren, J., Bergstrom, R., Lindgren, A., & Nyren, O. (1999). Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. *N Engl J Med, 340*(11), 825-831. doi:10.1056/nejm199903183401101

Lamouille, S., Xu, J., & Derynck, R. (2014). Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol, 15*(3), 178-196. doi:10.1038/nrm3758

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., . . . Tuschl, T. (2007). A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell, 129*(7), 1401-1414. doi:10.1016/j.cell.2007.04.040

Lapidot, T., Sirard, C., Vormoor, J., Murdoch, B., Hoang, T., Caceres-Cortes, J., . . . Dick, J. E. (1994). A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature, 367*(6464), 645-648. doi:10.1038/367645a0

Lau, W. M., Teng, E., Chong, H. S., Lopez, K. A., Tay, A. Y., Salto-Tellez, M., . . . Chan, S. L. (2014). CD44v8-10 is a cancer-specific marker for gastric cancer stem cells. *Cancer Res, 74*(9), 2630-2641. doi:10.1158/0008-5472.Can-13-2309

Le, D. T., Uram, J. N., Wang, H., Bartlett, B. R., Kemberling, H., Eyring, A. D., . . . Diaz, L. A., Jr. (2015). PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med, 372*(26), 2509-2520. doi:10.1056/NEJMoa1500596

Le Magnen, C., Bubendorf, L., Rentsch, C. A., Mengus, C., Gsponer, J., Zellweger, T., . . . Spagnoli, G. C. (2013). Characterization and clinical relevance of ALDHbright populations in prostate cancer. *Clin Cancer Res, 19*(19), 5361-5371. doi:10.1158/1078-0432.Ccr-12-2857

Lee, H. H., Seo, K. J., An, C. H., Kim, J. S., & Jeon, H. M. (2012). CD133 expression is correlated with chemoresistance and early recurrence of gastric cancer. *J Surg Oncol, 106*(8), 999-1004. doi:10.1002/jso.23178

Levina, V., Marrangoni, A. M., DeMarco, R., Gorelik, E., & Lokshin, A. E. (2008). Drug-selected human lung cancer stem cells: cytokine network, tumorigenic and metastatic properties. *PLoS One, 3*(8), e3077. doi:10.1371/journal.pone.0003077

Li, B., Lu, Y., Wang, H., Han, X., Mao, J., Li, J., . . . Song, B. (2016). miR-221/222 enhance the tumorigenicity of human breast cancer stem cells via modulation of PTEN/Akt pathway. *Biomed Pharmacother, 79*, 93-101. doi:10.1016/j.biopha.2016.01.045

Li, H. Z., Yi, T. B., & Wu, Z. Y. (2008). Suspension culture combined with chemotherapeutic agents for sorting of breast cancer stem cells. *BMC Cancer, 8*, 135. doi:10.1186/1471-2407-8-135

Li, J., Qin, S., Xu, J., Xiong, J., Wu, C., Bai, Y., . . . Yu, H. (2016). Randomized, Double-Blind, Placebo-Controlled Phase III Trial of Apatinib in Patients With Chemotherapy-Refractory Advanced or Metastatic Adenocarcinoma of the Stomach or Gastroesophageal Junction. *J Clin Oncol*, *34*(13), 1448-1454. doi:10.1200/jco.2015.63.5995

Li, K., Guo, X., Wang, Z., Li, X., Bu, Y., Bai, X., . . . Huang, Y. (2016). The prognostic roles of ALDH1 isoenzymes in gastric cancer. *Onco Targets Ther*, *9*, 3405-3414. doi:10.2147/ott.S102314

Li, N., Deng, W., Ma, J., Wei, B., Guo, K., Shen, W., . . . Luo, S. (2015). Prognostic evaluation of Nanog, Oct4, Sox2, PCNA, Ki67 and E-cadherin expression in gastric cancer. *Med Oncol*, *32*(1), 433. doi:10.1007/s12032-014-0433-6

Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., . . . Pei, D. (2010). A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell*, *7*(1), 51-63. doi:10.1016/j.stem.2010.04.014

Li, S., & Li, Q. (2014). Cancer stem cells and tumor metastasis (Review). *Int J Oncol*, *44*(6), 1806-1812. doi:10.3892/ijo.2014.2362

Li, Y., Choi, P. S., Casey, S. C., Dill, D. L., & Felsher, D. W. (2014). MYC through miR-17-92 suppresses specific target genes to maintain survival, autonomous proliferation, and a neoplastic state. *Cancer Cell*, *26*(2), 262-272. doi:10.1016/j.ccr.2014.06.014

Li, Z., & Guo, Z. (2019). Comparison of CDH1 Gene Hypermethylation Status in Blood and Serum among Gastric Cancer Patients. *Pathol Oncol Res*. doi:10.1007/s12253-019-00658-5

Liao, T. T., & Yang, M. H. (2017). Revisiting epithelial-mesenchymal transition in cancer metastasis: the connection between epithelial plasticity and stemness. *Mol Oncol, 11*(7), 792-804. doi:10.1002/1878-0261.12096

Lichner, Z., Saleh, C., Subramaniam, V., Seivwright, A., Prud'homme, G. J., & Yousef, G. M. (2015). miR-17 inhibition enhances the formation of kidney cancer spheres with stem cell/ tumor initiating cell properties. *Oncotarget, 6*(8), 5567-5581. doi:10.18632/oncotarget.1901

Lim, S., Becker, A., Zimmer, A., Lu, J., Buettner, R., & Kirfel, J. (2013). SNAI1-mediated epithelial-mesenchymal transition confers chemoresistance and cellular plasticity by regulating genes involved in cell death and stem cell maintenance. *PLoS One, 8*(6), e66558. doi:10.1371/journal.pone.0066558

Lin, D., Khan, U., Goetze, T. O., Reizine, N., Goodman, K. A., Shah, M. A., . . . Posey, J. A. (2019). Gastroesophageal Junction Adenocarcinoma: Is There an Optimal Management? *Am Soc Clin Oncol Educ Book, 39*, e88-e95. doi:10.1200/edbk_236827

Liu, B., Wu, S., Ma, J., Yan, S., Xiao, Z., Wan, L., . . . Mao, A. (2018). lncRNA GAS5 Reverses EMT and Tumor Stem Cell-Mediated Gemcitabine Resistance and Metastasis by Targeting miR-221/SOCS3 in Pancreatic Cancer. *Mol Ther Nucleic Acids, 13*, 472-482. doi:10.1016/j.omtn.2018.09.026

Liu, J., Ma, L., Xu, J., Liu, C., Zhang, J., Liu, J., . . . Zhou, Y. (2013). Spheroid body-forming cells in the human gastric cancer cell line MKN-45 possess cancer stem cell properties. *Int J Oncol, 42*(2), 453-459. doi:10.3892/ijo.2012.1720

Liu, J., Sun, X., Qin, S., Wang, H., Du, N., Li, Y., . . . Ren, H. (2016). CDH1 promoter methylation correlates with decreased gene expression and poor prognosis in

patients with breast cancer. *Oncol Lett, 11*(4), 2635-2643. doi:10.3892/ol.2016.4274

Liu, Y. J., Yan, P. S., Li, J., & Jia, J. F. (2005). Expression and significance of CD44s, CD44v6, and nm23 mRNA in human cancer. *World J Gastroenterol, 11*(42), 6601-6606. doi:10.3748/wjg.v11.i42.6601

Lu, H., Tran, L., Park, Y., Chen, I., Lan, J., Xie, Y., & Semenza, G. L. (2018). Reciprocal Regulation of DUSP9 and DUSP16 Expression by HIF1 Controls ERK and p38 MAP Kinase Activity and Mediates Chemotherapy-Induced Breast Cancer Stem Cell Enrichment. *Cancer Res, 78*(15), 4191-4202. doi:10.1158/0008-5472.Can-18-0270

Lu, L., Wu, M., Sun, L., Li, W., Fu, W., Zhang, X., & Liu, T. (2016). Clinicopathological and prognostic significance of cancer stem cell markers CD44 and CD133 in patients with gastric cancer: A comprehensive meta-analysis with 4729 patients involved. *Medicine (Baltimore), 95*(42), e5163. doi:10.1097/md.0000000000005163

Luther, J., Owyang, S. Y., Takeuchi, T., Cole, T. S., Zhang, M., Liu, M., . . . Kao, J. Y. (2011). Helicobacter pylori DNA decreases pro-inflammatory cytokine production by dendritic cells and attenuates dextran sodium sulphate-induced colitis. *Gut, 60*(11), 1479-1486. doi:10.1136/gut.2010.220087

Lynam-Lennon, N., Heavey, S., Sommerville, G., Bibby, B. A., Ffrench, B., Quinn, J., . . . Maher, S. G. (2017). MicroRNA-17 is downregulated in esophageal adenocarcinoma cancer stem-like cells and promotes a radioresistant phenotype. *Oncotarget, 8*(7), 11400-11413. doi:10.18632/oncotarget.13940

Ma, J., Fang, B., Zeng, F., Ma, C., Pang, H., Cheng, L., . . . Wang, Z. (2015). Down-regulation of miR-223 reverses epithelial-mesenchymal transition in

gemcitabine-resistant pancreatic cancer cells. *Oncotarget, 6*(3), 1740-1749. doi:10.18632/oncotarget.2714

MacDonagh, L., Gallagher, M. F., Ffrench, B., Gasch, C., Breen, E., Gray, S. G., . . . Barr, M. P. (2017). Targeting the cancer stem cell marker, aldehyde dehydrogenase 1, to circumvent cisplatin resistance in NSCLC. *Oncotarget, 8*(42), 72544-72563. doi:10.18632/oncotarget.19881

Madjd, Z., Erfani, E., Gheytanchi, E., Moradi-Lakeh, M., Shariftabrizi, A., & Asadi-Lari, M. (2016). Expression of CD133 cancer stem cell marker in melanoma: a systematic review and meta-analysis. *Int J Biol Markers, 31*(2), e118-125. doi:10.5301/jbm.5000209

Mani, S. A., Guo, W., Liao, M. J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., . . . Weinberg, R. A. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell, 133*(4), 704-715. doi:10.1016/j.cell.2008.03.027

Matsuoka, J., Yashiro, M., Sakurai, K., Kubo, N., Tanaka, H., Muguruma, K., . . . Hirakawa, K. (2012). Role of the stemness factors sox2, oct3/4, and nanog in gastric carcinoma. *J Surg Res, 174*(1), 130-135. doi:10.1016/j.jss.2010.11.903

Mayer, B., Jauch, K. W., Gunthert, U., Figdor, C. G., Schildberg, F. W., Funke, I., & Johnson, J. P. (1993). De-novo expression of CD44 and survival in gastric cancer. *Lancet, 342*(8878), 1019-1022. doi:10.1016/0140-6736(93)92879-x

Medema, J. P. (2013). Cancer stem cells: the challenges ahead. *Nat Cell Biol, 15*(4), 338-344. doi:10.1038/ncb2717

Medici, D., Hay, E. D., & Olsen, B. R. (2008). Snail and Slug promote epithelial-mesenchymal transition through beta-catenin-T-cell factor-4-dependent

expression of transforming growth factor-beta3. *Mol Biol Cell, 19*(11), 4875-4887. doi:10.1091/mbc.E08-05-0506

Mens, M. M. J., & Ghanbari, M. (2018). Cell Cycle Regulation of Stem Cells by MicroRNAs. *Stem Cell Rev Rep, 14*(3), 309-322. doi:10.1007/s12015-018-9808-y

Miyoshi, N., Fujino, S., Ohue, M., Yasui, M., Takahashi, Y., Sugimura, K., . . . Yano, M. (2018). The POU5F1 gene expression in colorectal cancer: a novel prognostic marker. *Surg Today, 48*(7), 709-715. doi:10.1007/s00595-018-1644-9

Mo, J. S., Park, H. W., & Guan, K. L. (2014). The Hippo signaling pathway in stem cell biology and cancer. *EMBO Rep, 15*(6), 642-656. doi:10.15252/embr.201438638

Mokrowiecka, A., Veits, L., Falkeis, C., Musial, J., Kordek, R., Lochowski, M., . . . Malecka-Panas, E. (2017). Expression profiles of cancer stem cell markers: CD133, CD44, Musashi-1 and EpCAM in the cardiac mucosa-Barrett's esophagus-early esophageal adenocarcinoma-advanced esophageal adenocarcinoma sequence. *Pathol Res Pract, 213*(3), 205-209. doi:10.1016/j.prp.2016.12.018

Moons, L. M., Kusters, J. G., van Delft, J. H., Kuipers, E. J., Gottschalk, R., Geldof, H., . . . Siersema, P. D. (2008). A pro-inflammatory genotype predisposes to Barrett's esophagus. *Carcinogenesis, 29*(5), 926-931. doi:10.1093/carcin/bgm241

Morel, A. P., Lievre, M., Thomas, C., Hinkal, G., Ansieau, S., & Puisieux, A. (2008). Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PLoS One, 3*(8), e2888. doi:10.1371/journal.pone.0002888

Mukohyama, J., Isobe, T., Hu, Q., Hayashi, T., Watanabe, T., Maeda, M., . . . Shimono, Y. (2019). miR-221 Targets QKI to Enhance the Tumorigenic Capacity of Human Colorectal Cancer Stem Cells. *Cancer Res, 79*(20), 5151-5158. doi:10.1158/0008-5472.Can-18-3544

Mulder, J. W., Kruyt, P. M., Sewnath, M., Oosting, J., Seldenrijk, C. A., Weidema, W. F., . . . Pals, S. T. (1994). Colorectal cancer prognosis and expression of exon-v6-containing CD44 proteins. *Lancet, 344*(8935), 1470-1472. doi:10.1016/s0140-6736(94)90290-9

Muro, K., Chung, H. C., Shankaran, V., Geva, R., Catenacci, D., Gupta, S., . . . Bang, Y. J. (2016). Pembrolizumab for patients with PD-L1-positive advanced gastric cancer (KEYNOTE-012): a multicentre, open-label, phase 1b trial. *Lancet Oncol, 17*(6), 717-726. doi:10.1016/s1470-2045(16)00175-3

NCRI. (2019). Cancer Factsheet Oesophagus. Retrieved from https://www.ncri.ie/sites/ncri/files/factsheets/Factsheet%20Oesophagus.pdf

Nechushtan, H., Hamamreh, Y., Nidal, S., Gotfried, M., Baron, A., Shalev, Y. I., . . . Peylan-Ramu, N. (2015). A phase IIb trial assessing the addition of disulfiram to chemotherapy for the treatment of metastatic non-small cell lung cancer. *Oncologist, 20*(4), 366-367. doi:10.1634/theoncologist.2014-0424

Nelson, W. J., & Nusse, R. (2004). Convergence of Wnt, beta-catenin, and cadherin pathways. *Science, 303*(5663), 1483-1487. doi:10.1126/science.1094291

Ni, J., Cozzi, P. J., Hao, J. L., Beretov, J., Chang, L., Duan, W., . . . Li, Y. (2014). CD44 variant 6 is associated with prostate cancer metastasis and chemo-/radioresistance. *Prostate, 74*(6), 602-617. doi:10.1002/pros.22775

Niehrs, C. (2012). The complex world of WNT receptor signalling. *Nat Rev Mol Cell Biol, 13*(12), 767-779. doi:10.1038/nrm3470

Nunes, T., Hamdan, D., Leboeuf, C., El Bouchtaoui, M., Gapihan, G., Nguyen, T. T., . . . Janin, A. (2018). Targeting Cancer Stem Cells to Overcome Chemoresistance. *Int J Mol Sci, 19*(12). doi:10.3390/ijms19124036

Ocana, O. H., Corcoles, R., Fabra, A., Moreno-Bueno, G., Acloque, H., Vega, S., . . . Nieto, M. A. (2012). Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. *Cancer Cell, 22*(6), 709-724. doi:10.1016/j.ccr.2012.10.012

Odenthal, M., Hee, J., Gockel, I., Sisic, L., Schmitz, J., Stoecklein, N. H., . . . Vallbohmer, D. (2015). Serum microRNA profiles as prognostic/predictive markers in the multimodality therapy of locally advanced adenocarcinomas of the gastroesophageal junction. *Int J Cancer, 137*(1), 230-237. doi:10.1002/ijc.29363

Otsubo, T., Akiyama, Y., Yanagihara, K., & Yuasa, Y. (2008). SOX2 is frequently downregulated in gastric cancers and inhibits cell growth through cell-cycle arrest and apoptosis. *Br J Cancer, 98*(4), 824-831. doi:10.1038/sj.bjc.6604193

Ouyang, H., Gore, J., Deitz, S., & Korc, M. (2014). microRNA-10b enhances pancreatic cancer cell invasion by suppressing TIP30 expression and promoting EGF and TGF-beta actions. *Oncogene, 33*(38), 4664-4674. doi:10.1038/onc.2013.405

Ozawa, M., Ichikawa, Y., Zheng, Y. W., Oshima, T., Miyata, H., Nakazawa, K., . . . Taniguchi, H. (2014). Prognostic significance of CD44 variant 2 upregulation in colorectal cancer. *Br J Cancer, 111*(2), 365-374. doi:10.1038/bjc.2014.253

Pan, Y., Shu, X., Sun, L., Yu, L., Sun, L., Yang, Z., & Ran, Y. (2017). miR196a5p modulates gastric cancer stem cell characteristics by targeting Smad4. *Int J Oncol, 50*(6), 1965-1976. doi:10.3892/ijo.2017.3965

Paulson, T. G., Xu, L., Sanchez, C., Blount, P. L., Ayub, K., Odze, R. D., & Reid, B. J. (2006). Neosquamous epithelium does not typically arise from Barrett's epithelium. *Clin Cancer Res, 12*(6), 1701-1706. doi:10.1158/1078-0432.Ccr-05-1810

Pectasides, E., Stachler, M. D., Derks, S., Liu, Y., Maron, S., Islam, M., . . . Catenacci, D. V. (2018). Genomic Heterogeneity as a Barrier to Precision Medicine in Gastroesophageal Adenocarcinoma. *Cancer Discov, 8*(1), 37-48. doi:10.1158/2159-8290.Cd-17-0395

Peltier, H. J., & Latham, G. J. (2008). Normalization of microRNA expression levels in quantitative RT-PCR assays: identification of suitable reference RNA targets in normal and cancerous human solid tissues. *Rna, 14*(5), 844-852. doi:10.1261/rna.939908

Peters, Y., Al-Kaabi, A., Shaheen, N. J., Chak, A., Blum, A., Souza, R. F., . . . Siersema, P. D. (2019). Barrett oesophagus. *Nat Rev Dis Primers, 5*(1), 35. doi:10.1038/s41572-019-0086-z

Pistore, C., Giannoni, E., Colangelo, T., Rizzo, F., Magnani, E., Muccillo, L., . . . Bonapace, I. M. (2017). DNA methylation variations are required for epithelial-to-mesenchymal transition induced by cancer-associated fibroblasts in prostate cancer cells. *Oncogene, 36*(40), 5551-5566. doi:10.1038/onc.2017.159

Plaks, V., Kong, N., & Werb, Z. (2015). The cancer stem cell niche: how essential is the niche in regulating stemness of tumor cells? *Cell Stem Cell, 16*(3), 225-238. doi:10.1016/j.stem.2015.02.015

Polyak, K., & Weinberg, R. A. (2009). Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat Rev Cancer, 9*(4), 265-273. doi:10.1038/nrc2620

Ponta, H., Sherman, L., & Herrlich, P. A. (2003). CD44: from adhesion molecules to signalling regulators. *Nat Rev Mol Cell Biol, 4*(1), 33-45. doi:10.1038/nrm1004

Prasetyanti, P. R., & Medema, J. P. (2017). Intra-tumor heterogeneity from a cancer stem cell perspective. *Mol Cancer, 16*(1), 41. doi:10.1186/s12943-017-0600-4

Prat, A., Parker, J. S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J. I., . . . Perou, C. M. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res, 12*(5), R68. doi:10.1186/bcr2635

Prieto-Vila, M., Takahashi, R. U., Usuba, W., Kohama, I., & Ochiya, T. (2017). Drug Resistance Driven by Cancer Stem Cells and Their Niche. *Int J Mol Sci, 18*(12). doi:10.3390/ijms18122574

Quail, D. F., Taylor, M. J., & Postovit, L. M. (2012). Microenvironmental regulation of cancer stem cell phenotypes. *Curr Stem Cell Res Ther, 7*(3), 197-216. doi:10.2174/157488812799859838

Quante, M., Bhagat, G., Abrams, J. A., Marache, F., Good, P., Lee, M. D., . . . Wang, T. C. (2012). Bile acid and inflammation activate gastric cardia stem cells in a mouse model of Barrett-like metaplasia. *Cancer Cell, 21*(1), 36-51. doi:10.1016/j.ccr.2011.12.004

Raimondi, C., Gradilone, A., Naso, G., Vincenzi, B., Petracca, A., Nicolazzo, C., . . . Gazzaniga, P. (2011). Epithelial-mesenchymal transition and stemness features in circulating tumor cells from breast cancer patients. *Breast Cancer Res Treat, 130*(2), 449-455. doi:10.1007/s10549-011-1373-x

Rawla, P., & Barsouk, A. (2019). Epidemiology of gastric cancer: global trends, risk factors and prevention. *Prz Gastroenterol, 14*(1), 26-38. doi:10.5114/pg.2018.80001

Reya, T., Morrison, S. J., Clarke, M. F., & Weissman, I. L. (2001). Stem cells, cancer, and cancer stem cells. *Nature, 414*(6859), 105-111. doi:10.1038/35102167

Rice, T. W., Patil, D. T., & Blackstone, E. H. (2017). 8th edition AJCC/UICC staging of cancers of the esophagus and esophagogastric junction: application to clinical practice. *Ann Cardiothorac Surg, 6*(2), 119-130. doi:10.21037/acs.2017.03.14

Rodriguez, E., Chen, L., Ao, M. H., Geddes, S., Gabrielson, E., Askin, F., . . . Li, Q. K. (2014). Expression of transcript factors SALL4 and OCT4 in a subset of non-small cell lung carcinomas (NSCLC). *Transl Respir Med, 2*(1), 10. doi:10.1186/s40247-014-0010-7

Roscigno, G., Quintavalle, C., Donnarumma, E., Puoti, I., Diaz-Lagares, A., Iaboni, M., . . . Condorelli, G. (2016). MiR-221 promotes stemness of breast cancer cells by targeting DNMT3b. *Oncotarget, 7*(1), 580-592. doi:10.18632/oncotarget.5979

Roy, M., Connor, J., Al-Niaimi, A., Rose, S. L., & Mahajan, A. (2018). Aldehyde dehydrogenase 1A1 (ALDH1A1) expression by immunohistochemistry is associated with chemo-refractoriness in patients with high-grade ovarian serous carcinoma. *Hum Pathol, 73*, 1-6. doi:10.1016/j.humpath.2017.06.025

Rubenstein, J. H., & Shaheen, N. J. (2015). Epidemiology, Diagnosis, and Management of Esophageal Adenocarcinoma. *Gastroenterology, 149*(2), 302-317.e301. doi:10.1053/j.gastro.2015.04.053

Runge, T. M., Abrams, J. A., & Shaheen, N. J. (2015). Epidemiology of Barrett's Esophagus and Esophageal Adenocarcinoma. *Gastroenterol Clin North Am, 44*(2), 203-231. doi:10.1016/j.gtc.2015.02.001

Ruscetti, M., Quach, B., Dadashian, E. L., Mulholland, D. J., & Wu, H. (2015). Tracking and Functional Characterization of Epithelial-Mesenchymal Transition

and Mesenchymal Tumor Cells during Prostate Cancer Metastasis. *Cancer Res, 75*(13), 2749-2759. doi:10.1158/0008-5472.Can-14-3476

Saad, R., Chen, Z., Zhu, S., Jia, P., Zhao, Z., Washington, M. K., . . . El-Rifai, W. (2013). Deciphering the unique microRNA signature in human esophageal adenocarcinoma. *PLoS One, 8*(5), e64463. doi:10.1371/journal.pone.0064463

Schech, A., Kazi, A., Yu, S., Shah, P., & Sabnis, G. (2015). Histone Deacetylase Inhibitor Entinostat Inhibits Tumor-Initiating Cells in Triple-Negative Breast Cancer Cells. *Mol Cancer Ther, 14*(8), 1848-1857. doi:10.1158/1535-7163.Mct-14-0778

Schmidt, J. M., Panzilius, E., Bartsch, H. S., Irmler, M., Beckers, J., Kari, V., . . . Scheel, C. H. (2015). Stem-cell-like properties and epithelial plasticity arise as stable traits after transient Twist1 activation. *Cell Rep, 10*(2), 131-139. doi:10.1016/j.celrep.2014.12.032

Schneider, J. L., & Corley, D. A. (2017). The Troublesome Epidemiology of Barrett's Esophagus and Esophageal Adenocarcinoma. *Gastrointest Endosc Clin N Am, 27*(3), 353-364. doi:10.1016/j.giec.2017.03.002

Secrier, M., Li, X., de Silva, N., Eldridge, M. D., Contino, G., Bornschein, J., . . . Fitzgerald, R. C. (2016). Mutational signatures in esophageal adenocarcinoma define etiologically distinct subgroups with therapeutic relevance. *Nat Genet, 48*(10), 1131-1141. doi:10.1038/ng.3659

SEER, N. (2019). Cancer Stat Facts: Esophageal Cancer. Retrieved from https://seer.cancer.gov/statfacts/html/esoph.html

Shajari, E., & Mollasalehi, H. (2020). Ribonucleic-acid-biomarker candidates for early-phase group detection of common cancers. *Genomics, 112*(1), 163-168. doi:10.1016/j.ygeno.2018.08.011

Sheedy, P., & Medarova, Z. (2018). The fundamental role of miR-10b in metastatic cancer. *Am J Cancer Res, 8*(9), 1674-1688.

Shen, C., Zhang, H., Wang, P., Feng, J., Li, J., Xu, Y., . . . Fang, D. (2016). Deoxycholic acid (DCA) confers an intestinal phenotype on esophageal squamous epithelium via induction of the stemness-associated reprogramming factors OCT4 and SOX2. *Cell Cycle, 15*(11), 1439-1449. doi:10.1080/15384101.2016.1175252

Shibue, T., & Weinberg, R. A. (2009). Integrin beta1-focal adhesion kinase signaling directs the proliferation of metastatic cancer cells disseminated in the lungs. *Proc Natl Acad Sci U S A, 106*(25), 10290-10295. doi:10.1073/pnas.0904227106

Shibue, T., & Weinberg, R. A. (2017). EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. *Nat Rev Clin Oncol, 14*(10), 611-629. doi:10.1038/nrclinonc.2017.44

Shipitsin, M., Campbell, L. L., Argani, P., Weremowicz, S., Bloushtain-Qimron, N., Yao, J., . . . Polyak, K. (2007). Molecular definition of breast tumor heterogeneity. *Cancer Cell, 11*(3), 259-273. doi:10.1016/j.ccr.2007.01.013

Siewert, J. R., & Stein, H. J. (1998). Classification of adenocarcinoma of the oesophagogastric junction. *Br J Surg, 85*(11), 1457-1459. doi:10.1046/j.1365-2168.1998.00940.x

Singh, A., & Settleman, J. (2010). EMT, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer. *Oncogene, 29*(34), 4741-4751. doi:10.1038/onc.2010.215

Skubitz, A. P., Taras, E. P., Boylan, K. L., Waldron, N. N., Oh, S., Panoskaltsis-Mortari, A., & Vallera, D. A. (2013). Targeting CD133 in an in vivo ovarian

cancer model reduces ovarian cancer progression. *Gynecol Oncol, 130*(3), 579-587. doi:10.1016/j.ygyno.2013.05.027

Smyth, E. C., Lagergren, J., Fitzgerald, R. C., Lordick, F., Shah, M. A., Lagergren, P., & Cunningham, D. (2017). Oesophageal cancer. *Nat Rev Dis Primers, 3*, 17048. doi:10.1038/nrdp.2017.48

Song, J. (2007). EMT or apoptosis: a decision for TGF-beta. *Cell Res, 17*(4), 289-290. doi:10.1038/cr.2007.25

Song, K. H., Choi, C. H., Lee, H. J., Oh, S. J., Woo, S. R., Hong, S. O., . . . Kim, T. W. (2017). HDAC1 Upregulation by NANOG Promotes Multidrug Resistance and a Stem-like Phenotype in Immune Edited Tumor Cells. *Cancer Res, 77*(18), 5039-5053. doi:10.1158/0008-5472.Can-17-0072

Spechler, S. J., & Souza, R. F. (2014). Barrett's esophagus. *N Engl J Med, 371*(9), 836-845. doi:10.1056/NEJMra1314704

Spender, L. C., O'Brien, D. I., Simpson, D., Dutt, D., Gregory, C. D., Allday, M. J., . . . Inman, G. J. (2009). TGF-beta induces apoptosis in human B cells by transcriptional regulation of BIK and BCL-XL. *Cell Death Differ, 16*(4), 593-602. doi:10.1038/cdd.2008.183

Sun, X., Martin, R. C. G., Zheng, Q., Farmer, R., Pandit, H., Li, X., . . . Li, Y. (2018). Drug-induced expression of EpCAM contributes to therapy resistance in esophageal adenocarcinoma. *Cell Oncol (Dordr), 41*(6), 651-662. doi:10.1007/s13402-018-0399-z

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell, 126*(4), 663-676. doi:10.1016/j.cell.2006.07.024

Tanaka, M., Kijima, H., Shimada, H., Makuuchi, H., Ozawa, S., & Inokuchi, S. (2015). Expression of podoplanin and vimentin is correlated with prognosis in esophageal squamous cell carcinoma. *Mol Med Rep, 12*(3), 4029-4036. doi:10.3892/mmr.2015.3966

Taube, J. H., Malouf, G. G., Lu, E., Sphyris, N., Vijay, V., Ramachandran, P. P., . . . Mani, S. A. (2013). Epigenetic silencing of microRNA-203 is required for EMT and cancer stem cell properties. *Sci Rep, 3*, 2687. doi:10.1038/srep02687

Thapa, R., & Wilson, G. D. (2016). The Importance of CD44 as a Stem Cell Biomarker and Therapeutic Target in Cancer. *Stem Cells Int, 2016*, 2087204. doi:10.1155/2016/2087204

ThermoFisher. (2019). Using TaqMan Endogenous Control Assays to select an endogenous control. In A. Biosciences (Ed.).

Tian, Y., Luo, A., Cai, Y., Su, Q., Ding, F., Chen, H., & Liu, Z. (2010). MicroRNA-10b promotes migration and invasion through KLF4 in human esophageal cancer cell lines. *J Biol Chem, 285*(11), 7986-7994. doi:10.1074/jbc.M109.062877

Tiwari, N., Gheldof, A., Tatari, M., & Christofori, G. (2012). EMT as the ultimate survival mechanism of cancer cells. *Semin Cancer Biol, 22*(3), 194-207. doi:10.1016/j.semcancer.2012.02.013

Tomita, H., Tanaka, K., Tanaka, T., & Hara, A. (2016). Aldehyde dehydrogenase 1A1 in stem cells and cancer. *Oncotarget, 7*(10), 11018-11032. doi:10.18632/oncotarget.6920

Tomizawa, Y., Wu, T. T., & Wang, K. K. (2012). Epithelial mesenchymal transition and cancer stem cells in esophageal adenocarcinoma originating from Barrett's esophagus. *Oncol Lett, 3*(5), 1059-1063. doi:10.3892/ol.2012.632

Toyokawa, G., Seto, T., Takenoyama, M., & Ichinose, Y. (2015). Insights into brain metastasis in patients with ALK+ lung cancer: is the brain truly a sanctuary? *Cancer Metastasis Rev, 34*(4), 797-805. doi:10.1007/s10555-015-9592-y

Tran, H. D., Luitel, K., Kim, M., Zhang, K., Longmore, G. D., & Tran, D. D. (2014). Transient SNAIL1 expression is necessary for metastatic competence in breast cancer. *Cancer Res, 74*(21), 6330-6340. doi:10.1158/0008-5472.Can-14-0923

Trivers, K. F., Sabatino, S. A., & Stewart, S. L. (2008). Trends in esophageal cancer incidence by histology, United States, 1998-2003. *Int J Cancer, 123*(6), 1422-1428. doi:10.1002/ijc.23691

Tsai, J. H., Donaher, J. L., Murphy, D. A., Chau, S., & Yang, J. (2012). Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer Cell, 22*(6), 725-736. doi:10.1016/j.ccr.2012.09.022

Tse, J. C., & Kalluri, R. (2007). Mechanisms of metastasis: epithelial-to-mesenchymal transition and contribution of tumor microenvironment. *J Cell Biochem, 101*(4), 816-829. doi:10.1002/jcb.21215

Tsujiura, M., Ichikawa, D., Komatsu, S., Shiozaki, A., Takeshita, H., Kosuga, T., . . . Otsuji, E. (2010). Circulating microRNAs in plasma of patients with gastric cancers. *Br J Cancer, 102*(7), 1174-1179. doi:10.1038/sj.bjc.6605608

Ulla, M., Cavadas, D., Munoz, I., Beskow, A., Seehaus, A., & Garcia-Monaco, R. (2010). Esophageal cancer: pneumo-64-MDCT. *Abdom Imaging, 35*(4), 383-389. doi:10.1007/s00261-009-9554-3

Ustaalioglu, B. B. O., Tilki, M., Surmelioglu, A., Bilici, A., Gonen, C., Ustaalioglu, R., . . . Aliustaoglu, M. (2017). The clinicopathologic characteristics and prognostic

factors of gastroesophageal junction tumors according to Siewert classification. *Turk J Surg, 33*(1), 18-24. doi:10.5152/ucd.2017.3379

van der Waals, L. M., Borel Rinkes, I. H. M., & Kranenburg, O. (2018). ALDH1A1 expression is associated with poor differentiation, 'right-sidedness' and poor survival in human colorectal cancer. *PLoS One, 13*(10), e0205536. doi:10.1371/journal.pone.0205536

van Hagen, P., Hulshof, M. C., van Lanschot, J. J., Steyerberg, E. W., van Berge Henegouwen, M. I., Wijnhoven, B. P., . . . van der Gaast, A. (2012). Preoperative chemoradiotherapy for esophageal or junctional cancer. *N Engl J Med, 366*(22), 2074-2084. doi:10.1056/NEJMoa1112088

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol, 3*(7), Research0034. doi:10.1186/gb-2002-3-7-research0034

Vandewoestyne, M., Goossens, K., Burvenich, C., Van Soom, A., Peelman, L., & Deforce, D. (2013). Laser capture microdissection: should an ultraviolet or infrared laser be used? *Anal Biochem, 439*(2), 88-98. doi:10.1016/j.ab.2013.04.023

Vasan, N., Baselga, J., & Hyman, D. M. (2019). A view on drug resistance in cancer. *Nature, 575*(7782), 299-309. doi:10.1038/s41586-019-1730-1

Vermeulen, L., De Sousa, E. M. F., van der Heijden, M., Cameron, K., de Jong, J. H., Borovski, T., . . . Medema, J. P. (2010). Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. *Nat Cell Biol, 12*(5), 468-476. doi:10.1038/ncb2048

Visvader, J. E., & Lindeman, G. J. (2008). Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. *Nat Rev Cancer, 8*(10), 755-768. doi:10.1038/nrc2499

Visvader, J. E., & Lindeman, G. J. (2012). Cancer stem cells: current status and evolving complexities. *Cell Stem Cell, 10*(6), 717-728. doi:10.1016/j.stem.2012.05.007

Vogel, C., & Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet, 13*(4), 227-232. doi:10.1038/nrg3185

Wakamatsu, Y., Sakamoto, N., Oo, H. Z., Naito, Y., Uraoka, N., Anami, K., . . . Yasui, W. (2012). Expression of cancer stem cell markers ALDH1, CD44 and CD133 in primary tumor and lymph node metastasis of gastric cancer. *Pathol Int, 62*(2), 112-119. doi:10.1111/j.1440-1827.2011.02760.x

Wang, J. Y., Wu, T., Ma, W., Li, S., Jing, W. J., Ma, J., . . . Nan, K. J. (2018). Expression and clinical significance of autophagic protein LC3B and EMT markers in gastric cancer. *Cancer Manag Res, 10*, 1479-1486. doi:10.2147/cmar.S164842

Wang, T., Hou, J., Li, Z., Zheng, Z., Wei, J., Song, D., . . . Cai, J. C. (2017). miR-15a-3p and miR-16-1-3p Negatively Regulate Twist1 to Repress Gastric Cancer Cell Invasion and Metastasis. *Int J Biol Sci, 13*(1), 122-134. doi:10.7150/ijbs.14770

Wang, X., Yang, S., Zhao, X., Guo, H., Ling, X., Wang, L., . . . Zhou, S. (2014). OCT3 and SOX2 promote the transformation of Barrett's esophagus to adenocarcinoma by regulating the formation of tumor stem cells. *Oncol Rep, 31*(4), 1745-1753. doi:10.3892/or.2014.3003

Wang, X. F., Zhang, X. W., Hua, R. X., Du, Y. Q., Huang, M. Z., Liu, Y., . . . Guo, W. J. (2016). Mel-18 negatively regulates stem cell-like properties through downregulation of miR-21 in gastric cancer. *Oncotarget, 7*(39), 63352-63361. doi:10.18632/oncotarget.11221

Wang, Y., Zhao, Y., Herbst, A., Kalinski, T., Qin, J., Wang, X., . . . Bruns, C. J. (2016). miR-221 Mediates Chemoresistance of Esophageal Adenocarcinoma by Direct Targeting of DKK2 Expression. *Ann Surg, 264*(5), 804-814. doi:10.1097/sla.0000000000001928

Watanabe, T., Okumura, T., Hirano, K., Yamaguchi, T., Sekine, S., Nagata, T., & Tsukada, K. (2017). Circulating tumor cells expressing cancer stem cell marker CD44 as a diagnostic biomarker in patients with gastric cancer. *Oncol Lett, 13*(1), 281-288. doi:10.3892/ol.2016.5432

Wei, S. C., Fattet, L., Tsai, J. H., Guo, Y., Pai, V. H., Majeski, H. E., . . . Yang, J. (2015). Matrix stiffness drives epithelial-mesenchymal transition and tumour metastasis through a TWIST1-G3BP2 mechanotransduction pathway. *Nat Cell Biol, 17*(5), 678-688. doi:10.1038/ncb3157

Wen, L., Chen, X. Z., Yang, K., Chen, Z. X., Zhang, B., Chen, J. P., . . . Hu, J. K. (2013). Prognostic value of cancer stem cell marker CD133 expression in gastric cancer: a systematic review. *PLoS One, 8*(3), e59154. doi:10.1371/journal.pone.0059154

Wenqi, D., Li, W., Shanshan, C., Bei, C., Yafei, Z., Feihu, B., . . . Daiming, F. (2009). EpCAM is overexpressed in gastric cancer and its downregulation suppresses proliferation of gastric cancer. *J Cancer Res Clin Oncol, 135*(9), 1277-1285. doi:10.1007/s00432-009-0569-5

Wu, S., Du, Y., Beckford, J., & Alachkar, H. (2018). Upregulation of the EMT marker vimentin is associated with poor clinical outcome in acute myeloid leukemia. *J Transl Med, 16*(1), 170. doi:10.1186/s12967-018-1539-y

Wu, W. L., Wang, W. Y., Yao, W. Q., & Li, G. D. (2015). Suppressive effects of microRNA-16 on the proliferation, invasion and metastasis of hepatocellular carcinoma cells. *Int J Mol Med, 36*(6), 1713-1719. doi:10.3892/ijmm.2015.2379

Wu, Y., & Wu, P. Y. (2009). CD133 as a marker for cancer stem cells: progresses and concerns. *Stem Cells Dev, 18*(8), 1127-1134. doi:10.1089/scd.2008.0338

Wu, Y. H., Chiu, W. T., Young, M. J., Chang, T. H., Huang, Y. F., & Chou, C. Y. (2015). Solanum Incanum Extract Downregulates Aldehyde Dehydrogenase 1-Mediated Stemness and Inhibits Tumor Formation in Ovarian Cancer Cells. *J Cancer, 6*(10), 1011-1019. doi:10.7150/jca.12738

Xi, X. P., Zhuang, J., Teng, M. J., Xia, L. J., Yang, M. Y., Liu, Q. G., & Chen, J. B. (2016). MicroRNA-17 induces epithelial-mesenchymal transition consistent with the cancer stem cell phenotype by regulating CYP7B1 expression in colon cancer. *Int J Mol Med, 38*(2), 499-506. doi:10.3892/ijmm.2016.2624

Xia, P., Song, C. L., Liu, J. F., Wang, D., & Xu, X. Y. (2015). Prognostic value of circulating CD133(+) cells in patients with gastric cancer. *Cell Prolif, 48*(3), 311-317. doi:10.1111/cpr.12175

Xie, G., Ji, A., Yuan, Q., Jin, Z., Yuan, Y., Ren, C., . . . Chen, L. (2014). Tumour-initiating capacity is independent of epithelial-mesenchymal transition status in breast cancer cell lines. *Br J Cancer, 110*(10), 2514-2523. doi:10.1038/bjc.2014.153

Xing, Y., Luo, D. Y., Long, M. Y., Zeng, S. L., & Li, H. H. (2014). High ALDH1A1 expression correlates with poor survival in papillary thyroid carcinoma. *World J Surg Oncol, 12*, 29. doi:10.1186/1477-7819-12-29

Yang, W., Wang, Y., Wang, W., Chen, Z., & Bai, G. (2018). Expression of Aldehyde Dehydrogenase 1A1 (ALDH1A1) as a Prognostic Biomarker in Colorectal Cancer Using Immunohistochemistry. *Med Sci Monit, 24*, 2864-2872. doi:10.12659/msm.910109

Yao, D., Dai, C., & Peng, S. (2011). Mechanism of the mesenchymal-epithelial transition and its relationship with metastatic tumor formation. *Mol Cancer Res, 9*(12), 1608-1620. doi:10.1158/1541-7786.Mcr-10-0568

Yashiro, M., Sunami, T., & Hirakawa, K. (2005). CD54 expression is predictive for lymphatic spread in human gastric carcinoma. *Dig Dis Sci, 50*(12), 2224-2230. doi:10.1007/s10620-005-3039-1

Ychou, M., Boige, V., Pignon, J. P., Conroy, T., Bouche, O., Lebreton, G., . . . Rougier, P. (2011). Perioperative chemotherapy compared with surgery alone for resectable gastroesophageal adenocarcinoma: an FNCLCC and FFCD multicenter phase III trial. *J Clin Oncol, 29*(13), 1715-1721. doi:10.1200/jco.2010.33.0597

Ye, J., Wang, Z., Zhao, J., Chen, W., Wu, D., Wu, P., & Huang, J. (2017). MicroRNA-141 inhibits tumor growth and minimizes therapy resistance in colorectal cancer. *Mol Med Rep, 15*(3), 1037-1042. doi:10.3892/mmr.2017.6135

Ye, Y., Zhang, S., Chen, Y., Wang, X., & Wang, P. (2018). High ALDH1A1 expression indicates a poor prognosis in gastric neuroendocrine carcinoma. *Pathol Res Pract, 214*(2), 268-272. doi:10.1016/j.prp.2017.10.015

Yoon, C., Park, D. J., Schmidt, B., Thomas, N. J., Lee, H. J., Kim, T. S., . . . Yoon, S. S. (2014). CD44 expression denotes a subpopulation of gastric cancer cells in which Hedgehog signaling promotes chemotherapy resistance. *Clin Cancer Res, 20*(15), 3974-3988. doi:10.1158/1078-0432.Ccr-14-0011

Yoshida, A., Dave, V., Han, H., & Scanlon, K. J. (1993). Enhanced transcription of the cytosolic ALDH gene in cyclophosphamide resistant human carcinoma cells. *Adv Exp Med Biol, 328*, 63-72. doi:10.1007/978-1-4615-2904-0_8

Yu, J. M., Sun, W., Hua, F., Xie, J., Lin, H., Zhou, D. D., & Hu, Z. W. (2015). BCL6 induces EMT by promoting the ZEB1-mediated transcription repression of E-cadherin in breast cancer cells. *Cancer Lett, 365*(2), 190-200. doi:10.1016/j.canlet.2015.05.029

Yu, M., Bardia, A., Wittner, B. S., Stott, S. L., Smas, M. E., Ting, D. T., . . . Maheswaran, S. (2013). Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science, 339*(6119), 580-584. doi:10.1126/science.1228522

Zanoni, A., Verlato, G., Baiocchi, G. L., Casella, F., Cossu, A., d'Ignazio, A., . . . Giacopuzzi, S. (2018). Siewert III esophagogastric junction adenocarcinoma: does TNM 8th save us? *Updates Surg, 70*(2), 241-249. doi:10.1007/s13304-018-0537-1

Zavadil, J., Bitzer, M., Liang, D., Yang, Y. C., Massimi, A., Kneitz, S., . . . Bottinger, E. P. (2001). Genetic programs of epithelial cell plasticity directed by transforming growth factor-beta. *Proc Natl Acad Sci U S A, 98*(12), 6686-6691. doi:10.1073/pnas.111614398

Zavros, Y. (2017). Initiation and Maintenance of Gastric Cancer: A Focus on CD44 Variant Isoforms and Cancer Stem Cells. *Cell Mol Gastroenterol Hepatol, 4*(1), 55-63. doi:10.1016/j.jcmgh.2017.03.003

Zeineddine, D., Hammoud, A. A., Mortada, M., & Boeuf, H. (2014). The Oct4 protein: more than a magic stemness marker. *Am J Stem Cells, 3*(2), 74-82.

Zeppernick, F., Ahmadi, R., Campos, B., Dictus, C., Helmke, B. M., Becker, N., . . . Herold-Mende, C. C. (2008). Stem cell marker CD133 affects clinical outcome in glioma patients. *Clin Cancer Res, 14*(1), 123-129. doi:10.1158/1078-0432.Ccr-07-0932

Zhang, G. J., Zhou, H., Xiao, H. X., Li, Y., & Zhou, T. (2013). Up-regulation of miR-224 promotes cancer cell proliferation and invasion and predicts relapse of colorectal cancer. *Cancer Cell Int, 13*(1), 104. doi:10.1186/1475-2867-13-104

Zhang, H., & Li, Z. (2019). microRNA-16 Via Twist1 Inhibits EMT Induced by PM2.5 Exposure in Human Hepatocellular Carcinoma. *Open Med (Wars), 14*, 673-682. doi:10.1515/med-2019-0078

Zhang, L., Guo, X., Zhang, D., Fan, Y., Qin, L., Dong, S., & Zhang, L. (2017). Upregulated miR-132 in Lgr5(+) gastric cancer stem cell-like cells contributes to cisplatin-resistance via SIRT1/CREB/ABCG2 signaling pathway. *Mol Carcinog, 56*(9), 2022-2034. doi:10.1002/mc.22656

Zhang, Y., Li, C. F., Ma, L. J., Ding, M., & Zhang, B. (2016). MicroRNA-224 aggrevates tumor growth and progression by targeting mTOR in gastric cancer. *Int J Oncol, 49*(3), 1068-1080. doi:10.3892/ijo.2016.3581

Zhao, D., Klempner, S. J., & Chao, J. (2019). Progress and challenges in HER2-positive gastroesophageal adenocarcinoma. *J Hematol Oncol, 12*(1), 50. doi:10.1186/s13045-019-0737-2

Zhao, L., Liu, J., Chen, S., Fang, C., Zhang, X., & Luo, Z. (2018). Prognostic significance of NANOG expression in solid tumors: a meta-analysis. *Onco Targets Ther, 11*, 5515-5526. doi:10.2147/ott.S169593

Zhong, Z. Y., Shi, B. J., Zhou, H., & Wang, W. B. (2018). CD133 expression and MYCN amplification induce chemoresistance and reduce average survival time in pediatric neuroblastoma. *J Int Med Res, 46*(3), 1209-1220. doi:10.1177/0300060517732256

Zhou, Q. Y., Peng, P. L., & Xu, Y. H. (2019). MiR-221 affects proliferation and apoptosis of gastric cancer cells through targeting SOCS3. *Eur Rev Med Pharmacol Sci, 23*(21), 9427-9435. doi:10.26355/eurrev_201911_19436

Appendix One: Fold Change Data for Paired Tumour Samples

Appendix Table 1.1 Directions of Change in Expression of 5 Genes in Paired Tumour Samples.

	NANOG	POU5F1	Vimentin	CDH1	Serpine1
1	↑	↓	↑	0	NA
2	↓	↓	↑	0	NA
3	↓	↓	0	0	NA
4	↓	↓	↓	0	NA
5	↑	↑	0	0	↓
6	↓	↓	0	0	0
7	↑	↑	0	↓	0
8	↑	↑	0	0	0
9	↓	↓	0	0	0
10	↓	↓	↑	0	↑
11	0	↓	0	0	0
12	↑	↑	↓	↓	0
13	↑	↑	↓	↓	0
14	↑	↑	0	0	0
15	↓	↓	0	0	0
16	↑	↓	↑	↓	0
17	↑	↑	0	0	0
18	↑	↑	0	0	0
19	↑	↓	↓	0	0
20	↓	↑	0	0	0
21	↓	↓	0	0	0
22	↓	↑	0	0	0
23	↓	↓	0	0	0
24	↓	↑	0	0	0
25	↑	↑	0	0	0
26	↑	↑	↓	0	0
27	↑	↓	0	0	0
28	↓	↑	0	0	0
29	↓	↑	0	0	0
30	0	↓	0	0	0
31	↑	↓	↓	0	0
32	↓	↑	0	0	0
33	↓	↑	0	0	0
34	↑	↓	0	↑	0
35	↑	↑	↓	↓	0
36	↑	↑	0	0	0
37	↑	↑	↓	↓	0
38	↓	↓	↓	0	0
39	↑	↑	↑	↓	0
40	↑	↑	0	0	0
41	↑	↓	↑	0	0
42	↑	↑	↑	0	0
43	NA	NA	NA	NA	NA
44	↓	↓	0	0	0
45	↓	↑	0	0	0
46	↑	↓	0	0	0
47	↑	↑	0	0	0
48	↓	↑	↓	0	0
49	↓	↓	↓	0	0
50	↑	↑	↓	0	0
51	↓	↑	↑	0	0
52	↑	↓	0	0	0
53	↑	↑	↓	0	0
54	↓	↓	0	0	0
55	↑	↑	0	0	0
56	↑	↑	0	0	0
57	NA	NA	NA	NA	NA
58	NA	NA	NA	NA	NA

Appendix Table 1.2 Patterns of Fold Change in Expression of 5 Genes in Paired Tumour Samples.

Fold Change	NANOG	POU5F1	VIM	CDH1	SERPIN
1	0.635578	-6.14084	NA	0	NA
2	0.315849	-0.2488	NA	0	NA
3	-2.17788	4.277314	0	0	NA
4	7.934712	13.54419	-267.063	0	NA
5	2.266959	1.10533	0	0	NA
6	0	1.032724	0	0	0
7	16.28798	-1.85929	0	NA	0
8	NA	0.512681	0	0	0
9	0.171035	-0.16899	0	0	0
10	NA	0.542256	6.67693	0	NA
11	0	-0.81286	0	0	0
12	-0.3709	-0.91603	NA	NA	0
13	3.125757	0.976166	NA	NA	0
14	-3.88207	0.066168	0	0	0
15	NA	-0.9692	0	0	0
16	-2.62772	0.38434	NA	0.726109	0
17	4.684521	-0.99548	0	0	0
18	NA	NA	0	0	0
19	1.34974	-1.17354	NA	0	0
20	2.452959	0.232825	0	0	0
21	1.254187	-0.28725	0	0	0
22	12.72344	3.436246	0	0	0
23	-0.30968	2.73825	0	0	0
24	-0.25486	0.71189	0	0	0
25	-0.49094	0.120471	0	0	0
26	-1.58225	0.197693	NA	0	0
27	0.735062	12.3061	0	0	0
28	1.273408	0.783474	0	0	0
29	NA	0.709273	0	0	0
30	0	-7.16857	0	0	0
31	-0.76542	2.645437		0	0
32	NA	-0.42627	0	0	0
33	1.14588	0.839632	0	0	0
34	0.651368	1.514028	0	NA	0
35	0.829427	0.589938	NA	NA	0
36	-0.6919	0.089202	0	0	0
37	-0.06519	0.231518	NA	NA	0
38	1.521038	-31.2297	NA	0	0
39	-2.03388	-0.07437	0	NA	0
40	-0.19321	0.45659	0	0	0

Appendix Table 1.2 Patterns of Fold Change in Expression of 5 Genes in Paired Tumour Samples.

41	**0.880112**	**1.250526**	**0.539014**	**0**	**0**
42	0.867177	0.542148	2.207717	0	0
43	NA	NA	NA	NA	NA
44	-1.29228	-2.92816	0	0	0
45	3.92292	-0.1168	0	0	0
46	0.903058	1.442796	0	0	0
47	NA	0	0	0	0
48	-0.85926	0.48406	NA	0	0
49	-2.49181	167.8793	NA	0	0
50	-0.95095	-1.96154	NA	0	0
51	1.11198	0.246568	NA	0	0
52	NA	0.147434	0	0	0
53	0.247503	0.254485	NA	0	0
54	-1.68928	2.409591	0	0	0
55	0.334525	0.286401	0	0	0
56	NA	0.944754	0	0	0
57	NA	NA	NA	NA	NA
58	NA	NA	NA	NA	NA

Appendix Table 1.3 Directions of Change in Expression of 9 miRNAs in Paired Tumour Samples.

	miR16	miR17	miR21	miR221	miR10b	miR223	miR200a	miR141	miR224
1	0	0	↑	↑	0	↑	↑	↑	0
2	0	↓	↑	0	0	0	↑	↑	0
3	↑	0	↑	↑	0	0	↑	↑	↑
4	0	0	↑	0	0	0	↑	0	↑
5	0	0	0	0	0	0	0	↑	0
6	↑	0	↑	↑	↑	0	↔	↑	↑
7	↓	0	↓	↑	↓	0	↓	↓	↑
8	0	0	↑	↑	0	0	↓	↑	↑
9	0	0	0	0	0	0	0	↑	0
10	↓	↓	↓	0	0	0	↓	↓	↓
11	0	↓	↓	↑	0	0	↑	↓	↓
12	0	0	↓	↓	0	↓	↓	↓	0
13	↑	↑	↓	↑	↑	↑	↑	↓	↓
14	↓	0	↓	↓	0	0	↓	↓	0
15	0	0	0	0	0	0	0	0	0
16	0	↑	↑	↑	0	0	↑	↑	↑
17	↓	0	↑	↔	0	0	↓	0	↓
18	0	0	↑	↑	↑	0	↑	↑	↑
19	0	0	↑	↑	0	↑	↑	↑	↑
20	0	0	↓	↓	0	0	↑	↑	↑
21	↓	0	↑	↑	0	↑	↓	↓	↓
22	↓	↓	↑	↑	↓	0	↑	NA	↑
23	↑	0	↓	↑	↓	0	↓	↓	↑
24	↑	↑	↓	↑	0	0	↓	↑	↓
25	↑	↑	0	↑	0	↓	↓	↓	↓
26	0	0	↓	↓	0	0	↓	0	↓
27	0	0	0	0	0	0	0	0	0
28	↓	0	↔	0	0	0	↓	0	0
29	0	0	↑	↑	0	0	↓	↑	↑
30	↑	↑	↑	↑	0	0	↑	↓	↓
31	0	↑	↓	↑	0	0	↑	↓	↓
32	↓	0	↓	0	↓	0	↑	0	↓
33	↑	↑	↓	0	0	0	↓	↓	↑
34	↑	↑	↓	↑	0	0	↓	↓	↑
35	↓	0	↑	↓	↓	0	↓	↓	↓
36	↓	↓	↑	↓	0	0	↑	↑	↓
37	↓	↓	↓	↓	0	0	↓	↑	↓
38	↓	0	↓	0	0	0	↓	0	↑
39	0	0	0	0	0	0	0	↓	0
40	0	0	↑	↑	0	0	↑	0	0
41	0	↑	↑	↓	0	0	↑	↑	↑
42	↓	↓	↓	↓	0	0	↑	↑	↓
43	↑	0	↑	↑	↑	0	↑	↑	↑
44	0	0	↓	0	0	0	0	↓	0
45	0	0	↓	↑	0	0	0	↑	0
46	0	0	0	↓	0	0	0	0	0
47	0	0	↑	0	0	0	↑	0	0
48	0	0	0	0	0	0	↑	0	0
49	↓	↓	↓	↓	0	↓	↑	↓	↑
50	0	↓	↓	0	0	0	↓	↓	↓
51	0	0	↑	0	0	0	0	↑	0
52	0	0	↑	↑	0	0	↑	0	0
53	0	0	↑	0	0	↓	↓	↑	↑
54	↑	↑	↓	↑	↑	0	↑	↓	↑
55	↓	↓	↓	↓	0	0	↓	0	↓
56	0	0	0	0	NA	0	0	NA	0
57	NA	NA	NA	NA	NA	NA	NA	NA	NA
58	NA	NA	NA	NA	NA	NA	NA	NA	NA

Appendix Table 1.4 Patterns of Fold Change in Expression of 9 miRNAs in Paired Tumour Samples.

Fold Change	miR16	miR17	miR21	miR221	miR10b	miR223	miR200a	miR141	miR224
1	0	0	NA	NA	0	NA	NA	NA	0
2	0	NA	NA	0	0	0	NA	0.878662	0
3	NA	0	1.696205	NA	0	0	NA	0.317666	NA
4	0	0	-2.43651	0	0	0	NA	0	NA
5	0	0	0	0	0	0	0	NA	0
6	NA	0	NA	NA	NA	0	0	NA	NA
7	0.138499	0	0.094396	NA	NA	0	8.453346	4.464214	NA
8	0	0	-3.69735	NA	0	0	NA	0.415727	NA
9	0	0	0	0	0	0	0	NA	0
10	149.3175	NA	-0.3418	0	0	0	NA	-0.633	NA
11	0	NA	0.44037	-17.3488	0	0	NA	-1.75024	NA
12	0	0	0.860041	NA	0	NA	NA	NA	0
13	0	0	NA	0	0	0	0	NA	NA
14	NA	0	-0.31127	NA	0	0	0	NA	0
15	0	0	0	0	0	0	0	0	0
16	0	0	93.33224	NA	0	0	NA	NA	NA
17	NA	0	1.427751	1	0	0	-2.81473	0	NA
18	0	0	NA	NA	NA	0	NA	NA	NA
19	0	0	3.546538	NA	0	NA	NA	NA	NA
20	0	0	0.92371	NA	0	0	-26.6101	0.168144	NA
21	-1.68418	0	8.010903	NA	0	NA	NA	1.281703	-1.18511
22	-0.95346	0.07529	-2.87828	-1.15079	NA	0	0.043755	NA	NA
23	NA	0	0.431799	NA	NA	0	NA	-4.692	NA
24	NA	NA	0.815345	NA	0	0	NA	NA	NA
25	NA	-19.0848	0	NA	0	NA	NA	-0.86034	NA
26	0	0	0.573361	NA	0	0	NA	0	NA
27	0	0	0	0	0	0	0	0	0
28	NA	0	1	0	0	0	NA	0	0
29	0	0	-19.7456	2.482868	0	0	0	NA	NA
30	NA	NA	0.715607	NA	0	0	-6.58883	-1.69625	NA
31	0	NA	-0.51283	NA	0	0	1.417998	1.055718	3.188187
32	NA	0	NA	0	NA	0	4.555576	0	1.55357
33	NA	NA	0.146794	0	0	0	NA	1.733775	NA
34	NA	NA	0.184398	NA	0	0	1.191231	-0.46881	-4.16372
35	NA	0	2.266563	NA	NA	0	-4.61839	2.747833	-3.87273
36	NA	NA	12.92308	NA	0	0	1.900501	NA	NA
37	NA	NA	NA	NA	0	0	4.116958	-0.38699	NA
38	NA	0	NA	0	0	0	NA	0	NA
39	0	0	0	0	0	0	0	NA	0
40	0	0	-27.0677	NA	0	0	NA	0	0

Appendix Table 1.4 Patterns of Fold Change in Expression of 9 miRNAs in Paired Tumour Samples.

41	**0**	**NA**	**-2.63807**	**NA**	**0**	**0**	**7.698302**	**-0.76784**	**NA**
42	-0.92736	NA	0.978215	NA	0	0	NA	0.309645	NA
43	NA	0	NA	NA	NA	0	NA	NA	NA
44	0	0	NA	0	0	0	0	NA	0
45	0	0	NA	NA	0	0	0	NA	0
46	0	0	0	NA	0	0	0	0	0
47	0	0	NA	0	0	0	NA	0	0
48	0	0	0	0	0	0	0	0	0
49	NA	NA	NA	NA	0	NA	-30.1251	NA	NA
50	0	NA	NA	0	0	0	NA	NA	NA
51	0	0	NA	0	0	0	0	NA	0
52	0	0	NA	NA	0	0	NA	0	0
53	0	0	2.59192	0	0	NA	0.286626	NA	NA
54	NA	NA	0.353013	NA	NA	0	1.210505	-10.3418	NA
55	NA	NA	-0.11487	NA	0	0	NA	0	NA
56	0	0	0	0	NA	0	0	NA	0
57	NA	NA	NA	NA	NA	NA	NA	NA	NA
58	NA	NA	NA	NA	NA	NA	NA	NA	NA

Appendix Table 1.5 Directions of Change in Expression of 7 Proteins in Paired Tumour Samples.

	EpCAM	Oct-4	CD34	E-cadherin	Vimentin	ALDH1A1 Total	CD133 %	CD133 Topology
1	↓	0	0	↔	0	↔	↓	↓
2	↓	↓	0	↔	0	↓	0	0
3	↓	0	0	↔	0	↓	0	0
4	NA	NA	NA	↔	0	↔	↑	↔
5	↓	↓	0	↔	0	↓	↑	↑
6	↓	0	0	↔	0	↓	0	0
7	↔	↔	0	↔	0	↓	↑	↑
8	↔	0	0	↔	0	↔	↔	↔
9	↑	↔	0	↔	0	↑	↓	↓
10	↔	0	0	↔	0	↑	↓	↓
11	↔	0	0	↔	0	↓	0	0
12	↔	0	0	↔	0	↔	0	0
13	↔	0	0	↔	0	↓	↑	↑
14	↓	↓	0	↔	0	↔	↔	↔
15	NA	NA	NA	NA	NA	↑	↑	↑
16	NA	0	0	↔	NA	↑	↔	↔
17	↔	↓	0	↔	0	↓	↔	↔
18	↔	0	0	↔	0	↔	↑	↑
19	NA	NA	NA	NA	NA	NA	NA	NA
20	↓	0	0	↔	0	↓	↔	↔
21	↔	0	0	↔	0	↔	↑	↔
22	↓	0	0	↔	0	↓	↔	↑
23	↓	0	0	↔	0	↓	0	0
24	NA	NA	NA	NA	NA	NA	NA	NA
25	NA	0	0	NA	NA	↓	0	0
26	NA	NA	NA	NA	NA	NA	NA	NA
27	↓	0	0	↔	0	↔	↑	↑
28	↔	0	0	↔	0	↔	0	0
29	NA	↓	0	↔	NA	↓	↑	↑
30	↓	0	0	↔	0	↑	↓	↔
31	↓	0	0	↔	0	↑	↔	↔
32	↓	↑	0	↓	↑	↑	0	0
33	↓	0	0	↔	0	↑	0	0
34	↔	↔	0	↔	0	0	0	0
35	↓	0	0	↔	0	↔	↔	↔
36	↔	↔	0	↔	0	NA	↓	↓
37	↓	0	0	↔	0	↑	↓	↓
38	↓	0	0	↓	0	↓	↑	↑
39	NA	NA	NA	NA	NA	NA	NA	NA
40	↓	0	0	↔	0	↔	↔	↔
41	↔	↔	0	↔	0	↑	↓	↓
42	↓	0	0	↔	0	↑	↑	↑
43	NA	NA	NA	NA	NA	NA	NA	NA
44	↓	0	0	↔	0	↔	0	0
45	↔	0	0	↔	0	↑	↓	↓
46	↓	↓	0	↔	↑	↔	↓	↓
47	↔	↔	0	↔	NA	↑	↓	↔
48	NA	NA	NA	NA	NA	NA	NA	NA
49	NA	NA	NA	NA	NA	NA	NA	NA
50	↔	↑	0	↔	0	↑	↓	↓
51	↓	↓	0	↔	↑	↑	↓	↓
52	↑	0	0	↔	0	↓	↓	↓
53	↓	↔	0	↔	0	NA	NA	NA
54	NA	↑	0	↔	0	↔	↓	↓
55	↓	↓	0	↔	0	↑	↓	↓
56	↓	0	0	↔	0	↔	↔	↔
57	NA	0	0	↔	0	↓	↑	↑
58	↓	0	0	↔	0	NA	↓	↓

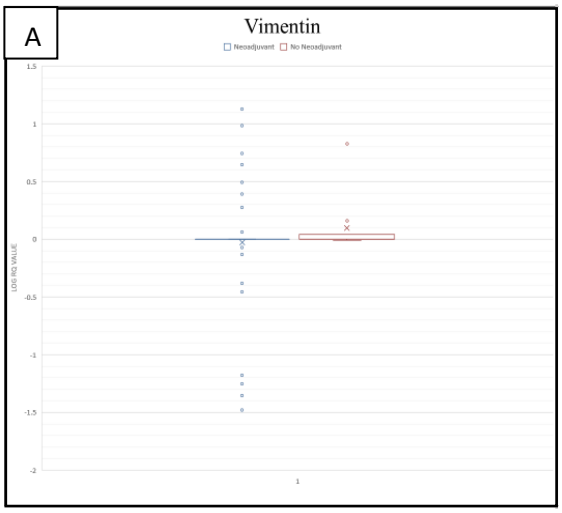
Appendix Table 1.6 Patterns of Fold Change in Expression of 7 Proteins in Paired Tumour Samples.

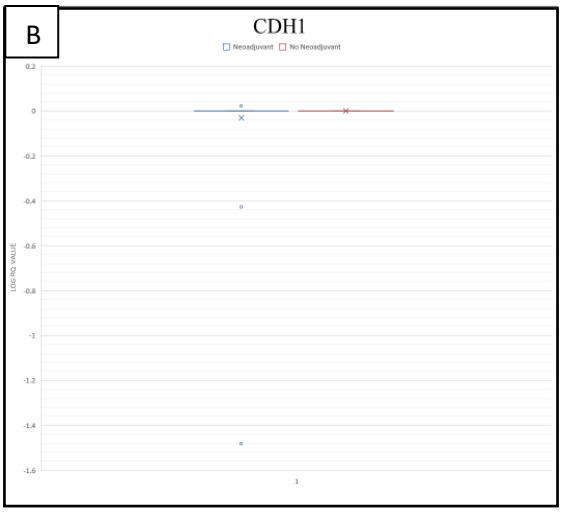
Fold Change	EpCAM	Oct-4	E-cadherin	Vimentin	ALDH1A1 Total	CD133 %	CD133 Topology
1	1.5	0	1	0	1	NA	NA
2	1.5	NA	1	0	NA	0	0
3	2	0	1	0	1.666667	0	0
4	NA	NA	1	0	1	0.5	1
5	2	NA	1	0	3	0	0
6	1.5	0	1	0	1.25	0	0
7	1	1	1	0	1.2	0	0
8	1	0	1	0	1	1	1
9	0.5	1	1	0	0.8	NA	NA
10	1	0	1	0	0.833333	NA	NA
11	1	0	1	0	1.25	0	0
12	1	0	1	0	1	0	0
13	1	0	1	0	1.333333	0	0
14	1.5	NA	1	0	1	1	1
15	NA	NA	NA	NA	0.666667	0	0
16	NA	0	1	NA	0.5	0	0
17	1	NA	1	0	1.2	1	1
18	1	0	1	0	1	0	0
19	NA	NA	NA	NA	NA	NA	NA
20	1.5	0	1	0	NA	1	1
21	1	0	1	0	1	0.75	1
22	2	0	1	0	NA	1	0.5
23	2	0	1	0	1.2	0	0
24	NA	NA	NA	NA	NA	NA	NA
25	NA	0	NA	NA	NA	0	0
26	NA	NA	NA	NA	NA	NA	NA
27	1.5	0	1	0	1	0	0
28	1	0	1	0	1	0	0
29	NA	NA	1	NA	1.25	0.666667	0.5
30	1.5	0	1	0	0.8	2	1
31	2	0	1	0	0.8	1	1
32	NA	NA	NA	NA	0.857143	0	0
33	1.5	0	1	0	0.6	0	0
34	1	1	1	0	0	0	0
35	2	0	1	0	1	0	0
36	1	1	1	0	NA	NA	NA
37	2	0	1	0	0.833333	NA	NA
38	2	0	NA	0	2	0	0
39	NA	NA	NA	NA	NA	NA	NA
40	3	0	1	0	1	1	1

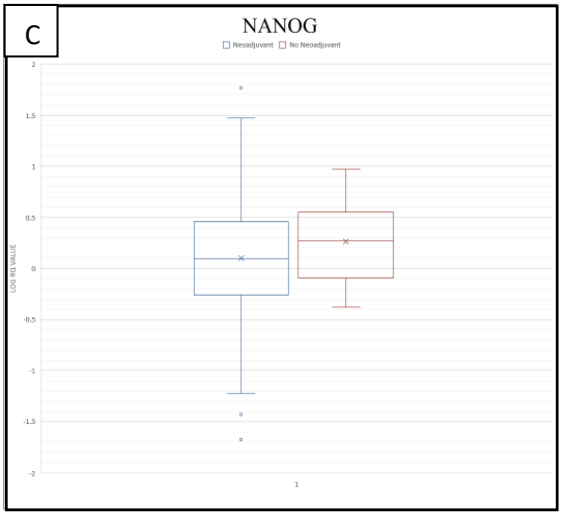
Appendix Table 1.6 Patterns of Fold Change in Expression of 7 Proteins in Paired Tumour Samples.

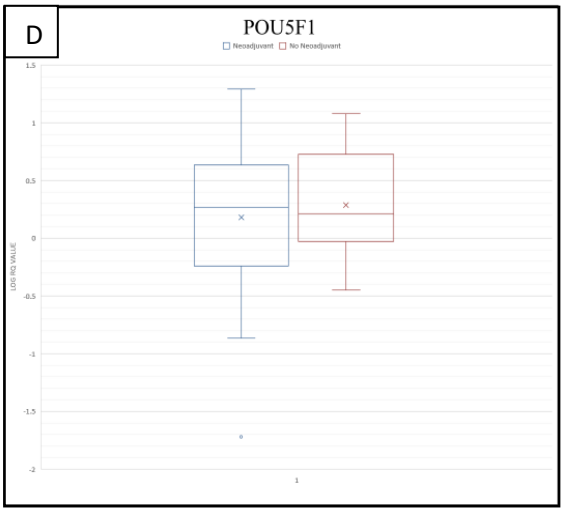
41	**1**	**1**	**1**	**0**	**NA**	**NA**	**NA**
42	1.5	0	1	0	NA	0	0
43	NA	NA	NA	NA	NA	NA	NA
44	2	0	1	0	1	0	0
45	1	0	1	0	0.8	NA	NA
46	1.5	NA	1	NA	1	NA	NA
47	1	1	1	NA	0.666667	NA	1
48	NA	NA	NA	NA	NA	NA	NA
49	NA	NA	NA	NA	NA	NA	NA
50	1	NA	1	0	0.75	NA	NA
51	2	NA	1	NA	0.5	2.5	2
52	0.5	0	1	0	NA	NA	NA
53	1	1	1	0	NA	NA	NA
54	NA	NA	1	0	1	5	2
55	1.5	NA	1	0	0.5	NA	NA
56	2	0	1	0	1	0	0
57	NA	0	1	0	1.666667	0	0
58	2	0	1	0	NA	NA	NA

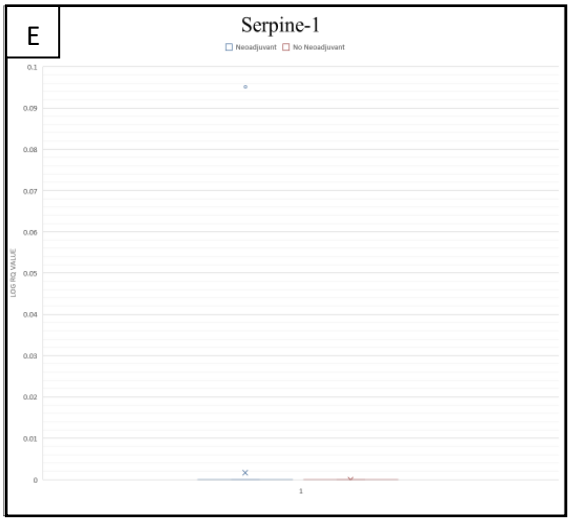
Appendix Two: Comparison of Expression Data in Patients with and Without Neoadjuvant Treatment

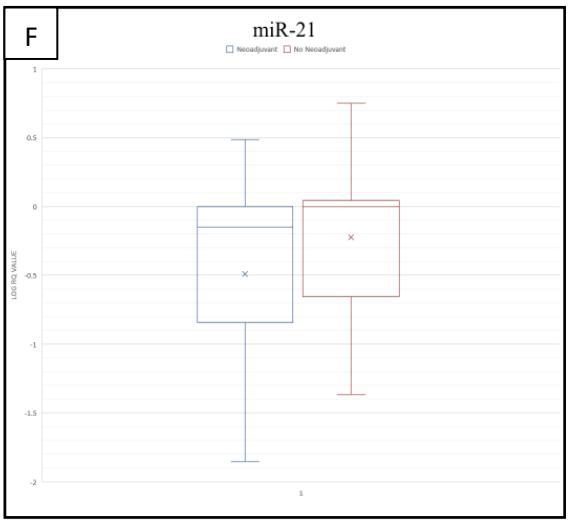

A
Vimentin
Neoadjuvant
No Neoadjuvant
LOG RQ VALUE
1.5
1
0.5
0
-0.5
-1
-1.5
-2
1

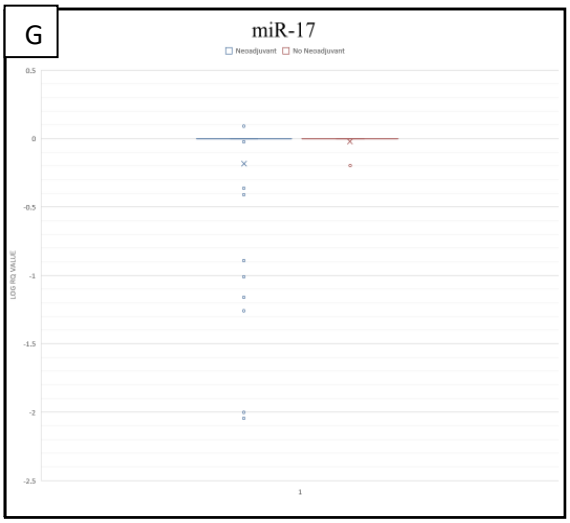

B
CDH1
Neoadjuvant
No Neoadjuvant
LOG RQ VALUE
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
-1.6
1

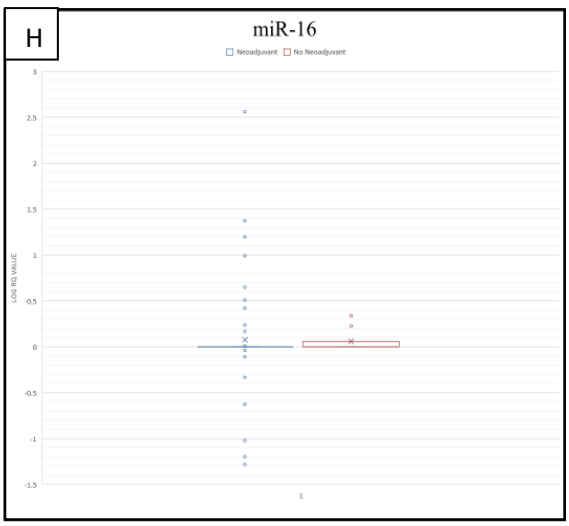

C
NANOG
Neoadjuvant
No Neoadjuvant
LOG RQ VALUE
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
1

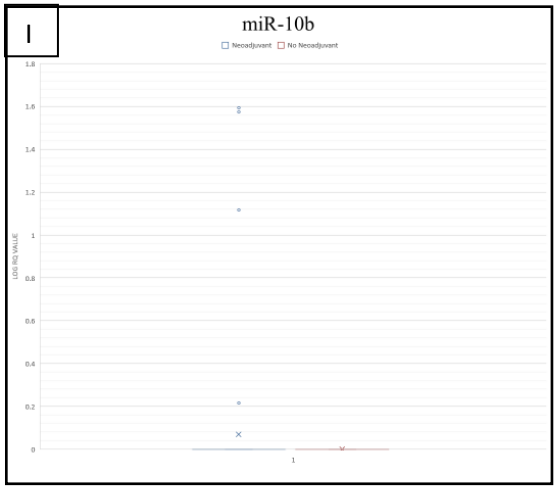

D
POU5F1
Neoadjuvant
No Neoadjuvant
LOG RQ VALUE
1.5
1
0.5
0
-0.5
-1
-1.5
-2
1

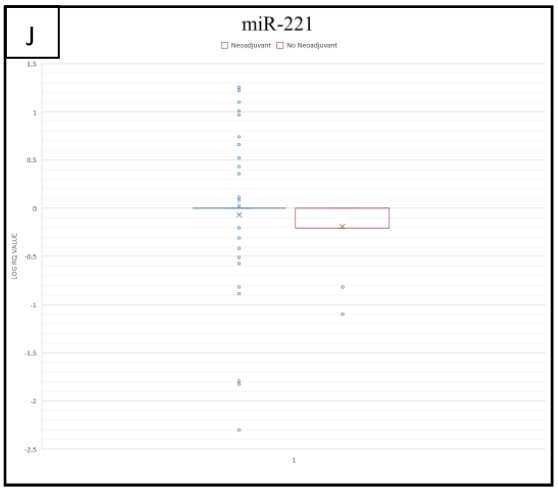

E
Serpine-1
Neoadjuvant
No Neoadjuvant
LOG RQ VALUE
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
1

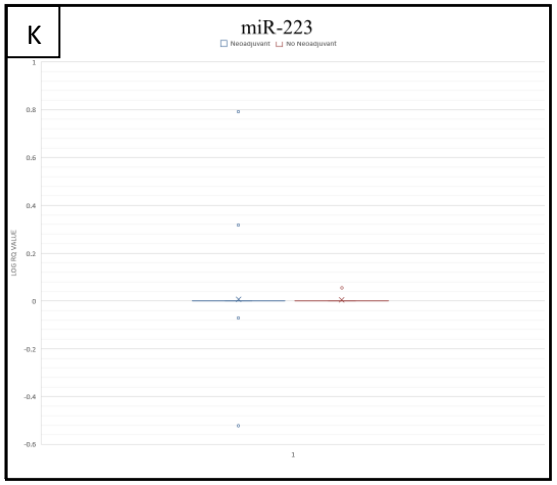

F
miR-21
Neoadjuvant
No Neoadjuvant
LOG RQ VALUE
1
0.5
0
-0.5
-1
-1.5
-2
1

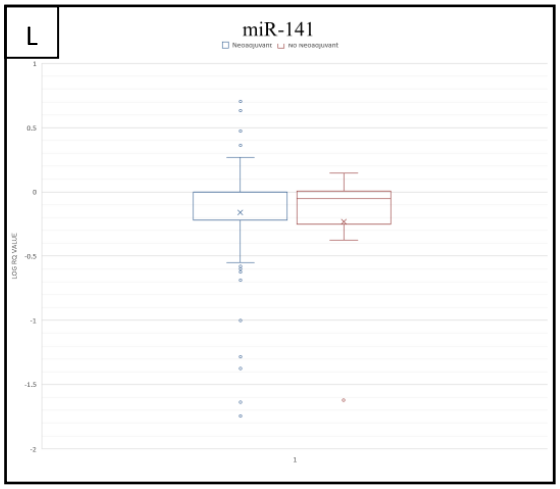

G
miR-17
Neoadjuvant
No Neoadjuvant
LOG RQ VALUE
0.5
0
-0.5
-1
-1.5
-2
-2.5
1

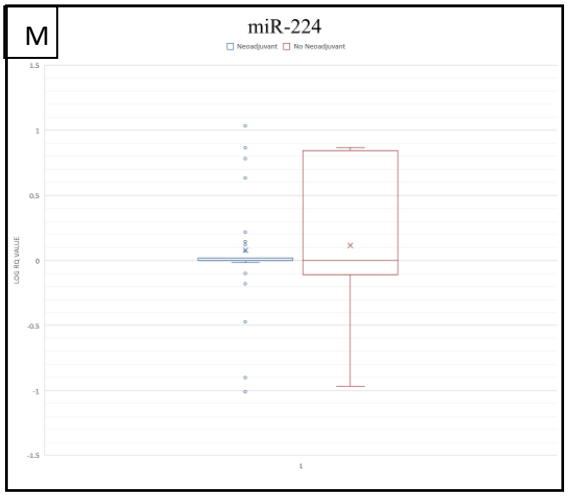

H
miR-16
Neoadjuvant
No Neoadjuvant
LOG RQ VALUE
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
1

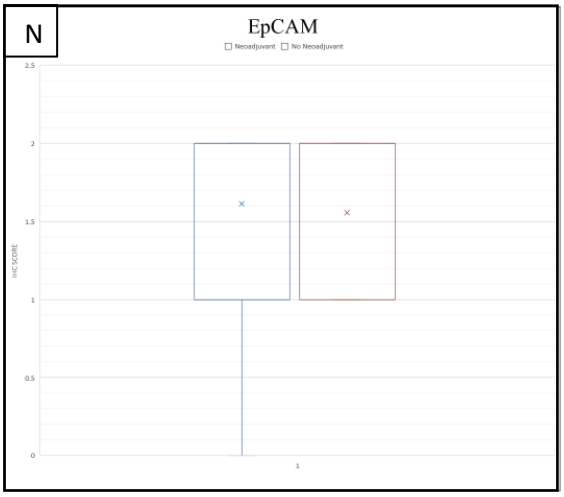

I
miR-10b
Neoadjuvant
No Neoadjuvant
LOG RQ VALUE

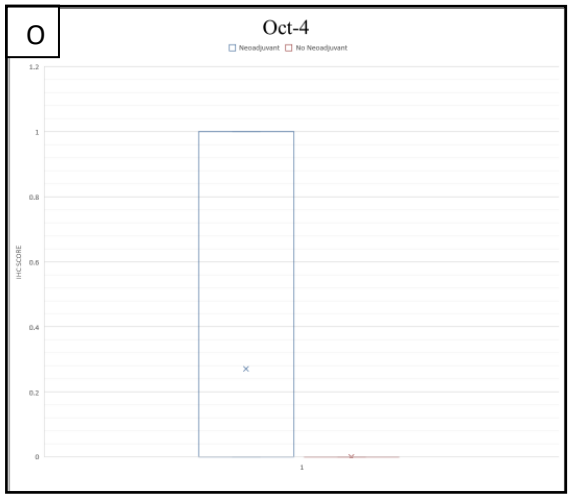

J
miR-221
Neoadjuvant
No Neoadjuvant
LOG RQ VALUE

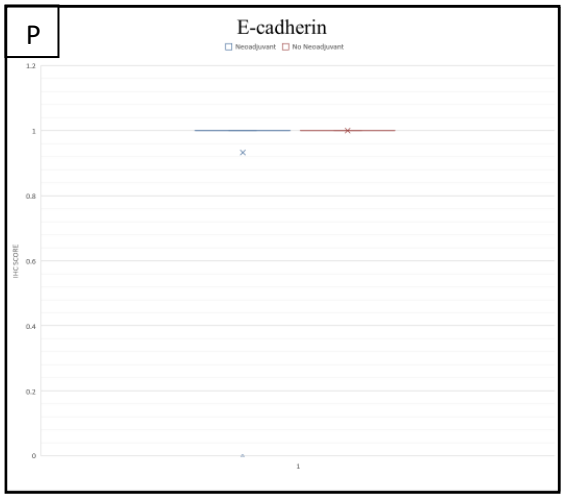

K
miR-223
Neoadjuvant
No Neoadjuvant
LOG RQ VALUE

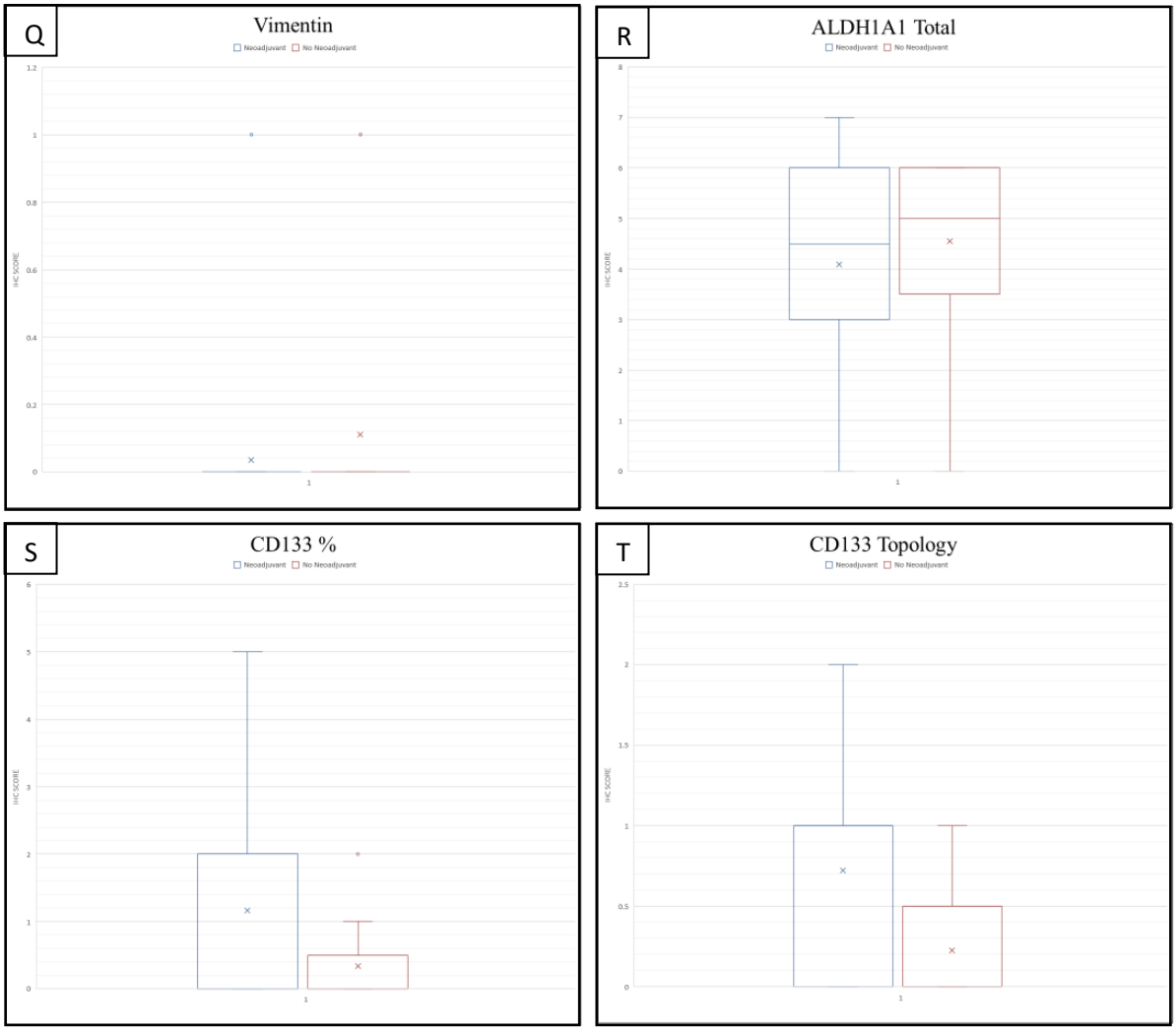

L
miR-141
Neoadjuvant
no neoadjuvant
LOG RQ VALUE


M
miR-224
Neoadjuvant
No Neoadjuvant
LOG RQ VALUE


N
EpCAM
Neoadjuvant
No Neoadjuvant
IHC SCORE


O
Oct-4
Neoadjuvant
No Neoadjuvant
IHC SCORE


P
E-cadherin
Neoadjuvant
No Neoadjuvant
IHC SCORE



Appendix Figure 2.1 Box and whisker plots showing the difference in log RQ values of mRNA, miRNA and expression in higher grade tumour samples between patients who received NAT and those who did not. A = Vimentin (mRNA), B = CDH1, C = NANOG, D = POU5F1, E = Serpine1, F = miR21, G = miR17, H = miR16, I = miR10b, J = miR221, K = miR223, L = miR141, M = miR224, N = EpCAM, O = Oct-4, P = E-cadherin, Q = Vimentin (protein), R = ALDH1A1 Total, S = CD133 Percentage, T = CD133 Topology. No statistically significant differential expression was observed any mRNA, miRNA or protein. Statistical significance: Wilcoxon matched pairs test, $p<0.05$. Data graphed as mean, median, range and interquartile range. X represents the mean. The central horizontal line represents the median. Where present, the top line represents the first quartile and the bottom line represents the third quartile. The whiskers denote the maximum and minimum values.

Appendix Three: miRNA Gene Targets

Appendix Table 3.1 Downstream Gene Targets of a Panel of 12 miRNAs.

miR-16	miR-17	miR-21	miR-221	miR-223	miR-224	miR-10b	miR-133b	miR-203a	miR-200a	miR-141	miR-103a
ACVR2A	ACVR1	DUSP9	ACVR1C	FGF2	ACVR1C	ACVR2A		BMPR1B	ACVR1C	ACVR1	
BMPR1B	ACVR2A	DVL1	ACVR2B	FZD3	ACVR2A	BMPR1B		ID4	ACVR2A	ACVR1C	
KRAS	ACVR2B	FZD3	JARID2	FZD3	FZD10	GSK3B		KRAS	KRAS	IL6ST	
SKIL	AKT2	PCGF6	POU5F1	IL6ST	FZD3	IGF1		PAX6	LIF	ISL1	
WNT10B	APC		SMAD3	PAX6	IL6ST	JAK3		PCGF2	NRAS	MAPK13	
	AXIN2		WNT16	PCGF3	MAPK1	LIFR		PIK3R1	PAX6	REST	
	BMI1		ZIC3	SMAD4	NODAL	SMAD2		SMAD3	PCGF5	SMAD3	
	BMPR1A			WNT5A	PCGF2	SMAD3		WNT10B	SMAD2	TCF3	
	COMMD3-BMI1				RAF1	SMAD4		WNT4	STAT3	WNT8A	
	FGFR1				RIF1	WNT8A			WNT10B	WNT9B	
	FGFR2				SMAD5						
	FZD3				WNT5B						
	LIF										
	MYC										
	PAX6										
	PCGF2										
	PIK3R1										
	PIK3R2										
	SMAD2										
	SMAD3										
	SMAD5										
	TCF3										
	WNT2B										
	WNT7B										
	WNT9B										
	ZFHX3										

Appendix Four: Associations between Molecular Expression Patterns and Clinical Outcomes

Appendix Table 4.1 Significant Associations between Clinical Outcomes and Increased Gene Expression.

Gene Expression	Differential Expression	Clinicopathological Features	Survival Outcomes
Vimentin	No	Low TRG	Poor DFS
CDH1	No		Good DFS
NANOG	No		
POU5F1	No		
Serpine-1	No		

Appendix Table 4.2 Significant Associations between Clinical Outcomes and Increased miRNA Expression.

miRNA Expression	Differential Expression	Clinicopathological Features	Survival Outcomes
miR-221	Yes	Disease recurrence	Poor OS and DFS
miR-21	No	High TRG Disease recurrence	
miR-16	No		Good OS
miR-224	No	High TRG Disease recurrence	
miR-141	No	HER2 positivity	
miR-10b	No		
miR-17	No		
miR-223	No		
miR-200a	No		

Appendix Table 4.3 Significant Associations and Increased Protein Expression.

Protein Expression	Differential Expression	Clinicopathological Features	Survival Outcomes
EpCAM	Yes		
E-cadherin	No		
Vimentin	No		
CD34	No		
Oct-4	No		Poor OS and DFS
ALDH1A1	No		Poor OS and DFS
CD133	No	Disease recurrence	Poor DFS