Student No: 20211808

Module Code: MEDS4006 UNUK

University of Nottingham

Assessed Coursework Cover Sheet for MRes & PGcert in Molecular Pathology, Bioinformatics & Diagnostics 201920

Module Title: Molecular Pathology Research Module

Coursework Title: Histopathological analysis of CD44 and c-Myc in colorectal cancer

Word Count (excluding references and appendices): 24927

This is to confirm that I submit this piece of assessed work in the full knowledge of the published guidelines on plagiarism and its consequences. By completing this coversheet I confirm that I have abided by the University of Nottingham's regulations concerning plagiarism.

X

Histopathological analysis of CD44 and c-Myc in colorectal cancer

Contents

Abstract

Colorectal cancer is the second highest cause of cancer deaths in the UK. The transcriptional regulator c-Myc and transmembrane protein CD44 are reported to be associated with cancer stem cells and epithelial to mesenchymal transition, which have both been linked to poor outcomes. A cohort of 1000 colorectal cancer patients was assessed by immunohistochemistry for expression of CD44, c-Myc and tumour stromal content. TMA blocks comprised of luminal, central and advancing edge cores were used to account for tumour heterogeneity. High nuclear c-Myc expression in tumour epithelial cells was associated with poor 5 year survival ($p < 0.001$). Surprisingly, it was also associated with a lack of metastasis and low tumour stage. In contrast to previous studies, membranous CD44 expression was associated with low primary tumour stage (p 0.001) and metastasis (p 0.003). Cytoplasmic CD44 staining was observed to follow two distinctive patterns, perhaps representing different variants. CD44 and c-Myc expression in the stroma were localised to lymphocytes and plasma cells, respectively. Both were associated with positive clinical outcomes, including low primary tumour stage and 5 year survival, consistent with previous studies regarding immune infiltration. Increased tumour stroma content was associated with worse outcomes, consistent with the "mesenchymal" consensus molecular subtype 4. CD44 expression and tumour stroma content were significantly correlated, supporting the role of CD44 in epithelial to mesenchymal transition. This study presented an opportunity to better understand the expression of c-Myc, CD44 and the tumour stromal content in colorectal cancer in a large number of patients.

Key findings:

- Positive nuclear c-Myc expression was associated with poor survival
- MLH1 deficiency led to increased c-Myc expression
- Stromal c-Myc expression was associated with improved survival
- High CD44 expression was associated with good clinical outcomes
- Stromal CD44 expression was associated with poor survival
- c-Myc+/CD44+ patients showed positive clinical outcomes
- Low stromal content was associated with high CD44 expression
- High stromal content was associated with poor outcome

Introduction

Colorectal cancer pathology

Colorectal cancer (CRC) is the third most common cancer in men and women, and the fourth most common cancer in the UK overall [1]. 44% of patients diagnosed with colorectal cancer die from the disease, and 10 year survival is 57%, with recurrence in 30% of patients [2]. Of those diagnosed with colorectal cancer, more than 95% of patients have adenocarcinoma, which originates in glandular cells within the colon or rectum. 75% are diagnosed at a local stage (I and II). Stage I tumours invade as far as the muscularis propria within the colon, and can usually be treated with surgery alone (85-95% of cases). Stage II tumours may infiltrate beyond the colon wall into adjacent organs, but not into the lymph nodes. Depending on the extent of invasion, 5 year overall survival rates can vary between 58-87%, and patients may be treated with surgery and fluorouracil (5FU)/leucovorin combination, although there is some debate as to the usefulness of chemotherapy at an early stage, when surgery alone has a reasonable success rate. Stage III tumours show some level of metastasis to the lymph nodes, and surgery and radiotherapy with adjuvant treatment has been recommended for these patients. This includes the use of fluoropyrimidine based chemotherapy (typically 5FU) in conjunction with oxaliplatin or targeted agents such as cetuximab [2].

The progression of normal colon into adenoma and eventually into an invasive adenocarcinoma occurs following a series of genetic and epigenetic changes. The three key pathways this follows in colorectal cancer are chromosomal instability (CIN), microsatellite instability (MSI), and CpG island methylator phenotype (CIMP). Chromosomal instability occurs in up to 70% of sporadic colorectal cancers, making it the most common pathway of genomic change. It leads to karyotypic changes in the cells, including loss of heterozygosity and aneuploidy [3]. The most common “canonical” sequence of changes was first described by Fearon and Vogelstein in 1990 [4] and has been expanded on over the last 30 years to show the genes and pathways involved (see figure 1). The initial change occurs when the APC tumour suppressor gene is mutated, typically by frameshift or nonsense mutation, leaving a truncated protein. APC usually blocks transition from G1 to S phase in the cell cycle, and disrupting this allows replication to progress unchecked. As APC regulates β-catenin, these mutations lead to constitutive overexpression of β-catenin and prolonged activation of the Wnt signalling pathway, important in determining cell fate and motility [5]. Mutations in β-catenin or hypermethylation of the APC promoter can also lead to this. The next canonical alteration occurs in KRAS [6] which is mutated in up to 50% of cases of colorectal cancer [7], typically at codon G12. After these mutations occur, loss of heterozygosity (LOH) at 18q is often

observed, particularly at 18q21, especially in advanced colorectal cancer. For loss of heterozygosity to occur, a minimum of one allele of a gene must be lost and the other is often mutated rendering it non-functional. One of the genes that is commonly lost in this step is DCC (Deleted in Colorectal Carcinoma), a tumour suppressor gene that is missing in approximately 70% of colorectal cancers, whose loss is negatively associated with survival [6]. However, DCC is rarely mutated and it is now considered that allelic loss of SMAD2 or SMAD4 is probably the more important target of LOH at 18q [8]. Mutations in PIK3CA occur late in the adenoma-carcinoma sequence. The final mutation that is associated with the adenoma-carcinoma progression is of TP53. The p53 protein is involved in the control of the cell cycle and apoptosis, inducing arrest of the cell cycle between G1 and S phase to facilitate DNA repair prior to replication. The mutation is thought to occur at the time of transformation from an adenoma to a carcinoma.

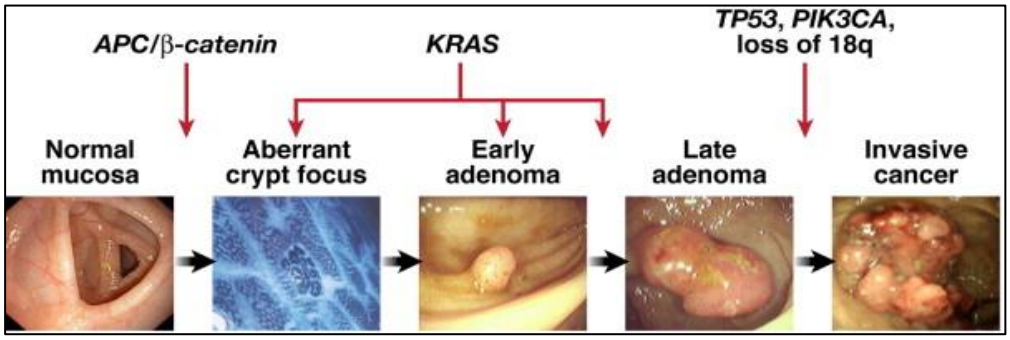


Figure 1: The adenoma-carcinoma sequence first described by Fearon and Vogelstein. An increase in mutations in APC, KRAS, TP53 and PIK3CA are accompanied by loss of heterozygosity at 18q. Tumour progression is associated with these changes. This image is taken from Pino et al, 2010 [3].

Up to 20% of colorectal cancers show an aberrant pattern of methylation associated with clinical outcomes. This subset of CRC is controversially described as molecularly distinct from other groups, including CIN, and is termed the "CpG island methylator phenotype" [9]. CIMP status can be determined by assessing the methylation of the promoter regions of the genes CACNA1G, CDKN2A, CRABP1, IGF2, MLH1, NEUROG1, RUNX3 and SOCS1. Patients with 5 or more of these promoters methylated are classified as CIMP positive [10]. CIMP tumours are associated with proximal colon disease and poor differentiation, much like MSI tumours.

Tumours with microsatellite instability account for roughly 15% of sporadic colorectal cancer. Microsatellite instability was first observed by Ionov *et al* in 1993, where they described the slippage of regions of the genome with tandem repeat sequences as an early event in colorectal cancer [11]. MSI tumours tend to carry more mutations than CIN tumours due to a failure to repair errors made by DNA polymerase during replication. MSI occurs in more than 95% of

patients with Lynch syndrome, a familial form of colorectal cancer that occurs due to faults in the mismatch repair genes. Tumours with microsatellite instability tend to be more frequent in female patients, the proximal colon, and may have mucinous histology [6]. They are often associated with a different mutation profile to that of tumour with CIN. There is a very strong association with genome wide hypermethylation and BRAF mutations (which can be seen in up to 60% of sporadic tumours with MSI). In addition, there are mutations within genes which contain repeat sequences in coding regions such as TGFBR2 and BAX [12, 13].

Inherited colorectal cancer

The vast majority of CRCs arise as sporadic tumours although it is thought that, even in these, there is a genetic contribution of 15% of risk [14]. There are however a number of family cancer syndromes in which germline mutations are associated with a high risk of CRC.

Lynch syndrome, associated with the MSI changes described above, is the most common familial syndrome associated with colorectal cancer and accounts for 3% of new cases. It is also associated with other cancers, the most frequent of which is endometrial cancer. Like spontaneous MSI tumours, phenotypically it is connected with cancer in the proximal colon, and is often poorly differentiated or mucinous. It is caused by germline deficiency in the mismatch repair (MMR) genes MLH1, MSH2, MSH6 or PSM2, or by deletion of EPCAM, which causes inactivation of MSH2. Colorectal cancer develops more rapidly in patients with Lynch syndrome due to the more frequent presence of polyps within the colon and poor mismatch repair ability [15]. As a result of this, patients under the age of 50 diagnosed with colorectal cancer are tested for Lynch syndrome. Testing for Lynch syndrome is performed by either immunohistochemistry (IHC), for loss of MMR protein expression or PCR, for the presence of MSI. Patients with MMR deficiency do not respond to treatment with 5FU [16], as they are insensitive to the mode of action used, but are more sensitive than MMR competent patients to drugs such as FOLFOX, a combination of folinic acid, fluorouracil and oxaliplatin [17].

Other inherited syndromes leading to colorectal cancer include Familial Adenomatous Polyposis (FAP), attenuated FAP, and Turcot syndrome [6]. FAP is a rare disease that affects less than 1 in 10,000 people, but those diagnosed with it have an almost 100% risk of colorectal cancer by the age of 50. It is characterised by the formation of hundreds of adenomatous polyps in the colon at a young age, and is treated with colectomy and regular follow ups to reduce the risk of cancer. Familial adenomatous polyposis is caused by inherited mutations in the APC gene, one of the earlier mutations to occur in

the Vogelstein model of colorectal carcinoma, which is why the cancer occurs in young patients [18].

The consensus molecular subtype (CMS) classification of CRC

Multiple groups have tried to categorise colorectal cancer into groups [19] [20] [21] [22] [23] [24] [25], in addition to characterisation by CIN, MSI or CIMP. The main subtyping that is currently observed is the system proposed by Guinney *et al* in 2015. They combined the data from 18 cohorts, normalised to reduce technical variation, and looked at the data using the subtyping of 6 different groups. They primarily focussed on transcriptomic data but also used proteomics, copy number variation and methylation data. They produced a system that split colorectal cancers into one of four molecular subtypes (CMS1-4) [26]. CMS1 consists of an "MSI immune" type, including most MSI and CIMP high tumours, characterised by widespread hypermethylation and a high frequency of BRAF mutations. Like other MSI tumours, these are associated with right-sided lesions and are more common in females. They also showed expression of genes associated with immune infiltration. CMS2 is the "canonical" subtype, with high chromosomal instability, loss of tumour suppressor genes and gain in oncogenes. These tumours are associated with left-sided lesions and upregulation of the Wnt and MYC pathways, and have the best survival after relapse. The third molecular subtype is CMS3, "metabolic". This subtype is associated with fewer somatic copy number alterations (SCNAs) and a high prevalence of CpG island phenotype. There is a high frequency of KRAS mutations and they are enriched for metabolic pathways. It shares similarities with the "metabolic" subtype described in gastric cancer, hence the name [27]. CMS4 is the "mesenchymal" subtype, which shows an upregulation of genes associated with the epithelial to mesenchymal transition (EMT). Epithelial to mesenchymal transition is important in altering the CRC cells from an organised epithelial type to a more mesenchymal type, allowing easier cell motility, leading to metastasis [28]. There is a high level of stromal infiltration and an overexpression of extracellular matrix proteins, and it is the group associated with the most advanced stages of diagnosis and the worst overall survival [26].

c-Myc

The MYC family of helix-loop-helix leucine zipper proteins consists of c-Myc, n-Myc and l-Myc, which are involved in multiple cellular functions, including cell growth, proliferation, metabolism and more [29]. c-Myc is the best characterised of these proteins, and is thought to regulate approximately 15% of genes. The oncogenic function of c-Myc was first demonstrated in its role in Burkitt's lymphoma in 1982. c-Myc is found on human chromosome 8q24, and in Burkitt's lymphoma is translocated into the immunoglobulin μ chain gene [30]. The sequence of the c-Myc gene was determined along with its structure

in 1983. It contains three exons and two introns, and is located on chromosome 8q24 [31]. This gene produces a 65 kDa, 439 amino acid phosphoprotein that is typically expressed in the nucleus [29].

c-Myc usually acts in a heterodimer with the bHLH zipper protein, MAX, to perform its cellular functions. The heterodimerisation of c-Myc and MAX allows them to recognise Enhancer-boxes (E-boxes) on their targets, which c-Myc alone would not have the affinity for [32]. The canonical sequence of E-boxes is CACGTG, and is recognisable by c-Myc-MAX, although some non-E-box sequences can be recognised when c-Myc-MAX are at high concentrations, and this may be involved in the oncogenic overexpression of c-Myc targets [33]. These interactions are supported by the c-Myc binding protein TRRAP, to upregulate gene expression and recruit histone acetyltransferases. MAX is also able to dimerise with the MAD protein, forming the MAD-MAX complex, competing with c-Myc-MAX in order to repress transcription [34]. In addition to c-Myc and MAD, MAX can bind MNT and MGA proteins to control transcription in a contextual manner [32].

c-Myc expression is associated with many different cellular processes (figure 2), including the opposing functions of proliferation and apoptosis. It has been theorised that this may depend on a threshold of expression, where it switches from a proliferative function to an apoptotic one. Or, it may be dependent on its interaction with other molecules [35]. c-Myc is one of the most frequently amplified oncogenes seen in many different human cancers including CRC. In addition, in colorectal cancer it is stimulated by the dysregulation of the Wnt signalling pathway. Loss of functional APC occurs in more than 75% of CRC patients [36], resulting in unchecked expression of β-catenin and Wnt signalling. The accumulation of β-catenin leads to activation of the TCF/LEF family of transcription factors, which are responsible for the activation of c-Myc [37]. c-Myc is rarely mutated in CRC, and instead overexpression is induced by β-catenin and TCF4 in colorectal tumours [38]. The relationship between c-Myc, LEF1 and nuclear β-catenin has been theorised to be important to maintain the prolific state of colorectal cancer cells [39]. The role of c-Myc in the Wnt signalling pathway is also connected to its purpose in maintaining a stem-state [40]. The polycomb protein BMI1 activates Wnt signalling, leading to upregulation of c-Myc. In turn, c-Myc binds to the E-box in BMI1, initiating a positive feedback loop of the Wnt signalling pathway [41].

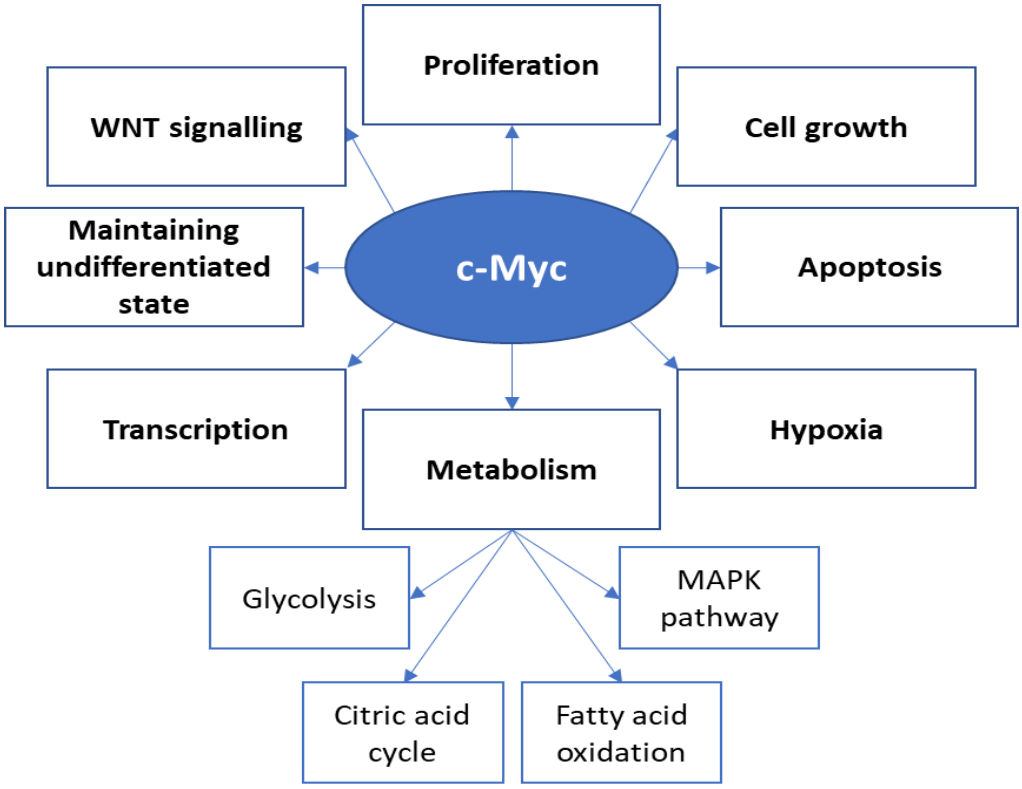


Figure 2: Pathways c-Myc is involved in in colorectal cancer. c-Myc plays a role in regulating approximately 15% of genes, in a multitude of pathways, some of which are shown here.

c-Myc in stem cell biology

c-Myc expression is tightly regulated in normal tissues, but its functions in embryonic stem cells are such that loss of c-Myc leads to embryonic lethality. This is brought about by early differentiation into progenitor-like cells, via proteins such as Oct4 and Nanog. The reliance of c-Myc on MAX to enable binding to its targets means that when MAX is lost, the undifferentiated state of cells is also lost. Levels of Sox2, Nanog and Oct3/4 gradually decrease when MAX is lost, driven by the upregulation of the MAPK pathway, whereas differentiation markers of endoderm, ectoderm and trophoderm, such as Gata6, Sox17, Nestin and Cdx2, are upregulated [42]. When the first induced pluripotent stem cells (iPSCs) were produced, c-Myc, along with Oct3/4, Klf4 and SOX2, was essential in inducing and maintaining the dedifferentiated state of the cells [43]. Chimeric mice from these iPSCs are prone to an increased risk of tumour formation, owing partly to the reactivation of c-Myc. Omitting c-Myc from the iPSC protocols reduces the efficiency of their generation, but greatly reduces the risk of tumourigenesis. It has been observed that l-Myc can be used to replace c-Myc in order to produce human iPSCs without increasing the risk of tumourigenesis [44].

The role of c-Myc in colorectal cancer cell metabolism was demonstrated by Satoh *et al* in 2017, in the correlation between c-Myc expression and more than 230 genes involved in metabolism. Many of these genes were upregulated, and involved in such pathways as purine/pyrimidine synthesis, glycolysis and the

MAPK pathway. Pathways that were downregulated in conjunction with decreased c-Myc expression included fatty acid oxidation and the citric acid cycle. The role of c-Myc in glycolysis suggests that it is a key player in producing the Warburg effect, where glucose metabolism is increased and fermented to lactate. When c-Myc was knocked out, the expression of many of the genes identified returned to normal. The role of c-Myc as a key regulator of metabolic reprogramming in CRC suggests that it has potential as a therapeutic target [45].

c-Myc has a role in the adaptation of cancer stem cells to a hypoxic environment, mediated by the Warburg effect [46]. PKM2 is an important driver of glycolysis, and plays an important role in the Warburg effect. Its expression is upregulated by c-Myc in cancer stem cells, and it activates hypoxia inducible factor 1 (HIF1) in a positive feedback loop. c-Myc also promotes glycolysis via glucose transporter type 1 (GLUT1), and pyruvate dehydrogenase kinase 1 (PDK1). The intracellular form of CD44 suppresses PKM2, promoting glycolysis and increasing the activity of the pentose phosphate pathway [46]. Triple negative breast cancer is heavily dependent on glutamine metabolism, mediated by amino acid transporters such as xCT/SLC7A11. The transmembrane glycoprotein CD44v stabilises this transporter in the cell membrane of cancer stem cells. CD44v increases the concentration of intracellular cysteine, aiding in the production of glutathione (GSH). GSH is involved in the elimination of reactive oxygen species, preventing apoptosis or differentiation of cancer stem cells [47]. The role of CD44 in colorectal cancer will be discussed further later in this thesis.

The role of c-Myc in colorectal cancer

c-Myc expression is associated with the canonical (CMS2) process of colorectal cancer development and better survival outcomes, including after relapse [26]. Overexpression of MYC mRNA is seen at all stages of CRC, from adenoma development to metastasis [38]. c-Myc has been observed to be linked to mismatch repair (MMR). In a study by Partlin *et al*, they demonstrated that c-Myc and MAX bind to MLH1 and MSH2 both *in vivo* and *in vitro* [48]. They used coprecipitation to show that c-Myc is found when MLH1 and MSH2 are pulled down. In another study by Bindra *et al,* 2007, they discuss whether hypoxia is a controlling factor in mismatch repair. They initially suggest that hypoxia led to the downregulation of MLH1 and MSH2, as the protein HIF1α displaces c-Myc from its binding site in a MAD-MAX like fashion. However, they then discussed the possibility that the downregulation of MLH1 and MSH2 occurs independent of HIF-1α expression, and relies only on the interaction of c-Myc and other MAX-associated proteins [49]. They also found that c-Myc was down-regulated under hypoxic conditions. c-Myc has been related to epithelial to mesenchymal transition (EMT), where it enhances SNAIL expression [50]. EMT is involved in the development of metastasis, leading to cancer progression and poor prognosis [51].

Given the role of overexpressed c-Myc in colorectal cancer, it would seem a good choice for therapeutic targeting. In the past scientists were reluctant to target c-Myc directly, due to the large number of genes under its transcriptional regulation, including in normal tissues, and the homology to its family members [52]. Direct inhibitors of c-Myc such as Omomyc [53], which binds to c-Myc binding sites and to MAX, and bromodomain and extraterminal domain inhibitors (BETi) [54], which bind the c-Myc super enhancer and repress it, are examples of strategies that could prevent the oncogenic expression of c-Myc and stop the proliferation of tumour cells.

CD44

CD44 (also known as Hermes, PGP1, Indian blood group and HCAM) is a transmembrane glycoprotein associated with colorectal cancer. It has multiple variants due to alternate splicing, the most common of which is the standard form CD44s. The other variants are denoted by CD44v. These different variants are associated with different cell types and disease phenotypes. The main ligand of CD44 is hyaluronic acid (HA) which is an extracellular matrix protein.

CD44 is found on chromosome 11p13 and has 20 exons. Alternative splicing of these exons in the pre-mRNA occurs to produce multiple isoforms. CD44s, the standard form of CD44, is found in lymphocytes and fibroblasts, and can also be known as CD44h (haematopoietic CD44). The structure of CD44 proteins can be seen in figure 3. It consists of an N-terminal domain at exons 1-5, which binds to the extracellular matrix via proteins such as hyaluronic acid (HA), its main ligand, and others such as fibronectin and laminin. There is then a stalk shaped structure which may be elongated to include the variable exons in CD44v proteins. At the C-terminal of the protein are a transmembrane region, associated with lipid raft binding, and a cytoplasmic tail [55]. This cytoplasmic tail contains important intracellular signalling motifs and mediates the interactions with the cytoskeleton. CD44 is commonly found at the basolateral end of polarised epithelial cells and on the leading edge of migrating cells.

Alternative splicing and post-translational modification are key factors in producing CD44 proteins with distinctive functions. CD44s occurs when the variable exons are spliced out, leaving exons 1-5 and 16-20 (excluding exon 18) to form an 85 kDa protein [56]. Alternate forms of CD44 include CD44v6, commonly associated with colorectal cancer, CD44v3, seen in cancer stem cells in head and neck cancers, and CD44v8-10, associated with gastric cancer [57]. CD44 is prone to both N and O glycosylation [58], with O glycosylation in particular being associated with CRC. The addition of these glycosylate groups, as well as other post-translational modifications such as phosphorylation, contributes to the wide range of functions and high molecular weight of CD44 [56]. Aberrant O-glycosylation of CD44 is associated with dysregulated expression and delivery, and a higher expression of CD44 being observed in exosomes [59].

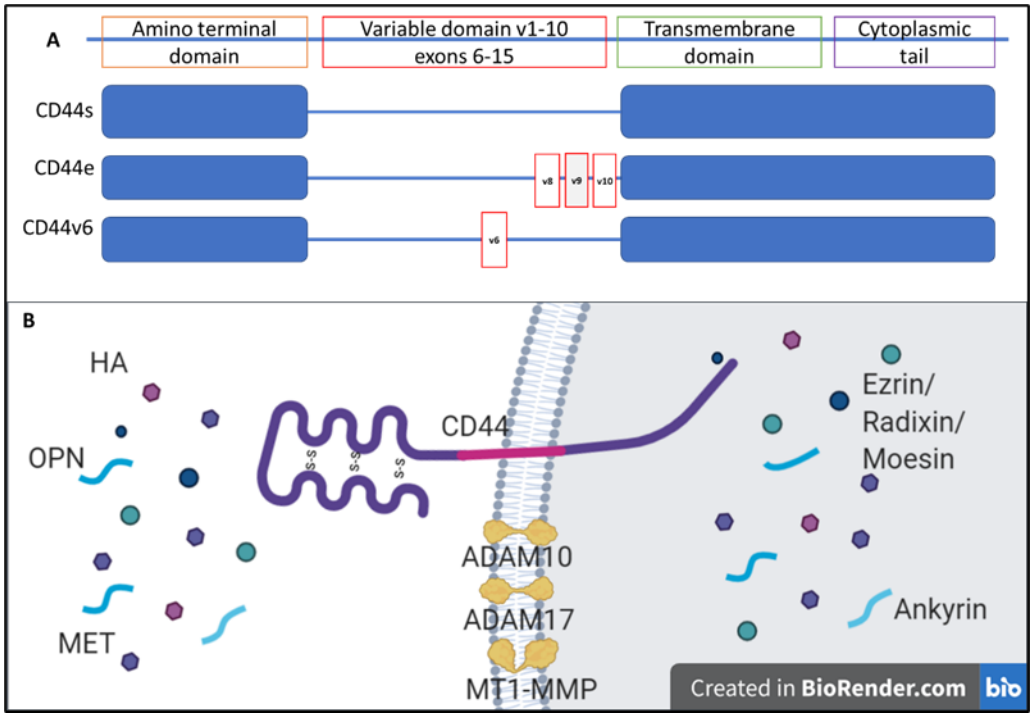


Figure 3: The structure and subcellular location of CD44. A shows the key domains in CD44, the amino terminal domain, the variable domain which consists of 10 exons that undergo alternate splicing to produce multiple proteins, the transmembrane and cytoplasmic domains. Key isoforms of CD44 in CRC include the common form of CD44 is CD44s, an epithelial form (CD44e), and CD44v6, which is associated with poor prognosis in adenocarcinoma. B shows the CD44 protein anchored into the cell membrane. The extracellular domains interact with HA, OPN and MET, whereas the intracellular domains interact with ERM and ankyrin. CD44 may undergo proteolytic cleavage by ADAM10, ADAM17 and MT1-MMP to release the extracellular or intracellular parts of the proteins into their environments.

CD44 is expressed in multiple cell types, including epithelial, lymphocytic, granulocytes and neuronal cells [60]. The main ligand of CD44, hyaluronic acid, is a key component of the extracellular matrix (ECM). HA is a glycosaminoglycan (GAG) protein, which functions as a structural protein and plays a role in wound healing and inflammation, cell motility and proliferation [61]. The interaction of CD44 with HA is reflective of its involvement in cell motility, and may be conducive of cells moving through the extracellular matrix [55]. The N-terminal domain of CD44 contains other protein-binding domains that have led to the conclusion that CD44 binds other proteins in the ECM, including other GAG proteins, growth factors, cytokines and matrix metalloproteinases, although this has yet to be confirmed [60].

Intracellular functions of CD44

In addition to the interactions with hyaluronic acid and the extracellular matrix, CD44 may play a role in extravasation, by binding to E-selectin. It has been demonstrated that both CD44s and CD44v isoforms can mediate this interaction, on hematopoietic cells and colon cancer cells respectively [62]. Cellular migration of both leukocytes and cancer cells can be mediated by CD44-HA interaction. One important mechanism that allows this to occur is the proteolytic cleavage of CD44 by metalloproteinases as seen in figure 3. Okamoto *et al* (1999) have shown that this cleavage differs in mode of action

between leukocytes and cancer cells, but occurs in the same location regardless of CD44 isoform, and the difference is most likely due to post-translational modifications. They found a 25 kDa protein that specifically reacted to their CD44*cyto* antibody but not their CD44*ecto* antibody. The expression of this protein was also increased in those cell lines that had a higher level of soluble CD44 in the supernatant [63]. In a later publication, they demonstrated the importance of cleavage in multiple tumour types, including colorectal carcinomas. They found that 90% of colon carcinoma tumours showed cleavage of CD44, where their matched normal partners did not [64].

CD44 is subject to proteolytic cleavage, and therefore can exist in three forms, a full-length, membrane bound protein, a soluble extracellular cleavage product and an intracellular fragment (CD44ICD). Ectodomain cleavage of CD44 is stimulated by membrane-associated matrix metalloproteinases and Ras proteins. CD44 ectodomain cleavage is dependent on ADAM10 Ca^{2+} stimulation or ADAM17 activation by Rac, and is associated with detachment from the ECM. Another metalloproteinase associated with CD44 ectodomain cleavage is MT1-MMP. This protein is expressed in many tumour types and is localised to the leading edge of the cell in the lamellipodia. Further cleavage of CD44 occurs and depends, as in many other type I transmembrane proteins, on γ-secretase. This releases the CD44 intracellular domain (ICD), which can migrate to the nucleus and activate gene transcription [65].

The cytoplasmic domains of CD44 contain residues for binding to intracellular proteins. These include ezrin/radixin/moesin (ERM) proteins and ankyrin. The ERM proteins bind to long actin filaments, and ankyrin binds to the protein spectrin. The network of ankyrin, spectrin and other accessory proteins and short actin oligomers is thought to play a role in protein and lipid assembly for signalling and other intracellular functions. These interactions are not thought to play a part in the cell adhesion function of CD44, but may be important with regards to hyaluronic acid signalling [66]. The binding site of ankyrin (amino acids 304-318) is downstream of the ERM binding site (amino acids 292-300), and binding of these proteins is phosphorylation dependent via Rho A and PKC, respectively [55]. Ankyrin and ERM proteins facilitate clustering of CD44 within the cell membrane to optimise E-selectin interaction, important in cell rolling [66].

CD44 in the colon and colorectal cancer

Three CD44 isoforms are commonly studied in the gut epithelium. CD44s is the most commonly expressed isoform, found in lymphocytes and fibroblasts, in addition to epithelial cells and in colorectal cancer. It consists of the first five exons and the final five exons. CD44v6 and CD44v4-10 are also commonly expressed in the colon epithelium [67]. CD44v4-10 is highly expressed in colon stem cells, and is involved in the initiation of adenoma. CD44v is also associated with familial adenomatous polyposis in mouse models [68]. Given its

subcellular location and involvement in metastasis, it has been suggested that CD44 is under the influence of the Wnt signalling pathway. Wnt signalling is often upregulated in CRC, which may account for the high levels of CD44 seen in patients [67]. This pathway is important in epithelial to mesenchymal transition (EMT), as demonstrated by the increase of β-catenin in poorly differentiated tumour cells and cancer stem cells [69]. CD44 is described as being a marker of stem cells both in normal and in malignant colon [70]. Along with Lgr5 and EpCAM, it has been used to identify a CSC population that is responsible for increased EMT [71]. CD44 isoform switching has also been implicated in EMT. Studies have shown that CD44s is associated with mesenchymal-like cells, and CD44v associated with epithelial-like cells [72].

CD44v6 is a particularly important isoform in CRC. It is known to be associated with metastasis and poor outcome of patients, and binds to hepatocyte growth factor receptor (MET) in the presence of hepatocyte growth factor (HGF) and osteopontin (OPN) produced in the tumour microenvironment [73]. OPN is involved in multiple biological processes including bone formation, inflammation, immune response and tumourigenesis. CD44v6 specifically has an OPN binding domain that allows the two proteins to interact, mediating metastasis. It exhibits pro-metastatic effects on tumours by enabling tumour cells to disseminate via hypoxia-induced angiogenesis and extracellular matrix remodelling. In addition, tissues which physiologically produce OPN may be targets for tumour homing, such as the bone or liver [74]. In a study by Todaro *et al*, they demonstrated that CD44v6 positive cancer stem cells led to a more aggressive and invasive phenotype both *in vivo* and *in vitro*. Expression of CD44v6 was increased in response to HGF, OPN and SDF-1 exposure, leading to enhanced migration and metastasis. These cells also showed constitutive activation of the PI3K/AKT pathway, and confirmed that CD44v6 expression correlates with Wnt/β-catenin pathway activation [73].

CD44v6 clinical effect

In many cancers, a high percentage of CD44v6 positive cells is indicative of a poor clinical outcome, likely due to the association with increased metastatic potential. CD44v6 overexpression has been shown to be related to stage and lymph node metastasis, as well as distant metastasis, and is particularly associated with lung and liver metastasis [67, 75]. The interaction with OPN may facilitate this, as described above. It has been suggested that p53 exhibits an inhibitor function on CD44, and that mutation of p53, common to CRC, allows the anti-apoptotic and proliferative activities of CD44v6 to run unchecked [76]. CD44v6 shows a stronger affinity for HA than CD44s, which is important for chemoresistance. Their interaction stimulates PI3K/Akt activation, which stimulates the production of P-glycoprotein, which allows the tumour cells to pump toxins, including therapeutic agents, out and into the lumen [77].

Project overview

The aims of this project are to understand the importance of two biomarkers, c-Myc and CD44, in colorectal cancer. 1000 patients provided tumour samples from locations including luminal, central, advancing edge and normal colon. These samples were then stained immunohistochemically with c-Myc and CD44 antibodies, and the slides scanned. Those images were provided for histopathological analysis. Upon completion of the histological analysis, the two biomarkers were subjected to statistical analysis alongside the clinical data, to determine if c-Myc or CD44 may have prognostic value. Tumour stroma content was similarly assessed to determine if it impacted clinical outcomes. Finally, bioinformatic analysis was performed using the Cancer Genome Atas (TCGA) colorectal cancer data, to analyse the pathways and processes that these biomarkers are involved in.

Methods

Patient cohort

This study was performed on a Tissue MicroArray created from a cohort of 1000 sequential patients treated for CRC at the Nottingham University Hospitals between 2008 and 2014. Specimens were obtained from colorectal cancer tissues and normal colonic mucosa and the following data about the patients were retrieved: sex, age at diagnosis, survival data, cause of death, tumour site, tumour size, lymph node involvement, metastases, final stage, dukes, stage, grade, recurrence, vascular invasion, perineural invasion, lymphovascular invasion, peritumoral lymphocytes, tumour edge, tumour budding, MLH1/PMS2/MSH2/MSH6 expression and MMR status. The creation of the TMA and the retrieval of the clinic-pathological data was undertaken by a senior research assistant in the lab.

Histopathology methods

Materials

Table 1: Materials used in IHC analysis

Item	Supplier	Catalogue number
Tissue Microarrayer	Beecher Instruments	MTA-1
X-tra Slides	Leica Biosystems	3800050
Monoclonal Mouse Anti-human CD44 Clone DF1485 Concentration 1:50	Agilent	M7082201-2
Recombinant Rabbit Anti-c-Myc antibody clone Y69 Concentration 1:20	Abcam	Ab32072
Benchmark ULTRA IHC/ISH System	Roche	N750-BMKU-FS05342716001
Pannoramic Slide Scanner	3DHistech	-

Methods

Colorectal cancer Tissue MicroArray (TMA) blocks were prepared according to standard procedures as previously described [78]. Formalin fixed, paraffin embedded (FFPE) tissue blocks and their corresponding haematoxylin and eosin (H&E) slides were retrieved. Regions of interested were annotated on the H&E slides by a pathologist, to identify the areas to be cored. Tissue cores with a diameter of 0.6 mm were extracted from the original block and transferred to a recipient block by a Tissue Microarrayer (Beecher Instruments, US). Four cores were taken from 1000 colorectal cancer samples including three from the tumour (luminal, central and advancing edge) and one from the normal adjacent area, and split across 15 TMAs with controls. Three regions from the tumour were taken, to allow interrogation of intratumoural heterogeneity, common in colorectal cancer [79]. TMA blocks were cut at a thickness of 4 µm and mounted on slides (Surgipath X-tra Adhesive, Leica, Germany). These slides were then stained with CD44 (Agilent, US) and c-Myc (Abcam, UK) antibodies on a Benchmark Ultra (Roche, US) and imaged with a 3D Histech slide scanner producing .mxrs format files. Images were analysed in QuPath v0.2.0-m11. Immunostaining was undertaken by the Histopathology department of the Queen's Medical Centre using standard techniques. Scanning was performed at X20 magnification.

Image analysis

QuPath Set up

QuPath-0.2.0-m11 [80] was used to open and analyse the TMA images. This is a free software which allows the user to perform manual or automated scoring of tissue samples. In this case only manual analyses were done, due to the nature of the antigens targeted; their presence in multiple cellular compartments meant scoring by eye minimised conflicting results. The process for creating a project (to save all images from one biomarker into one analysis file) and opening images was performed using commands within the application as follows:

Create new project > Select Folder > Add images > Choose files > Set image type > Rotate image > Import

Images were rotated 180° (see command above) as the majority of TMAs were originally displayed upside down. Cores were labelled with a column label (1-12) and a row label (A-Z) to facilitate accurate analysis and identification of the cores of interest within the slide. Core size was set to 0.75 mm for all TMA slide images. TMA segregation was performed using the following commands:

TMA > TMA dearrayer > TMA core diameter (0.75mm) > Column labels (1-12) > Row labels (A-Z) > Run

Labelling the TMAs in this way gave the advantage that the cores could be annotated for core presence/absence by selecting the cores by double clicking

until highlighted yellow, right clicking and selecting “Set core valid” to mark the core for analysis or “set core missing” to mark the core as missing or nonscorable. Where the software had not recognised the correct positioning of cores (for example due to faint staining), the TMA selection boxes could be moved to the correct location by double-clicking and dragging the box to the desired location. This allowed easier visualisation of the cores within the image. In addition, some cores were larger than the boxes (figure 4f), and placing the boxes over the cores meant that the same sized region was used for all analyses. To reduce bias from this process, the boxes were placed in the middle of the cores (where possible) and that area was then scored.

Exclusion criteria

Cores were excluded as seen in figure 4 if:

- Cores were absent (figure 4 H)
- Cores were rendered nonscorable by:
    - Folding or tissue lifting leading to insufficient epithelial content (figure 4 A & E)
    - Out of focus (blurry) images (figure 4 G)
    - Cores overlapped/were too close together to be discernible
    - Cores were jumbled and unassignable (figure 4 F)
    - Epithelial content was less than 15% (figure 4 B)
    - Unexpected tissue type was present e.g. normal colon present where cancer was expected and vice versa (figure 4 C & D)

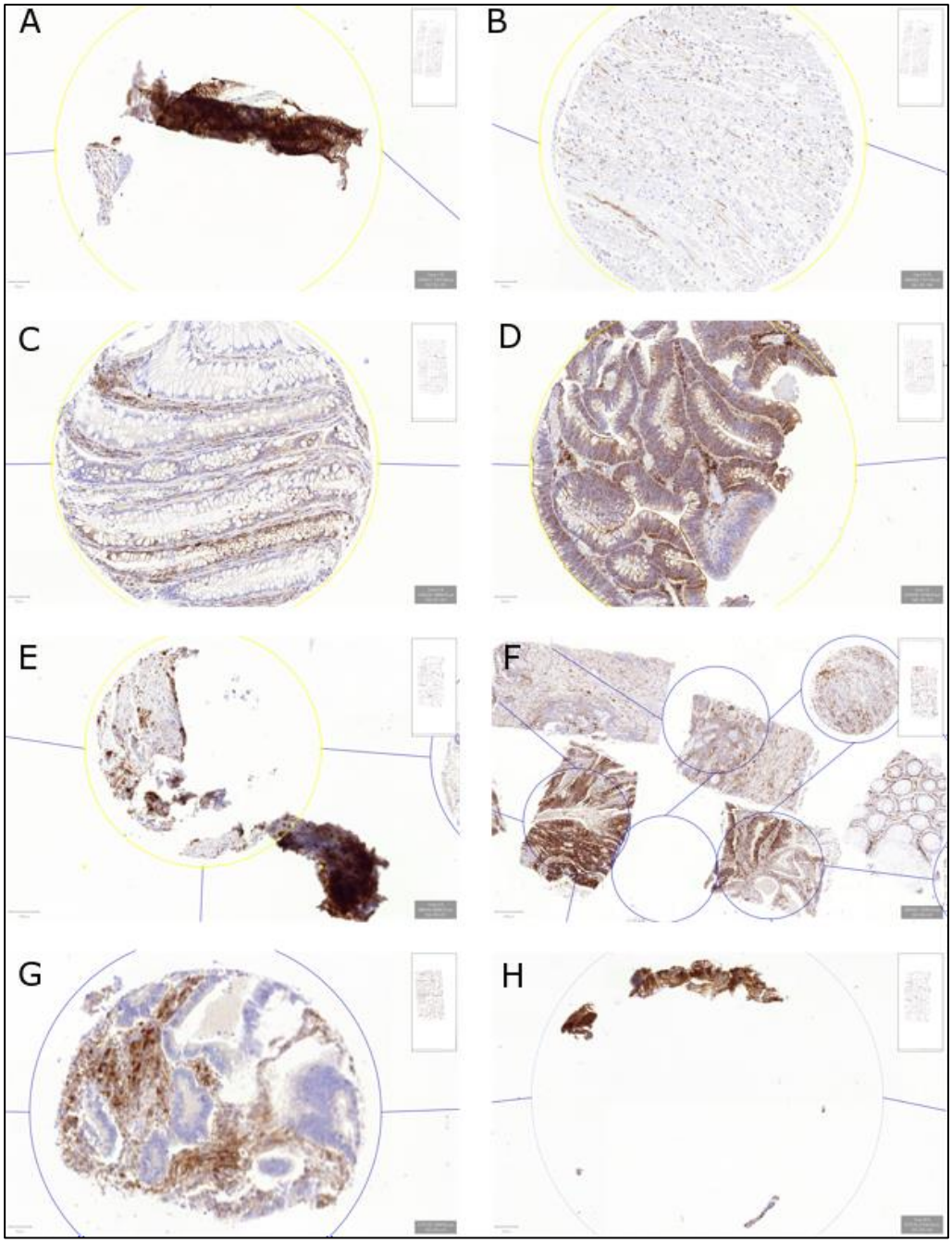


Figure 4: Exclusion criteria for QuPath image analysis. Cores were excluded for a number of different reasons including folding (A), insufficient epithelial content (B), incorrect epithelial type i.e. presence of normal colon (C) where tumour is expected or presence of tumour (D) where normal colon is expected. Other reasons for exclusion included tissue lifting (E), cores being unassignable (F), out of focus images (G), or lack of tissue (H).

Histoscore calculation

The histoscore (H-score) is a widely used method of assessment of immunostaining which allows heterogeneity of staining to be incorporated into the final score. The method usually depends on identifying 4 bands of intensity staining i.e. 0 = no staining, 1=weak staining, 2 = moderate staining, 3 = strong staining. The percentage of cells in each band is multiplied by the value of the

band. The histoscores (H-score) for each region of interest were calculated as follows:

Histoscore = (0 x (%0 cells)) + (1 x (% 1+ cells)) + (2 x (% 2+ cells)) + (3 x (% 3+ cells))

This methodology allows for a minimum score of 0 and a maximum score of 300. For this study, in the case of CD44 cytoplasmic staining, there was a maximum score of 200, because the staining was consistently low intensity.

Histoscores for every core in the tumour regions of luminal tumour, central tumour and advancing edge tumour were calculated and averaged to account for intratumoural heterogeneity. The H-scores of each tumour core were averaged assuming they met the analysis criteria described above. The H-score of the normal tissue was kept separate for comparison. H-scores for each cellular compartment (i.e. nuclear, cytoplasmic, membranous) and for stroma were generated individually. Examples of typical staining for each biomarker can be seen in figure 5. Cytoplasmic scoring for c-Myc was assessed as percentage positive cells as the staining was homogeneous between samples.

CD44 H-score assessment

CD44 staining was assessed for the following conditions:

- Cytoplasmic staining: scored from 0-2
- Membrane staining: scored from 0-3
- Nuclear staining: scored from 0-3
- Stroma staining: scored from 0-3

c-Myc H-score assessment

c-Myc staining was assessed in the following conditions:

- Nuclear staining: scored from 0-3
- Cytoplasmic staining: percentage positive cells
- Stromal staining: scored from 0-3

Epithelial content assessment:

The epithelial content of each tumour region was assessed prior to biomarker analysis. This was necessary to determine if the cores were suitable for biomarker analysis, and allowed the assessment of tumour epithelial content with clinical outcomes. In addition, it gave an opportunity to compare the two sets of slides, as epithelial content may vary from section to section as the tumour changes.

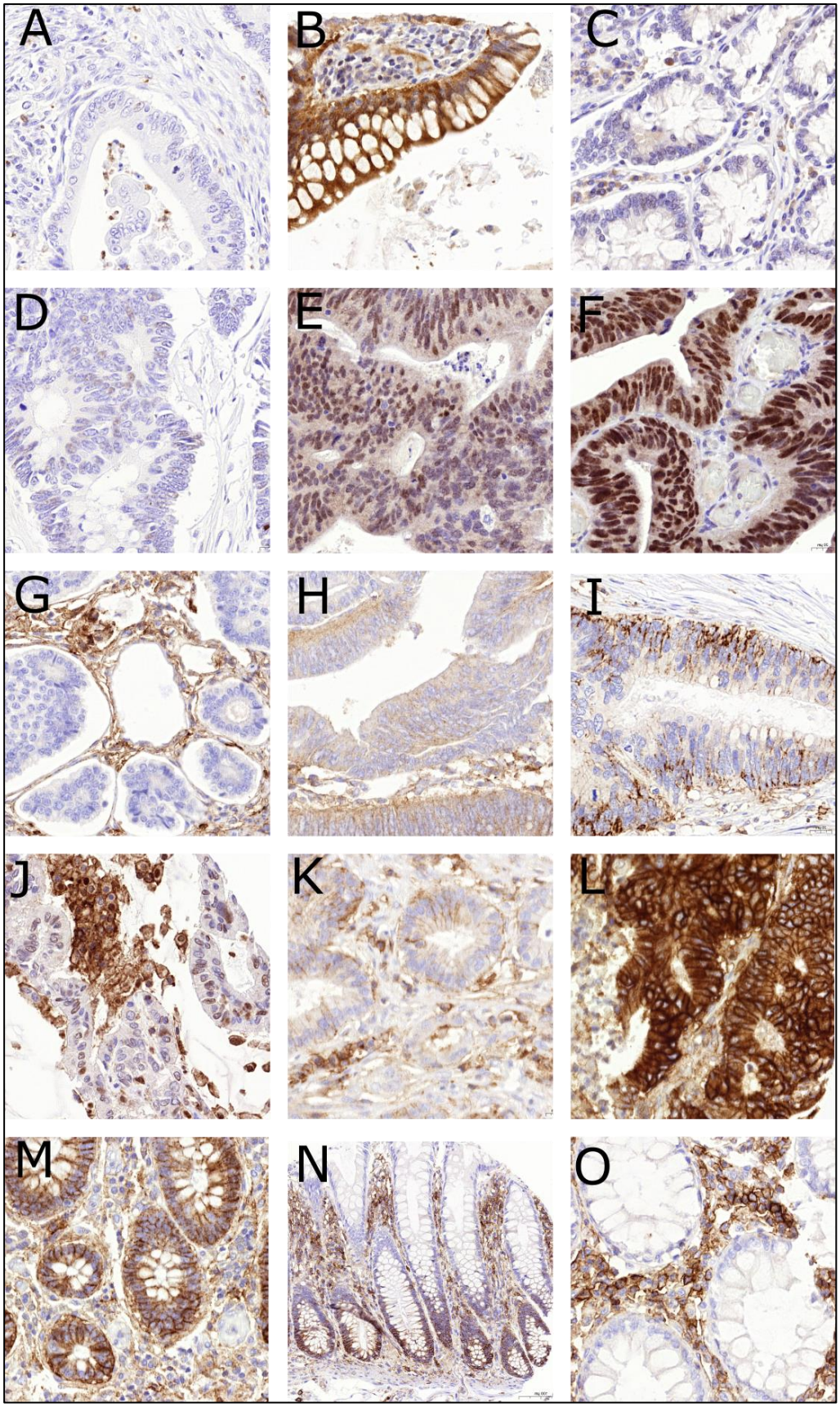


Figure 5: Examples of expression of c-Myc (A-F) and CD44 (G-O) in colorectal carcinoma (A-L) and normal colon (M-O). c-Myc: Negative epithelial staining can be seen in A. Cytoplasmic staining in c-Myc (B) was scored for percentage of positive cells. Plasma cells were positive for c-Myc as seen in the stroma of C. D-F show nuclear c-Myc staining varying from 1 (low intensity) to 3 (high intensity). CD44: G shows a tumour

sample with no epithelial CD44 staining. Cytoplasmic staining of CD44 was represented as a homogenous stain (H) or as a granular stain (I). Nuclear staining (M) was rarely observed for CD44. Immune cells in the stroma were frequently stained as in J. CD44 membranous staining in tumour samples can be seen in K (low intensity 1) and L (high intensity 3), and in normal samples (M, moderate intensity, 2). . Epithelial staining in the normal colon showed a stratified pattern of staining for CD44 (N), with the lower part of the crypts being positive and the upper parts were negative. CD44 staining was also observed in the stroma, particularly in normal samples, as in O.

Statistical analysis

All statistical analyses were performed using IBM SPSS Statistics v26 licensed from the University of Nottingham.

The H-score is prone to subjective variation, due to each pathologist scoring staining slightly differently. H-score reliability was assessed in two ways. First, the inter-observer reliability was assessed by an independent pathologist. Dr Abhik Mukherjee gave H-scores for several tumours for each biomarker. When compared to my scores, these were within 10% of each other. Traditionally, a variability in scores of 10% between pathologists is accepted as a high Kappa score. To ensure that intra-observer reliability was consistent throughout the analysis, the H-scores for a number of tumours were recalculated on another day, to calculate the Kappa statistic. For epithelial content and nuclear, cytoplasmic and membranous staining (where applicable) in both proteins, these were all highly correlated, above 0.8.

Additional variables were computed from existing variables as presented in table 2. This was done using Transform>Recode into different variables. Clinical variables were re-coded in order to perform Chi-square analysis, or perform analyses that depended on bivariate or categorical variable input.

Table 2: Variables computed from existing variables into categorical variables

Name	Values	Cut-off
Clinical variables		
Agedichotomised	1, young; 2, old	Below/above 50
GradeDichotomized	1, low; 2, high	1&2, 3
Sitedichotomized	1, Rightside; 2, Leftside	Right colon, other and left colon, rectum
c-Myc variables		
NucAvHScoreMedianSplit	0, Below; 1, Above	Below/above 54
NuclearLowestTertile	1, low; 2, high	Below/above 26.33
NuclearHighestTertile	1, low; 2, high	Below/above 85.33
Nuclearpositive	1, negative; 2, positive	0, any other value
Nuclearalltertiles	1, lower; 2, middle; 3, higher	0>26.33, 26.33>85.33, 85.33>300
CytoAvPositiveMedian	None	Below/above 2
CytoAvPositiveonly	None	0, any other value
StromalAverageMediansplit	1, below; 2, above	Below/above 6.67
StromalAverageHscorePositiveOnly	None	0, any other value
Name	**Values**	**Cut-off**

StromalAverageHScoreLowtertile	None	Below/above 3.5
StromalAverageHScoreHightertile	None	Below/above 10.3
CombinedNuclMedCyto75th	None	See c-Myc results
CD44 variables		
MembraneHScorePositivesonly	None	0, any other value
MeanMembraneHScore75thPerc	None	Below/above 57.5
MeanCytoHScoreMedianSplit	None	Below/above 58.33
MeanCytoHScoreTertilesSplit	None	0<35, 35<80, 80<200
MeanNuclearHscorePosNeg	None	0, any other value
MeanStromaHScoreMedialSplit	None	Below/above 47.5
MeanStromaHScoreTerts	None	0<35, 35<60, 60<300
NuclearAverageHScore	None	Below/above 12.6
LumMemHS75th	None	Below/above 85
CentralMemHS75th	None	Below/above 75
AEMemHS75th	None	Below/above 75
Other variables		
AvTumEpiMedian	None	Below/above 53.33
AverageStromaMedianSplit	None	Below/above 47.5
cMycCD44combined	None	See "CD44 and c-Myc interaction"
*continuous variables were split into categorical variables using cut-offs in the right-hand column of this table to allow analyses that could not be performed using continuous data		

In order to assess the distribution of the data, histograms were prepared using analyze>frequencies. The mean, median and mode were computed, as were the quartile and tertile splits of the data. Distribution of the data was assessed by eye, and correlation between the average H-score and individual tumour compartments or normal colon was assessed using Spearman's or Pearson's rank correlation (analyze>correlation>bivariate). The average scores of each cellular compartment were also assessed to see if increased expression in one compartment led to increased expression in another, using correlation analysis. Chi-squared analysis (analyze>descriptive statistics>crosstabs) was used, with adjusted residuals, to determine the relationship between the clinical variables and biomarker expression. Mann-Whitney U tests were used to assess the relationship between the biomarkers and tumour stroma content.

Kaplan-Meier and Cox regression methods were used to determine the association of each biomarker with survival. Kaplan-Meier analysis was performed using 3 year survival, 5 year survival and death by colorectal cancer. Cox regression analysis was performed using biomarker status, and with multivariate analysis to determine if survival was truly associated with biomarker expression or if clinical variables were affecting the results.

Chi-square analysis was repeated for each biomarker with the additional splitting of MMR status added, in the most biologically relevant cellular

compartments. This allowed for any confounding results from MMR status, which influences CRC outcomes, to be accounted for.

Tumour epithelium content analysis

In order to achieve the most representative figures to allow comparison of tumour epithelium content with the clinical variables, the mean tumour epithelium content was calculated in Microsoft Excel as follows:

((luminal (CD44) x luminal (c-Myc))/2) x ((central (CD44) x central (c-Myc))/2) x ((advancing edge (CD44) x advancing edge (c-Myc))/2))/3

Where one value in one region of interest was not available, the value from the other slide would be used. Where both values were not available for a tumour region of interest, the mean would be calculated from the other regions of interest using the following formulae:

Calculating average from individual region of interest:

=IF(A2,IF(F2,AVERAGE(A2,F2),A2)IF(F2,F2,""))

Calculating average from multiple regions of interest:

=AVERAGE(K2:M2)

This prevented unequal weighting of values and gave a tumour epithelium content which was most representative of the true overall tumour content.

Bioinformatic analysis

The proteins used as biomarkers in this project may be associated with different pathways and known pathogenesis. By testing their association with other genes in a known dataset (the Cancer Genome Atlas (TCGA)), it should enable us to understand more about their roles in colorectal cancer. TCGA CRC differential expression data was provided by Nigel Mongan. The base mean expression, log2 fold change data, log fold change standard deviations, statistic, p values and adjusted p values were provided for 60,488 genes, labelled with Ensembl gene IDs. To look for biologically significant changes to gene expression, results were filtered for genes with:

- More than ±1 log2 fold change (to shown truly biologically relevant changes)
- p value below 0.05
- adjusted p value below 0.05

To interrogate the data for biologically important changes, the Ensembl gene IDs were inputted into WebGestalt [81]. Here, the pathways c-Myc and CD44 are involved in were interrogated in KEGG, Panther and Reactome. Other analyses investigating chromosomal locations and gene ontology were alsp performed (data not shown).

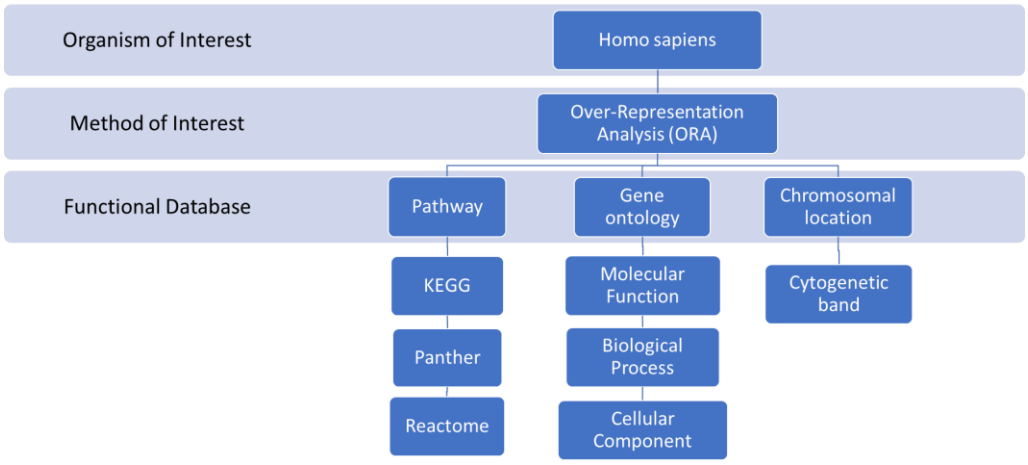


Figure 6: Parameters used to interrogate differentially expressed genes associated with c-Myc or CD44 in WebGestalt. Pathways were analysed to understand and confirm which pathways were enriched when CD44 and c-Myc were over-expressed. Gene ontology and chromosomal were interrogated to understand what kind of relationships that these genes of interest had with those that were over-expressed in their presence.

Results

Patient cohort

1000 patients were included in this analysis. The clinical features of these patients are described below. Over half (56.8%) were male, and the majority were over the age of 50 (93.7%). Site was re-coded into two variables, “left” and “right”. Left-sided tumours included “left” and “rectum” tumours, and right-sided tumours included “right” and “other” tumours. The location of tumours was split almost equally, with 49% of tumours being right-sided. 72.7% of patients survived for three years, but less than half (43%) survived to five years. 28.7% of patients experienced recurrence, and 11.9% of patients experienced metastasis. Most patients were in stage 2 (40.2%) or stage 3 (31.9%) at the end of the study. A higher primary tumour stage was generally observed, with only 7.4% of patients at T1. Lymph node metastasis was present in less than half of patients (N0 57%). The majority of patients were experiencing grade 2 colorectal cancer (88.6%), meaning the tumours were moderately differentiated [82]. The grades were re-coded from 1 (well-differentiated), 2 (moderately-differentiated) and 3 (poorly-differentiated) to (1 & 2) and 3, in order to classify differentiated and undifferentiated tumours more plainly. Vascular invasion was more or less equally split between absent or present (48.3%), and the presence of perineural invasion (18.3%) and intratumoural lymphovascular invasion (28.7%) were observed less often than not. Not all tumours were classified for tumour edge, but of those that were, 39.5% were classified as a “pushing” tumour edge, meaning that the tumour grows outward in a uniform pattern, and 36.7% were classified as “infiltrative”, meaning that the tumour grows into the surrounding tissue in a diffuse pattern [83]. 2.2% of patients could not be classified for mismatch repair status, but most patients were proficient (81.8%) and just 16% were deficient.

Overall the frequencies of the clinical features match the published data showing that there is no bias in the cohort. There was slightly higher occurrence of primary tumour T1 disease than may be expected from an historical cohort but this is probably because the NHS Bowel Cancer Screening Program was introduced over the period of time of this cohort (2008-2014).

Table 3: Patient stratification

Clinical Variable	Condition	Number (%)
Gender	Female	432 (43.2)
	Male	568 (56.8)
3 year Survival	Alive	727 (72.7)
	Dead	272 (27.2)
5 year Survival	Alive	430 (43)
	Dead	570 (57)
Synchronous metastasis	No metastasis	881 (88.1)
	With metastasis	119 (11.9)
Recurrence	No Recurrence	713 (71.3)
	Recurrence	287 (28.7)
Cause of death	Alive	624 (62.4)
	CRC	178 (17.8)
	Other	118 (11.8)
	Unknown	80 (8)
Survival Status	Alive	624 (62.4)
	Dead	376 (37.6)
Site	Right	461 (46.1)
	Left	363 (36.3)
	Rectum	147 (14.7)
	Other	29 (2.9)
Final stage	1	161 (16.1)
	2	402 (40.2)
	3	319 (31.9)
	4	118 (11.8)
Primary tumour	T1	74 (7.4)
	T2	106 (10.6)
	T3	526 (52.6)
	T4	294 (29.4)
Lymph node metastasis	N0	570 (57)
	N1	243 (24.3)
	N2	161 (16.1)
Metastasis	No metastasis	881 (88.1)
	With metastasis	119 (11.9)
Vascular invasion	Absent	502 (50.2)
	Present	483 (48.3)
Grade	1	20 (2)

Clinical Variable	Condition	Number (%)
	2	886 (88.6)
	3	93 (9.3)
Perineural invasion	Absent	783 (78.3)
	Present	183 (18.3)
Intramural lymphovascular invasion	Absent	641 (64.1)
	Present	287 (28.7)
Tumour edge	Infiltrative	395 (39.5)
	Pushing	367 (36.7)
Tumour budding score	Low	508 (50.8)
	High	243 (24.3)
Peritumoural lymphocytes	Inconspicuous	549 (54.9)
	Conspicuous	217 (21.7)
MMR Status	Unknown	22 (2.2)
	Proficient	818 (81.8)
	Deficient	160 (16)
Age dichotomised	Young	63 (6.3)
	Old	937 (93.7)
Grade dichotomised	Low	906 (90.6)
	High	93 (9.3)
Site dichotomised	Right side	490 (49)
	Left side	510 (51)

Statistical analysis of c-Myc

Average c-Myc expression for nuclear and stromal staining was calculated by taking the mean of the H-scores of each region of the tumour (luminal, central and advancing edge) and cytoplasmic staining was calculated as percentage of positive cells. To ensure the maximum number of data points, means were calculated where possible, and where only one location was present, this measurement was included in the results.

Nuclear staining

All distributions of nuclear c-Myc staining showed a skew to the right as seen in figure 7, due to the high frequency of zeroes for c-Myc expression. Therefore, medians and interquartile ranges (IQR) were presented. The median of the nuclear average H-score was 54 (IQR 12.5-102.5), which did not differ greatly from the medians of the luminal, central and advancing edge nuclear H-score medians (55 (IQR 10-110), 53 (IQR 10-100) and 50 (IQR 10-105) respectively). The correlation of each tumour region with nuclear average H-score was assessed using Spearman's rank correlation, which shows that the tumour regions were very strongly positively correlated (>0.9) with the average nuclear H-score ($p < 0.001$).

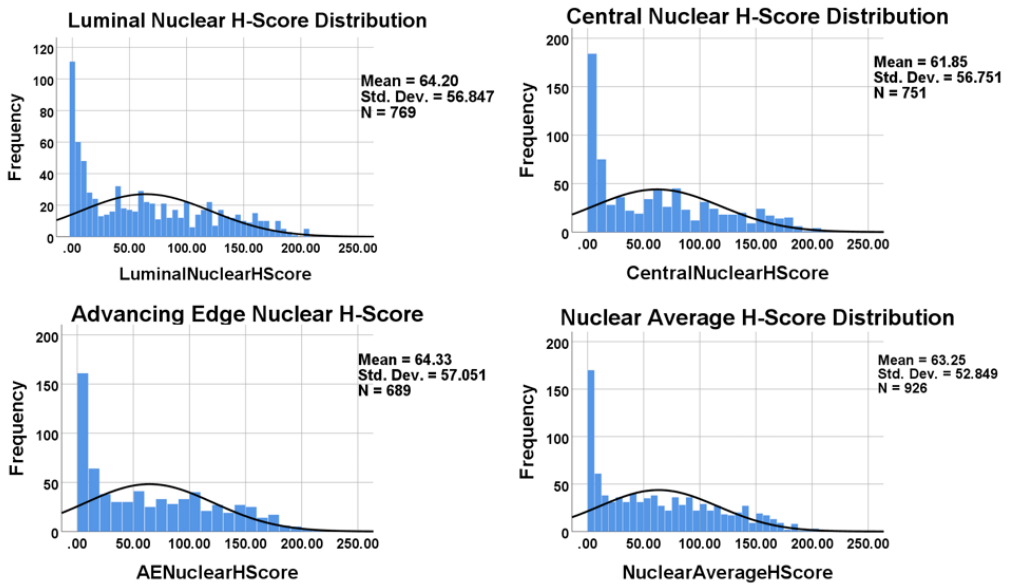


Figure 7: Distribution of nuclear H-scores. All nuclear H-scores were right-skewed, and average nuclear H-score was representative of the individual tumour regions.

Associations between c-Myc nuclear staining and clinical variables

The primary functions of c-Myc are performed in the nucleus and this is where the most biologically relevant staining was expected to be observed. Nuclear average H-score was re-coded into new negative and positive expression variables by the median (54) cut-off. Due to the skewed distribution of the data, additional splits were also performed in order to increase the number of significant observations that could be made. These included nuclear lowest tertile, nuclear highest tertile, and nuclear positive. To establish which variables may be linked to expression of c-Myc, Chi-squared analysis was performed with the clinical variables described in table 4.

When the data was split by the median expression, the variables 5 year survival and site dichotomised were significantly associated with nuclear average H-score. 5 year survival was negatively associated with c-Myc expression ($p < 0.001$), and patients with positive c-Myc expression were more likely to have died after 5 years. Tumour location was also significantly associated with c-Myc expression. Negative c-Myc expression showed an association with right-sided tumours (p 0.03).

When the data was dichotomised by the lowest tertile (above and below 26.33) poor 5 year survival was significantly associated with positive c-Myc expression ($p < 0.001$), as seen with the median split. The presence of intratumoural lymphovascular invasion was associated with positive expression of c-Myc (p 0.015). Splitting the data by highest tertile (above and below 85.33) once again showed a negative association between c-Myc expression and survival ($p < 0.001$), with positive expression of c-Myc being indicative of death. Primary tumour stage showed a negative association with c-Myc expression (p 0.006),

with positive c-Myc expression associated with T1-3. T4 was associated with negative nuclear c-Myc expression.

Table 4: Chi-square analysis of nuclear c-Myc positive and negative samples							
Clinical variable	**Condition**	**Negative**	**Positive**	**Total**	**Adjusted residual**		**X^2 (p value)**
3 year Survival	Alive	353 (76.2)	329 (71.2)	682 (73.7)	1.7	-1.7	3.020 (0.082)
	Dead	110 (23.8)	133 (28.8)	243 (26.3)	-1.7	1.7	
	Total	463 (100)	462 (100)	925 (100)			
5 year Survival	Alive	240 (51.7)	166 (35.9)	406 (43.8)	4.8	-4.8	**23.45 (<0.001)**
	Dead	224 (48.3)	296 (64.1)	520 (56.2)	-4.8	4.8	
	Total	464 (100)	462 (100)	926 (100)			
Synchronous met	No metastasis	412 (88.8)	410 (88.7)	822 (88.8)	0	0	0.001 (0.981)
	With metastasis	52 (11.2)	52 (11.3)	104 (11.2)	0	0	
	Total	464 (100)	462 (100)	926 (100)			
Recurrence	No Recurrence	336 (72.4)	327 (70.8)	663 (71.6)	0.6	-0.6	0.304 (0.581)
	Recurrence	128 (27.6)	135 (29.2)	263 (28.4)	-0.6	0.6	
	Total	464 (100)	462 (100)	926 (100)			
Cause of death	Alive	300 (64.7)	283 (61.3)	583 (63)	1.1	-1.1	6.899 (0.075)
	CRC	69 (14.9)	92 (19.9)	161 (17.4)	-2	2	
	Other	53 (11.4)	59 (12.8)	112 (12.1)	-0.6	0.6	
	Unknown	42 (9.1)	28 (6.1)	70 (7.6)	1.7	-1.7	
	Total	464 (100)	462 (100)	926 (100)			
Survival Status	Alive	300 (64.7)	283 (61.3)	583 (63)	1.1	-1.1	1.147 (0.284)
	Dead	164 (35.3)	179 (38.7)	343 (37)	-1.1	1.1	
	Total	464 (100)	462 (100)	926 (100)			
Final stage	1	70 (15.1)	83 (18)	153 (16.5)	-1.2	1.2	2.610 (0.456)
	2	200 (43.1)	178 (38.5)	378 (40.8)	1.4	-1.4	
	3	142 (30.6)	150 (32.5)	292 (31.5)	-0.6	0.6	
	4	52 (11.2)	51 (11)	103 (11.1)	0.1	-0.1	
	Total	464 (100)	462 (100)	926 (100)			
Primary tumour	T1	26 (5.6)	44 (9.5)	70 (7.6)	-2.3	2.3	6.934 (0.074)
	T2	53 (11.4)	47 (10.2)	100 (10.8)	0.6	-0.6	
	T3	241 (51.9)	249 (53.9)	490 (52.9)	-0.6	0.6	
	T4	144 (31)	122 (26.4)	266 (28.7)	1.6	-1.6	
	Total	464 (100)	462 (100)	926 (100)			
Lymph node metastasis	N0	274 (60.2)	261 (58.7)	535 (59.4)	0.5	-0.5	1.420 (0.492)
	N1	105 (23.1)	117 (26.3)	222 (24.7)	-1.1	1.1	
	N2	76 (16.7)	67 (15.1)	143 (15.9)	0.7	-0.7	
	Total	455 (100)	445 (100)	900 (100)			
Metastasis	No metastasis	412 (88.8)	410 (88.7)	822 (88.8)	0	0	0.001 (0.981)

	With metastasis	52 (11.2)	52 (11.3)	104 (11.2)	0	0	
	Total	464 (100)	462 (100)	926 (100)			
Clinical variable	**Condition**	**Negative**	**Positive**	**Total**	**Adjusted residual**		**X^2 (p value)**
Vascular invasion	Absent	228 (49.6)	244 (54.1)	472 (51.8)	-1.4	1.4	1.877 (0.171)
	Present	232 (50.4)	207 (45.9)	439 (48.2)	1.4	-1.4	
	Total	460 (100)	451 (100)	911 (100)			
Grade	1	9 (1.9)	9 (2)	18 (1.9)	0	0	0.574 (0.751)
	2	421 (90.7)	412 (89.4)	833 (90.1)	0.7	-0.7	
	3	34 (7.3)	40 (8.7)	74 (8)	-0.8	0.8	
	Total	464 (100)	461 (100)	925 (100)			
Perineural invasion	Absent	364 (80)	360 (82.2)	724 (81.1)	-0.8	0.8	0.699 (0.403)
	Present	91 (20)	78 (17.8)	169 (18.9)	0.8	-0.8	
	Total	455 (100)	438 (100)	893 (100)			
Intramural lymphovascular invasion	Absent	297 (71.7)	301 (66.6)	598 (69.1)	1.6	-1.6	2.678 (0.102)
	Present	117 (28.3)	151 (33.4)	268 (30.9)	-1.6	1.6	
	Total	414 (100)	452 (100)	866 (100)			
Tumour edge	Infiltrative	155 (53.3)	210 (50.2)	365 (51.5)	0.8	-0.8	0.629 (0.428)
	Pushing	136 (46.7)	208 (49.8)	344 (48.5)	-0.8	0.8	
	Total	291 (100)	418 (100)	709 (100)			
Tumour budding score	Low	197 (69.9)	272 (65.1)	469 (67)	1.3	-1.3	1.745 (0.187)
	High	85 (30.1)	146 (34.9)	231 (33)	-1.3	1.3	
	Total	282 (100)	418 (100)	700 (100)			
Peritumoural lymphocytes	Inconspicuous	217 (73.8)	290 (69)	507 (71)	1.4	-1.4	1.905 (0.168)
	Conspicuous	77 (26.2)	130 (31)	207 (29)	-1.4	1.4	
	Total	294 (100)	420 (100)	714 (100)			
MMR Status		3 (0.6)	3 (0.6)	6 (0.6)	0	0	4.763 (0.092)
	Proficient	399 (86)	373 (80.7)	772 (83.4)	2.1	-2.1	
	Deficient	62 (13.4)	86 (18.6)	148 (16)	-2.2	2.2	
	Total	464 (100)	462 (100)	926 (100)			
Age dichotomised	Young	25 (5.4)	31 (6.7)	56 (6)	-0.8	0.8	0.712 (0.399)
	Old	439 (94.6)	431 (93.3)	870 (94)	0.8	-0.8	
	Total	464 (100)	462 (100)	926 (100)			
Grade dichotomised	Low	430 (92.7)	421 (91.3)	851 (92)	0.8	-0.8	0.572 (0.449)
	High	34 (7.3)	40 (8.7)	74 (8)	-0.8	0.8	
	Total	464 (100)	461 (100)	925 (100)			
Site dichotomised	Right side	242 (52.2)	208 (45)	450 (48.6)	2.2	-2.2	**4.716 (0.03)**
	Left side	222 (47.8)	254 (55)	476 (51.4)	-2.2	2.2	
	Total	464 (100)	462 (100)	926 (100)			
**Significant p values (<0.05) are highlighted in bold*							

Mismatch repair status

Patients who were c-Myc positive showed a higher association with dMMR than pMMR. c-Myc expression has been known to interact with mismatch repair status [48, 49]. To ensure that MMR status was not confounding the results, the Chi-square analysis was repeated, with the data split into pMMR and dMMR. Some samples did not have MMR status available (data not shown). Repetition of the analysis with MMR status yielded additional results. 5 year survival was significant as before, but the split demonstrated that patients who had negative nuclear c-Myc expression and were pMMR, were statistically more likely to survive ($p<0.001$) than those who were dMMR (p 0.033). 3 year survival in MMR deficient patients was significantly higher in c-Myc positive patients (p 0.039) but not in MMR proficient patients (p 0.426). Lower primary tumour stage was associated with positive c-Myc expression (p 0.019), and right-sided tumours were associated with low expression of nuclear c-Myc (p 0.003). These results were previously masked in the general analysis.

As c-Myc is believed to be regulated by MLH1 and MSH2, the relationship between c-Myc expression and the status of these two genes was investigated. Median expression of c-Myc in proficient MLH1 (pMLH1) patients was 50 (interquartile range 89.67) and the median expression of MLH1 deficient (dMLH1) patients was 70.8 (interquartile range 80.2). Mann-Whitney U test was used to assess if the difference in expression of c-Myc between pMLH1 and dMLH1 patients was significant. dMLH1 patients were found to have higher expression of c-Myc than pMLH1 patients (p 0.012). MSH2 has been reported to be regulated by the c-Myc MAX complex, and so the nuclear average H-score for MSH2 proficient (pMSH2) and deficient (dMSH2) patients were compared. As only 9 patients were dMSH2, these results may not be truly representative of the population. The median expression of pMSH2 patients was 52.59 (interquartile range 88.88) and the median expression of dMSH2 patients was 75 (interquartile range 82.5). Mann-Whitney U analysis revealed that there was no significant difference between the two groups (p 0.731).

Survival analysis of c-Myc expression

In the Chi-squared analysis above, interestingly, 3 year survival was not significantly different between the two groups (p 0.082) whereas 5 year survival was ($p <0.001$). To understand the importance of c-Myc expression on overall survival, Kaplan-Meier survival analysis was performed as seen in figure 8. Nuclear average H-scores split by median expression and tertiles were compared with 3 year survival and 5 year survival. When the data was split by median expression, 3 year survival was not significantly different (p 0.099) but 5 year survival ($p <0.001$) and death relating to CRC (p 0.006) were. After 5 years, c-Myc negative patients had a 51.7% chance of survival, whereas c-Myc positive patients had a 35.9% chance of survival. When death was related to colorectal cancer, there was an approximately 86% chance of survival in the c-Myc negative group after 5 years, whereas the positive group had a worse

chance of survival of approximately 77%. Median survival could not be calculated as most cases either survived or death was not related to CRC. When the data was divided into three tertiles, 5 year survival was significant (p 0.001). The lowest tertile had a 53.2% chance of survival, whereas the middle tertile had a 42.7% chance and the highest expressing group had a 35.6% chance of survival.

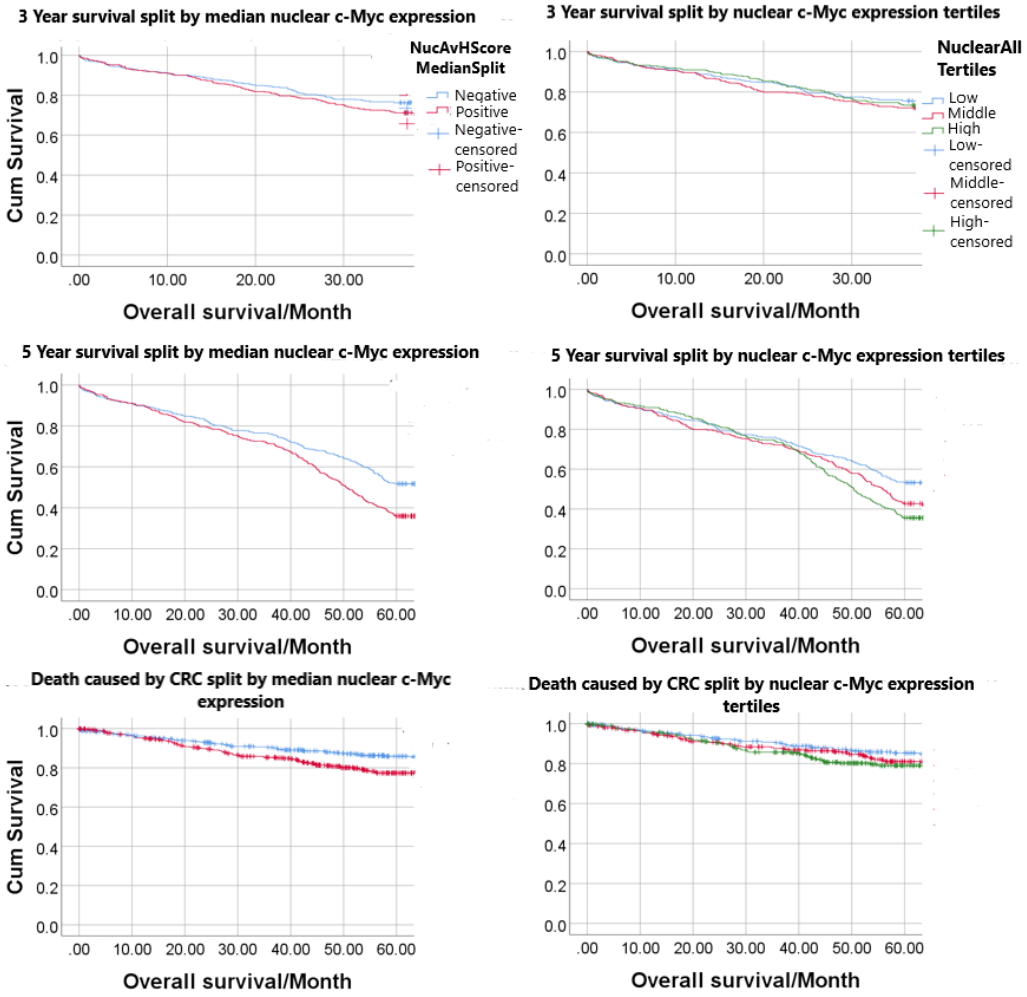


Figure 8: Kaplan-Meier analysis of nuclear c-Myc expression split by median (left) and tertiles (right). 5 year survival was significantly associated with nuclear c-Myc expression (middle), as was death relating to CRC (bottom). 3 year survival was not significantly associated with nuclear c-Myc (top).

In order to further understand the relationship between c-Myc nuclear expression and survival, Cox regression analysis was performed using c-Myc nuclear average median split, in a multivariate analysis with the variables that c-Myc had shown significant association with. This was to determine whether c-Myc expression was truly related to survival or whether one of the clinical variables was confounding the analysis. Primary tumour stage, lymph node metastasis, grade, intratumoural lymphovascular invasion, perineural invasion, tumour edge, grade dichotomised and site dichotomised variables were assessed alongside c-Myc nuclear expression. Death related to colorectal cancer was assessed. The hazard ratio of nuclear average H-score in univariate

analysis revealed that death was 1.346 times more likely in c-Myc positive patients (p 0.007). Multivariate analysis revealed an even stronger difference (hazard ratio 1.6, p 0.001). In addition to this, primary tumour size, lymph node metastasis, grade and perineural invasion were significantly associated with survival. T3 tumours were significantly associated with death relating to CRC (p 0.001) as were N2 tumours (metastasis to >6 nodes) (p 0.003). Presence of perineural invasion (p 0.047) and pushing tumour edge (p 0.039) were indicative of increased risk of death by CRC. Intratumoural lymphovascular invasion and site dichotomised were not significantly associated with survival status.

Table 5: Cox regression analysis of c-Myc nuclear expression and significantly associated variables

Variable	**Hazard ratio**	**95.0% CI**		**p value**
		Lower	**Upper**	
Nuc Average H-Score Median Split	1.346	1.084	1.67	**0.007**
Clinical variable	**Hazard ratio**	**95.0% CI**		
		Lower	**Upper**	**P value**
Primary tumour 1T1*	-	-	-	**0.000**
Primary tumour T2	1.451	0.397	5.304	0.573
Primary tumour T3	2.607	0.817	8.325	0.106
Primary tumour T4	7.077	2.194	22.830	**0.001**
Lymph node metastasisN0	-	-	-	**0.005**
Lymph node metastasisN1	1.050	0.763	1.444	0.765
Lymph node metastasisN2	1.695	1.202	2.391	**0.003**
Grade 1	-	-	-	**0.003**
Grade2	0.756	0.235	2.429	0.638
Grade 3	1.486	0.440	5.019	0.524
Perineural invasion	1.384	1.004	1.909	**0.047**
Intramural lymphovascular invasion	0.765	0.572	1.022	0.070
Tumour edge	0.729	0.539	0.984	**0.039**
Site dichotomised	0.883	0.679	1.149	0.354
Grade dichotomised**				-
Nuclear Average H-Score Median Split	1.600	1.226	2.087	**0.001**

**Significant associations (p <0.05) highlighted in bold*
***Could not be determined*

Cytoplasmic expression of c-Myc

Luminal, central and advancing edge cytoplasmic percentage positive cells were averaged to produce the cytoplasmic average positive score. The data

was skewed to the right for all cytoplasmic data including normal samples, as seen in figure 9. Median expression of average cytoplasmic staining was 2 (IQR 0-15). Luminal, central and advancing edge medians were all 0, with IQRs of 0-15, 0-10 and 0-10 respectively. Spearman's rank correlation was used to analyse the similarity of the average data to the individual tumour samples. Luminal, central and advancing edge samples were highly positively correlated with the averaged cytoplasmic data (>0.75, p<0.001). Average cytoplasmic percentage positive was split by positive/negative expression due to low median expression.

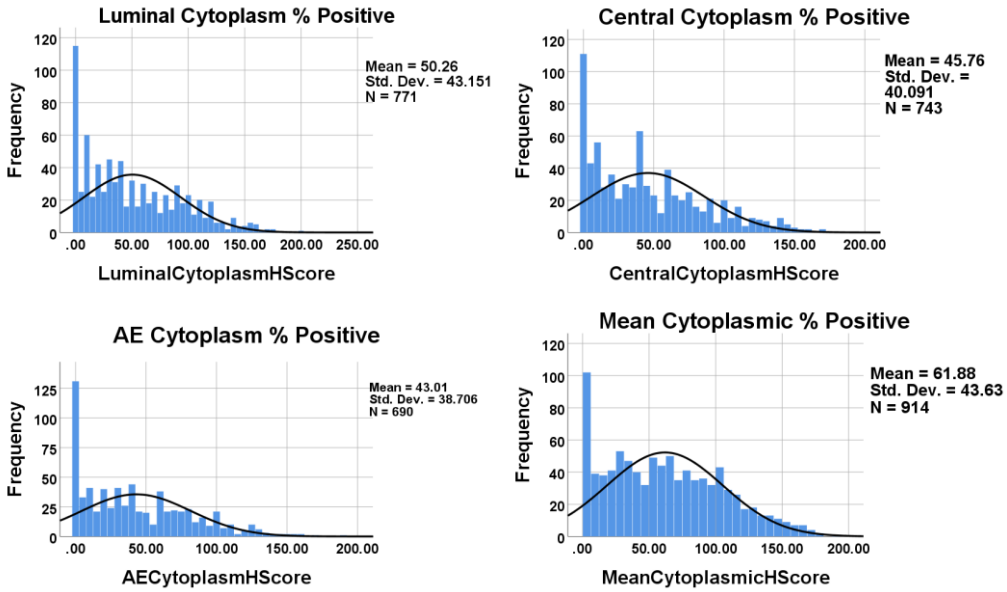


Figure 9: Distribution of percentage positive cytoplasmic staining for luminal, central, advancing edge and average tumours.

Association with clinical variables

Chi-squared analysis was performed as for nuclear H-score data. When the data was split by negative versus positive cytoplasmic c-Myc expression, increased lymph node metastasis was associated with negative cytoplasmic c-Myc expression (p 0.003). The presence of perineural invasion (p 0.024) was again associated with negative cytoplasmic c-Myc. In addition, site dichotomised was also significantly associated with cytoplasmic c-Myc expression (p 0.026). Tumours that originated in the left side of the colon were associated with positive cytoplasmic c-Myc expression.

Table 6: Chi-squared analysis of cytoplasmic expression of c-Myc expression against clinical variables							
Clinical variable	**Condition**	**Negative**	**Positive**	**Total**	**Adjusted residual**		**X^2 (p value)**
Gender	Female	152 (42.8)	253 (44.3)	405 (43.7)	-0.4	0.4	0.198 (0.656)
	Male	203 (57.2)	318 (55.7)	521 (56.3)	0.4	-0.4	
	Total	355 (100)	571 (100)	926 (100)			
3 year Survival	Alive	258 (72.7)	424 (74.4)	682 (73.7)	-0.6	0.6	0.330 (0.566)
	Dead	97 (27.3)	146 (25.6)	243 (26.3)	0.6	-0.6	
	Total	355 (100)	570 (100)	925 (100)			
5 year Survival	Alive	146 (41.1)	260 (45.5)	406 (43.8)	-1.3	1.3	1.727 (0.189)
	Dead	209 (58.9)	311 (54.5)	520 (56.2)	1.3	-1.3	
	Total	355 (100)	571 (100)	926 (100)			
Synchronous metastasis	No metastasis	318 (89.6)	504 (88.3)	822 (88.8)	0.6	-0.6	0.378 (0.539)
	With metastasis	37 (10.4)	67 (11.7)	104 (11.2)	-0.6	0.6	
	Total	355 (100)	571 (100)	926 (100)			
Recurrence	No Recurrence	253 (71.3)	410 (71.8)	663 (71.6)	-0.2	0.2	0.031 (0.86)
	Recurrence	102 (28.7)	161 (28.2)	263 (28.4)	0.2	-0.2	
	Total	355 (100)	571 (100)	926 (100)			
Cause of death	Alive	216 (60.8)	367 (64.3)	583 (63)	-1.1	1.1	2.336 (0.506)
	CRC	65 (18.3)	96 (16.8)	161 (17.4)	0.6	-0.6	
	Other	42 (11.8)	70 (12.3)	112 (12.1)	-0.2	0.2	
	Unknown	32 (9)	38 (6.7)	70 (7.6)	1.3	-1.3	
	Total	355 (100)	571 (100)	926 (100)			
Survival Status	Alive	216 (60.8)	367 (64.3)	583 (63)	-1.1	1.1	1.103 (0.294)
	Dead	139 (39.2)	204 (35.7)	343 (37)	1.1	-1.1	
	Total	355 (100)	571 (100)	926 (100)			
Final stage	1	55 (15.5)	98 (17.2)	153 (16.5)	-0.7	0.7	3.121 (0.373)
	2	139 (39.2)	239 (41.9)	378 (40.8)	-0.8	0.8	
	3	124 (34.9)	168 (29.4)	292 (31.5)	1.8	-1.8	
	4	37 (10.4)	66 (11.6)	103 (11.1)	-0.5	0.5	
	Total	355 (100)	571 (100)	926 (100)			
Primary tumour	T1	26 (7.3)	44 (7.7)	70 (7.6)	-0.2	0.2	0.704 (0.872)
	T2	36 (10.1)	64 (11.2)	100 (10.8)	-0.5	0.5	
	T3	186 (52.4)	304 (53.2)	490 (52.9)	-0.3	0.3	
	T4	107 (30.1)	159 (27.8)	266 (28.7)	0.8	-0.8	
	Total	355 (100)	571 (100)	926 (100)			
Lymph node metastasis*	N0	192 (55.5)	343 (61.9)	535 (59.4)	-1.9	1.9	**11.43 (0.003)**
	N1	81 (23.4)	141 (25.5)	222 (24.7)	-0.7	0.7	
	N2	73 (21.1)	70 (12.6)	143 (15.9)	3.4	-3.4	
	Total	346 (100)	554 (100)	900 (100)			

Metastasis	No metastasis	318 (89.6)	504 (88.3)	822 (88.8)	0.6	-0.6	0.378 (0.539)
	With metastasis	37 (10.4)	67 (11.7)	104 (11.2)	-0.6	0.6	
	Total	355 (100)	571 (100)	926 (100)			
Clinical variable	**Condition**	**Negative**	**Positive**	**Total**	**Adjusted residual**		**X^2 (p value)**
Vascular invasion	Absent	173 (49.4)	299 (53.3)	472 (51.8)	-1.1	1.1	1.292 (0.256)
	Present	177 (50.6)	262 (46.7)	439 (48.2)	1.1	-1.1	
	Total	350 (100)	561 (100)	911 (100)			
Perineural invasion	Absent	266 (77.3)	458 (83.4)	724 (81.1)	-2.3	2.3	**5.127 (0.024)**
	Present	78 (22.7)	91 (16.6)	169 (18.9)	2.3	-2.3	
	Total	344 (100)	549 (100)	893 (100)			
Intramural lymphovascular invasion	Absent	226 (70.2)	372 (68.4)	598 (69.1)	0.6	-0.6	0.308 (0.579)
	Present	96 (29.8)	172 (31.6)	268 (30.9)	-0.6	0.6	
	Total	322 (100)	544 (100)	866 (100)			
Tumour edge	Infiltrative	115 (51.3)	250 (51.5)	365 (51.5)	-0.1	0.1	0.003 (0.959)
	Pushing	109 (48.7)	235 (48.5)	344 (48.5)	0.1	-0.1	
	Total	224 (100)	485 (100)	709 (100)			
Tumour budding score	Low	141 (64.4)	328 (68.2)	469 (67)	-1	1	0.987 (0.321)
	High	78 (35.6)	153 (31.8)	231 (33)	1	-1	
	Total	219 (100)	481 (100)	700 (100)			
Peritumoural lymphocytes	Inconspicuous	158 (69)	349 (72)	507 (71)	-0.8	0.8	0.663 (0.415)
	Conspicuous	71 (31)	136 (28)	207 (29)	0.8	-0.8	
	Total	229 (100)	485 (100)	714 (100)			
MMR Status		2 (0.6)	4 (0.7)	6 (0.6)	-0.3	0.3	0.066 (0.968)
	Proficient	296 (83.4)	476 (83.4)	772 (83.4)	0	0	
	Deficient	57 (16.1)	91 (15.9)	148 (16)	0	0	
	Total	355 (100)	571 (100)	926 (100)			
Age dichotomised	Young	16 (4.5)	40 (7)	56 (6)	-1.6	1.6	2.405 (0.121)
	Old	339 (95.5)	531 (93)	870 (94)	1.6	-1.6	
	Total	355 (100)	571 (100)	926 (100)			
Grade dichotomised	Low	321 (90.4)	530 (93)	851 (92)	-1.4	1.4	1.948 (0.163)
	High	34 (9.6)	40 (7)	74 (8)	1.4	-1.4	
	Total	355 (100)	570 (100)	925 (100)			
Site dichotomised	Right side	189 (53.2)	261 (45.7)	450 (48.6)	2.2	-2.2	**4.969 (0.026)**
	Left side	166 (46.8)	310 (54.3)	476 (51.4)	-2.2	2.2	
	Total	355 (100)	571 (100)	926 (100)			

**Significant p values (<0.05) highlighted in bold*

Survival analysis of cytoplasmic c-Myc expression

3 year survival, 5 year survival and death related to colorectal cancer were assessed with cytoplasmic c-Myc staining. No significant difference could be observed between survival of positive or negative cytoplasmic c-Myc expressors, as seen in the Kaplan-Meier diagrams in figure 10. Cox regression analysis was done to assess whether 5 year survival stratified by cytoplasmic staining was affected by other variables. Perineural invasion, lymph node metastasis and site dichotomised were analysed with cytoplasmic staining. In the multivariate analysis, increased lymph node invasion led to increased survival (p <0.001), whereas the presence of perineural invasion was associated with increased risk of death (p 0.007), as were tumours that originated in the right side of the colon (p 0.026). Cytoplasmic staining was not significantly associated with survival outcome either in univariate or multivariate analysis.

Table 7: Cox regression analysis of c-Myc cytoplasmic staining with significantly associated variables

Variable	**Hazard ratio**	**95.0% CI**		**p value**
		Lower	**Upper**	
Cytoplasmic average Positive only	0.937	0.755	1.164	0.558
Clinical variable	**Hazard ratio**	**95.0% CI for Hazard ratio**		**p value**
		Lower	**Upper**	
Lymph node metastasis N0*	-			**<0.001**
Lymph node metastasis N1	0.459	0.361	0.585	**<0.001**
Lymph node metastasis N2	0.568	0.438	0.737	**<0.001**
Perineural invasion	1.363	1.089	1.705	**0.007**
Site dichotomised	0.816	0683	0.975	**0.026**
Cytoplasmic average positive only	1.020	0.850	1.223	0.832

**Significant associations highlighted in bold*

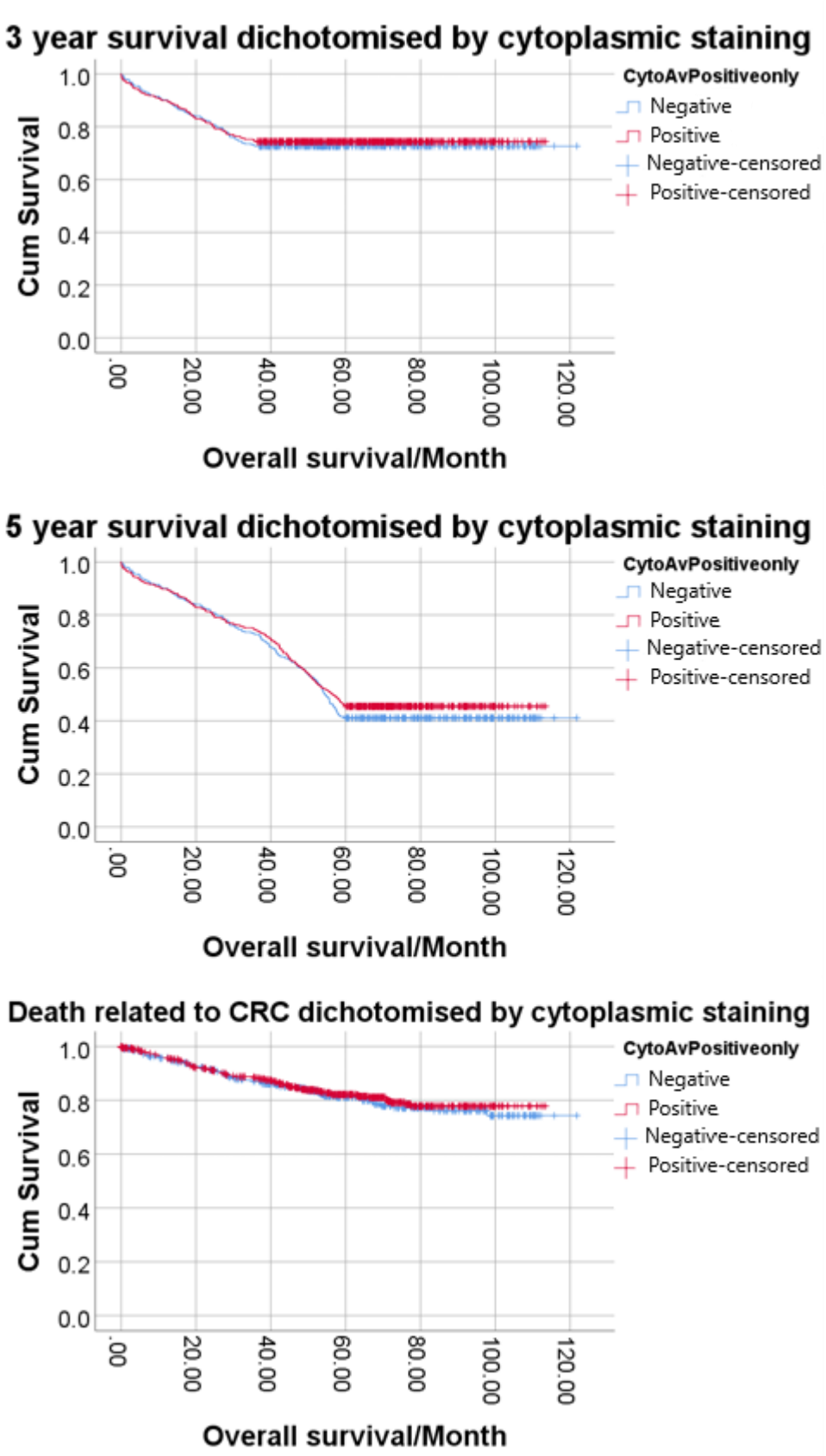


Figure 10: Kaplan-Meier analysis of survival after 3 and 5 years, and death relating to colorectal cancer, stratified by c-Myc cytoplasmic average percentage positive cells.

Stromal expression of c-Myc

The stromal H-scores were not normally distributed, as seen below in figure 11, and the median average expression of c-Myc in the stroma was 6.67 (IQR 2.5-13.5). Luminal, central and advancing edge medians were all 5 (IQR 2-14, 1-13 and 1-13.25 respectively). The luminal, central and advancing edge H-scores were strongly positively correlated with the average H-score (>0.75, p <0.001), demonstrating homogeneity of the staining.

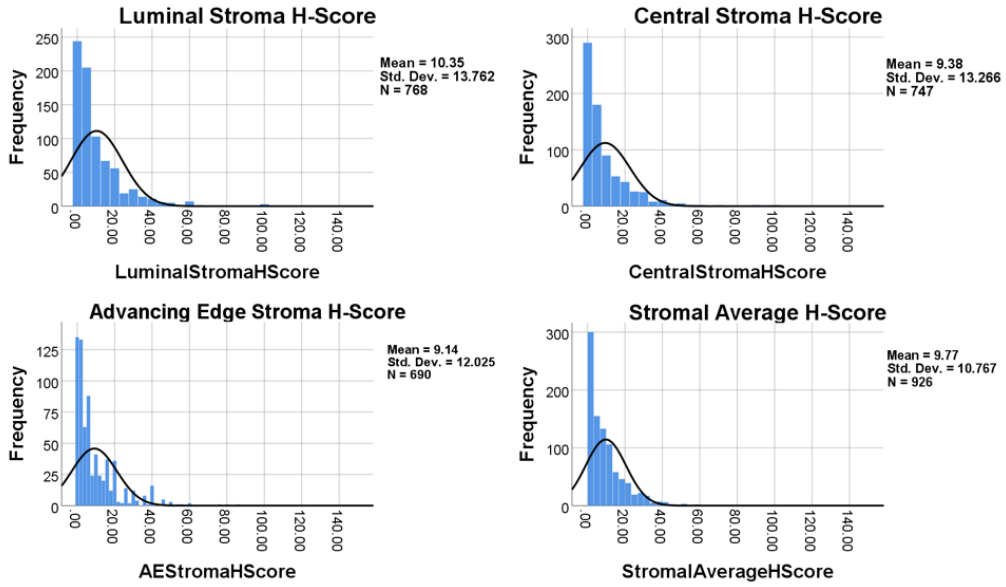


Figure 11: Distribution of c-Myc expression in the stroma. The data was not normally distributed, but was highly correlated between tumour regions and average H-score.

Stromal c-Myc H-score Chi-squared analysis

Chi-square analysis was performed with the clinical variables against the c-Myc stromal average H-score split by median (6.67) and by lowest and highest tertiles (data not shown). 3 year survival (p 0.012) and survival status (p 0.002) were associated with positive c-Myc stromal expression. Positive c-Myc stromal expression was associated with lower final stage (p 0.007), primary tumour stage (p 0.001) and lymph node metastasis (p 0.019). The absence of vascular invasion was associated with positive stromal c-Myc expression (p 0.018), as were conspicuous peritumoural lymphocytes (p <0.001).

Table 8: Stromal average H-score median split

Clinical variable	Condition	Negative	Positive	Total	Adjusted residual		X^2 (p value)
Gender	Female	212 (45.5)	193 (42)	405 (43.7)	1.1	-1.1	1.177 (0.278)
	Male	254 (54.5)	267 (58)	521 (56.3)	-1.1	1.1	
	Total	466 (100)	460 (100)	926 (100)			
3 year Survival*	Alive	326 (70.1)	356 (77.4)	682 (73.7)	-2.5	2.5	**6.334 (0.012)**
	Dead	139 (29.9)	104 (22.6)	243 (26.3)	2.5	-2.5	

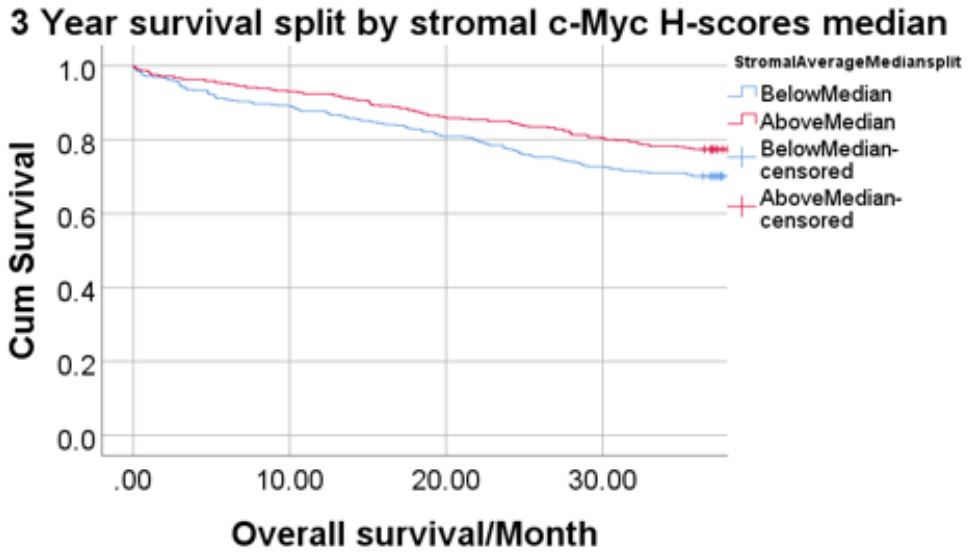
	Total	465 (100)	460 (100)	925 (100)			
Clinical variable	**Condition**	**Negative**	**Positive**	**Total**	**Adjusted residual**		**X^2 (p value)**
5 year Survival	Alive	197 (42.3)	209 (45.4)	406 (43.8)	-1	1	0.939 (0.333)
	Dead	269 (57.7)	251 (54.6)	520 (56.2)	1	-1	
	Total	466 (100)	460 (100)	926 (100)			
Synchronous metastasis	No Metastasis	411 (88.2)	411 (89.3)	822 (88.8)	-0.6	0.6	0.307 (0.579)
	With metastasis	55 (11.8)	49 (10.7)	104 (11.2)	0.6	-0.6	
	Total	466 (100)	460 (100)	926 (100)			
Recurrence	No Recurrence	322 (69.1)	341 (74.1)	663 (71.6)	-1.7	1.7	2.882 (0.09)
	Recurrence	144 (30.9)	119 (25.9)	263 (28.4)	1.7	-1.7	
	Total	466 (100)	460 (100)	926 (100)			
Cause of death	Alive	271 (58.2)	312 (67.8)	583 (63)	-3	3	**11.62 (0.009)**
	CRC	91 (19.5)	70 (15.2)	161 (17.4)	1.7	-1.7	
	Other	59 (12.7)	53 (11.5)	112 (12.1)	0.5	-0.5	
	Unknown	45 (9.7)	25 (5.4)	70 (7.6)	2.4	-2.4	
	Total	466 (100)	460 (100)	926 (100)			
Survival Status	Alive	271 (58.2)	312 (67.8)	583 (63)	-3	3	**9.285 (0.002)**
	Dead	195 (41.8)	148 (32.2)	343 (37)	3	-3	
	Total	466 (100)	460 (100)	926 (100)			
Final stage	1	58 (12.4)	95 (20.7)	153 (16.5)	-3.4	3.4	**12.23 (0.007)**
	2	193 (41.4)	185 (40.2)	378 (40.8)	0.4	-0.4	
	3	160 (34.3)	132 (28.7)	292 (31.5)	1.8	-1.8	
	4	55 (11.8)	48 (10.4)	103 (11.1)	0.7	-0.7	
	Total	466 (100)	460 (100)	926 (100)			
Primary tumour	T1	20 (4.3)	50 (10.9)	70 (7.6)	-3.8	3.8	**17.39 (0.001)**
	T2	48 (10.3)	52 (11.3)	100 (10.8)	-0.5	0.5	
	T3	248 (53.2)	242 (52.6)	490 (52.9)	0.2	-0.2	
	T4	150 (32.2)	116 (25.2)	266 (28.7)	2.3	-2.3	
	Total	466 (100)	460 (100)	926 (100)			
Lymph node metastasis	N0	255 (55.8)	280 (63.2)	535 (59.4)	-2.3	2.3	**7.961 (0.019)**
	N1	115 (25.2)	107 (24.2)	222 (24.7)	0.4	-0.4	
	N2	87 (19)	56 (12.6)	143 (15.9)	2.6	-2.6	
	Total	457 (100)	443 (100)	900 (100)			
Metastasis	No metastasis	411 (88.2)	411 (89.3)	822 (88.8)	-0.6	0.6	0.307 (0.579)
	With metastasis	55 (11.8)	49 (10.7)	104 (11.2)	0.6	-0.6	
	Total	466 (100)	460 (100)	926 (100)			
Vascular invasion	Absent	221 (47.9)	251 (55.8)	472 (51.8)	-2.4	2.4	**5.604 (0.018)**
	Present	240 (52.1)	199 (44.2)	439 (48.2)	2.4	-2.4	
	Total	461 (100)	450 (100)	911 (100)			

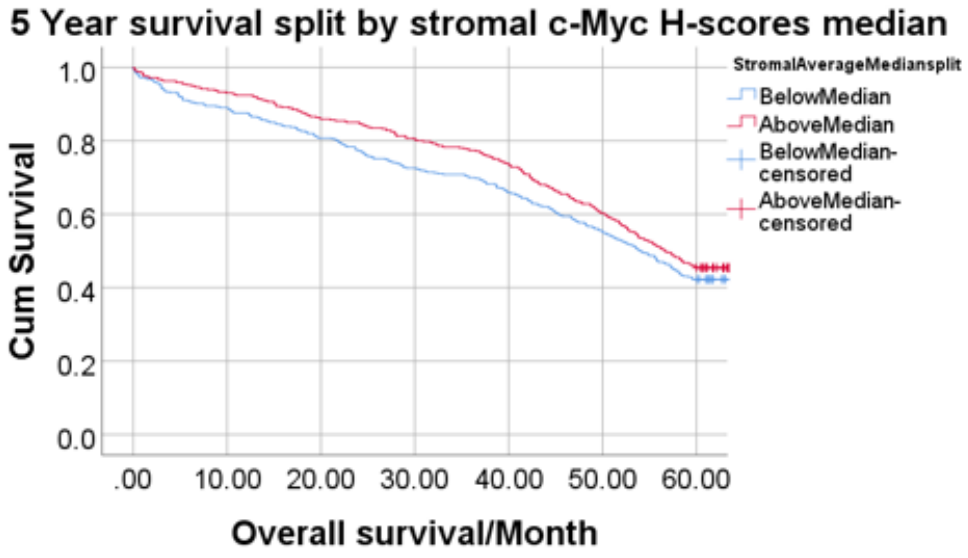
Clinical variable	Condition	Negative	Positive	Total	Adjusted residual		X^2 (p value)
Perineural invasion	Absent	361 (79.5)	363 (82.7)	724 (81.1)	-1.2	1.2	1.464 (0.226)
	Present	93 (20.5)	76 (17.3)	169 (18.9)	1.2	-1.2	
	Total	454 (100)	439 (100)	893 (100)			
Intramural lymphovascular invasion	Absent	292 (69.4)	306 (68.8)	598 (69.1)	0.2	-0.2	0.036 (0.85)
	Present	129 (30.6)	139 (31.2)	268 (30.9)	-0.2	0.2	
	Total	421 (100)	445 (100)	866 (100)			
Tumour edge	Infiltrative	170 (54)	195 (49.5)	365 (51.5)	1.2	-1.2	1.404 (0.236)
	Pushing	145 (46)	199 (50.5)	344 (48.5)	-1.2	1.2	
	Total	315 (100)	394 (100)	709 (100)			
Tumour budding score	Low	203 (66.3)	266 (67.5)	469 (67)	-0.3	0.3	0.107 (0.743)
	High	103 (33.7)	128 (32.5)	231 (33)	0.3	-0.3	
	Total	306 (100)	394 (100)	700 (100)			
Peritumoural lymphocytes	Inconspicuous	254 (80.4)	253 (63.6)	507 (71)	4.9	-4.9	**24.18 (<0.001)**
	Conspicuous	62 (19.6)	145 (36.4)	207 (29)	-4.9	4.9	
	Total	316 (100)	398 (100)	714 (100)			
MMR Status		1 (0.2)	5 (1.1)	6 (0.6)	-1.7	1.7	7.225 (0.027)
	Proficient	402 (86.3)	370 (80.4)	772 (83.4)	2.4	-2.4	
	Deficient	63 (13.5)	85 (18.5)	148 (16)	-2.1	2.1	
	Total	466 (100)	460 (100)	926 (100)			
Age dichotomised	Young	27 (5.8)	29 (6.3)	56 (6)	-0.3	0.3	0.106 (0.745)
	Old	439 (94.2)	431 (93.7)	870 (94)	0.3	-0.3	
	Total	466 (100)	460 (100)	926 (100)			
Grade dichotomised	Low	427 (91.6)	424 (92.4)	851 (92)	-0.4	0.4	0.174 (0.677)
	High	39 (8.4)	35 (7.6)	74 (8)	0.4	-0.4	
	Total	466 (100)	459 (100)	925 (100)			
Site dichotomised	Right side	220 (47.2)	230 (50)	450 (48.6)	-0.8	0.8	0.721 (0.396)
	Left side	246 (52.8)	230 (50)	476 (51.4)	0.8	-0.8	
	Total	466 (100)	460 (100)	926 (100)			
**Significant p values (<0.05) highlighted in bold*							

Survival analyses of stromal c-Myc expression

Kaplan-Meier analysis was performed using 3 year survival, 5 year survival and death caused by colorectal cancer as defining events, stratified by positive and negative stromal c-Myc H-score and by highest tertile. 3 year survival was significantly higher for positive stromal c-Myc (p 0.009), with negative stromal c-Myc expression having an approximately 70% chance of death after 3 years and c-Myc positive patients having a roughly 77% chance of survival. There was no significant difference between 3 year survival if the data was stratified by

highest tertile (p 0.094). There was no significant difference between survival at 5 years regardless if the split was by median (p 0.133) or highest tertile (p 0.255). Death relating to CRC was significantly different when the data was stratified by highest tertile (p 0.026), with stromal c-Myc negative patients having a 79% chance of survival and stromal c-Myc positive having an 86% chance of survival. When split by median (p 0.062) there was no significant difference.





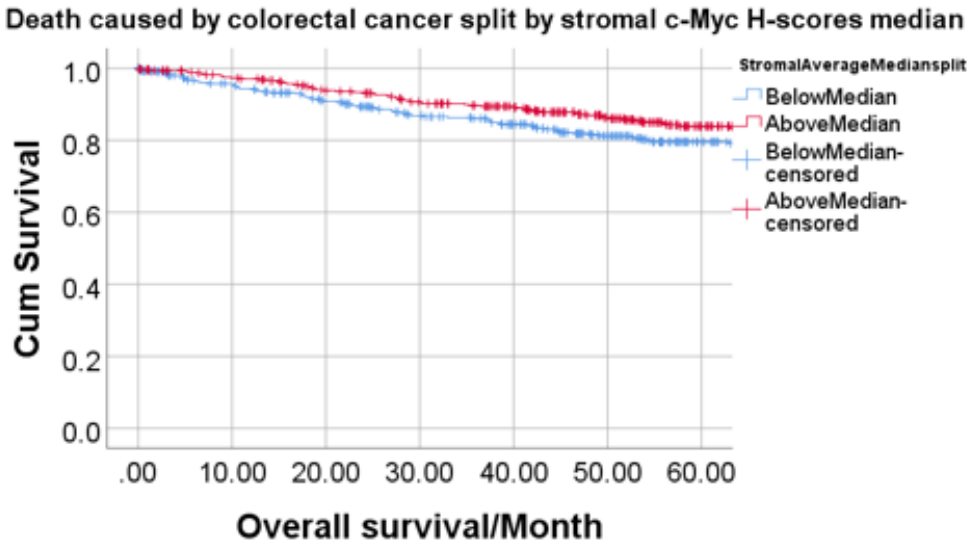


Figure 12: Kaplan-Meier analysis of stromal c-Myc expression. 3 year survival and death relating to colorectal cancer were significantly related to c-Myc stromal expression.

Table 9: Cox regression analysis of stromal c-Myc expression and significantly associated variables				
Variable	**Hazard ratio**	**95.0% CI**		**p value**
		Lower	**Upper**	
Stromal Average Median split	0.753	0.608	0.933	**0.01**
Clinical Variable	**Hazard ratio**	**95.0% CI**		**p value**
		Lower	**Upper**	
Final Stage 1*				**<0.001**
Final Stage 2	0.001	0	6.82E+20	0.807
Final Stage 3	0.001	0	5.01E+20	0.798
Final Stage 4	0.002	0	1.17E+21	0.822
Primary tumour T1				**<0.001**
Primary tumour T2	1.218	0.334	4.449	0.765
Primary tumour T3	1550.767	0	1E+27	0.793
Primary tumour T4	3135.546	0	2.03E+27	0.774
Lymph node metastasis N0				**0.004**
Lymph node metastasisN1	1.193	0.677	2.103	0.54
Lymph node metastasis N2	2.017	1.138	3.577	**0.016**
Vascular invasion	1.095	0.802	1.496	0.569
Peritumoural lymphocytes	0.821	0.602	1.119	0.212
Recurrence	2.971	2.168	4.072	**<0.001**
MMR Status				**<0.001**
MMR Status Proficient	1.165	0.161	8.4	0.88
MMR Status Deficient	2.329	0.318	17.075	0.406
Site dichotomised	0.88	0.675	1.148	0.345
Stromal Average Median split	0.832	0.645	1.074	0.158
**Significant associations highlighted in bold*				

Cox regression analysis was performed with stromal c-Myc expression and in univariate analysis was significantly associated with 5 year survival (p 0.01). Increased primary tumour stage (p<0.001) and lymph node metastasis (p 0.004) were associated with poorer survival, as was recurrence (p <0.001). Stromal c-Myc expression was not significantly associated with survival in the multivariate analysis (p 0.158), despite there being a higher risk of death in the c-Myc negative group in the univariate analysis (p 0.01). This suggests that c-Myc stromal expression is confounded by other clinical variables.

Does the staining intensity of one cellular compartment affect the staining in another area?

c-Myc is expected to be expressed in the nucleus, as it plays an important role in controlling transcription and increasing expression of 15% of the genome. The analysis performed here however shows that c-Myc can also be expressed in the cytoplasm. It is important to understand if the expression of one cellular compartment is related to the expression in another, in order to be able to effectively use c-Myc as a biomarker. The scatter plot in figure 13 below shows that there is a moderate correlation between nuclear average H-score and

cytoplasmic average positive percentage. The Spearman's rank correlation showed a moderately positive (0.429), significant ($p<0.001$) correlation. This suggests that patients who are nuclear c-Myc positive may also be positive for cytoplasmic c-Myc. The median H-score of 54 was used as the cut-off for nuclear expression as it showed a significant association with survival. The 75th percentile of positive cytoplasmic staining was used to avoid including non-specific background staining. This is shown in the following groups:

Table 10: Combination of c-Myc expression in the nucleus and in the cytoplasm			
		Nuclear (%)	
		Below median	Above median
Cytoplasmic (%)	Below 75th percentile	486 (48.6)	272 (27.2)
	Above 75th percentile	49 (0.49)	193 (19.3)

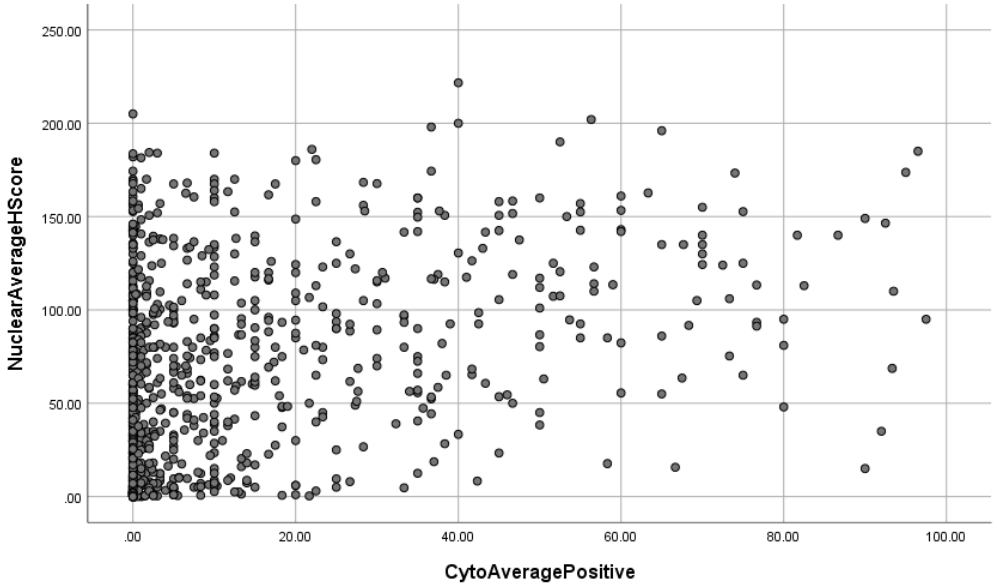


Figure 13: Scatter plot showing the relationship between c-Myc nuclear and cytoplasmic H-scores.

Statistical analysis of CD44

Membrane staining of CD44

Average membrane H-score was calculated from luminal, central and advancing edge samples. 86 were excluded due to insufficient data for all three individual tumour regions. The distribution of all membrane H-scores showed a right-skew as seen in figure 14. The median expression of membrane average H-score was 15 (IQR 0-57.5), and the average H-score was highly correlated (Spearman's rank >0.8) with all three tumour regions. Luminal, central and advancing edge samples all showed a median of 20 (IQR 0-85, 0-75 and 0-75, respectively). In the normal colon, expression of CD44 tended to be observed in the base of the crypts only (as seen in figure 5), associated with stem cells. This is expected given the role CD44 plays in colonic stem cell maintenance [68].

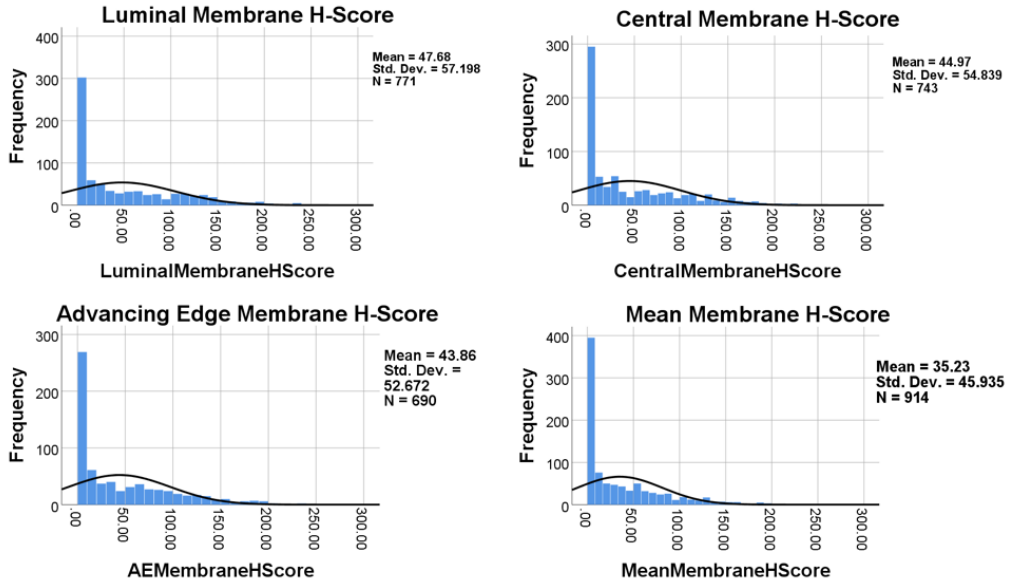


Figure 14: Distribution of CD44 membrane staining in individual tumour regions and average H-scores.

Membrane Chi-square analysis

As the median H-score of CD44 was so low in the membrane, it is possible that any staining artefact present could influence the results. For this reason, the data was assessed with an average H-score cut-off of 57.5 (75th percentile). Mismatch repair status was very strongly associated with expression of CD44 ($p < 0.001$) and high CD44 expression was associated with MMR deficient tumours. Mismatch repair status is important in determining treatment of colorectal cancer [84], and may influence which pathways are involved in tumourigenesis. Because of this, the data were categorised into mismatch repair proficient and deficient prior to Chi-squared analysis. The results are displayed in the table below alongside the ungrouped data.

The analysis returned a number of significant associations with high membranous CD44 expression without differentiating between mismatch

repair status, including with females, (p 0.009), poor 5 year survival (p 0.001) and no metastasis (p 0.003). Lack of recurrence was weakly associated with high CD44 expression (p 0.004), as were right-sided tumours (p 0.011). High final stage and primary tumour stage were associated with low CD44 expression (p 0.001 and <0.001 respectively). High CD44 expression was associated with the absence of vascular invasion (p <0.001), and a pushing tumour edge (p <0.001), as well as showing weak association with conspicuous peritumoural lymphocytes (p 0.04).

When the data were categorised by the MMR status into proficient (pMMR) or deficient (dMMR), several of the clinical variables showed differences between the two groups. In pMMR patients, 3 year survival was significantly higher in high expressors of CD44 than in patients with low CD44 expression (p 0.032), whereas it did not differ in the uncategorised data. 5 year survival was significantly higher in low expressors of CD44 (p 0.01), but less so than when the data was uncategorised (p 0.001). Metastasis (p 0.015) was significantly associated with low CD44 expression in pMMR patients, although less so than in the uncategorised data (p 0.003). The association between low primary tumour stage and high CD44 expression increased in the pMMR group compared to the uncategorised data (p <0.001). Vascular invasion (p <0.001) was associated with low CD44 membrane expression in the proficient patients, as was an infiltrative tumour edge (p 0.016).

Only gender was specific to MMR deficient patients, with female patients having higher expression of membranous CD44 (p 0.013). Grade dichotomised showed a stronger association between high grade (poorly differentiated tumours) and low CD44 expression in dMMR patients (p 0.008) as opposed to pMMR patients (p 0.034). Some associations which were significant before distinguishing between MMR status were lost in this analysis. This includes recurrence, which although marginally significant without differentiating for MMR status (p 0.04), was not significantly associated with CD44 expression in pMMR (p 0.073) or dMMR (p 0.955) patients alone. Peritumoural lymphocytes also lost their significance after mismatch repair status was accounted for. Finally, right-sided tumours were significantly associated with high CD44 membranous expression prior to MMR status split, at which point almost perfectly opposite expression could be observed in MMR proficient (p 0.348) and deficient (p 0.377) groups (see table 11). Only grade dichotomised was significantly associated with CD44 expression in the two MMR groups but not in the overall analysis. In pMMR patients, high CD44 expression was enriched in well differentiated tumours and low CD44 expression was enriched poorly differentiated tumours (p 0.034), whereas in dMMR patients high CD44 expression was enriched in patients with low grade tumours and low CD44 expression was enriched in high grade, poorly differentiated tumours (p 0.008). Age, tumour budding score, cause of death, lymph node metastasis, perineural

invasion and intratumoural vascular invasion did not show any association with CD44 membrane expression.

Table 11: Chi-square analysis of membranous CD44 expression and clinical variables								
MMR Status	**Clinical variable**	**Condition**	**Low**	**High**	**Total**	**Adjusted residual**		**X^2 (p value)**
	MMR Status	Proficient	604 (88.3)	153 (69.2)	757 (83.6)	6.7	-6.7	**44.421 (<0.001)**
		Deficient	80 (11.7)	68 (30.8)	148 (16.4)	-6.7	6.7	
		Total	684 (75.6)	221 (24.4)	905 (100)			
Proficient	**Gender**	Female	250 (41.4)	72 (47.1)	322 (42.5)	-1.3	1.3	1.605 (0.205)
		Male	354 (58.6)	81 (52.9)	435 (57.5)	1.3	-1.3	
		Total	604 (100)	153 (100)	757 (100)			
Deficient	**Gender**	Female	33 (41.3)	42 (61.8)	75 (50.7)	-2.5	2.5	**6.189 (0.013)**
		Male	47 (58.8)	26 (38.2)	73 (49.3)	2.5	-2.5	
		Total	80 (100)	68 (100)	148 (100)			
Total	**Gender**	Female	285 (41.3)	115 (51.3)	400 (43.8)	-2.6	2.6	**6.919 (0.009)**
		Male	405 (58.7)	109 (48.7)	514 (56.2)	2.6	-2.6	
		Total	690 (100)	224 (100)	914 (100)			
Proficient	**3 year Survival**	Alive	434 (71.9)	123 (80.4)	557 (73.6)	-2.1	2.1	**4.578 (0.032)**
		Dead	170 (28.1)	30 (19.6)	200 (26.4)	2.1	-2.1	
		Total	604 (100)	153 (100)	757 (100)			
Deficient	**3 year Survival**	Alive	58 (72.5)	46 (68.7)	104 (70.7)	0.5	-0.5	0.260 (0.61)
		Dead	22 (27.5)	21 (31.3)	43 (29.3)	-0.5	0.5	
		Total	80 (100)	67 (100)	147 (100)			
Total	**3 year Survival**	Alive	497 (72)	171 (76.7)	668 (73.2)	-1.4	1.4	1.858 (0.173)
		Dead	193 (28)	52 (23.3)	245 (26.8)	1.4	-1.4	
		Total	690 (100)	223 (100)	913 (100)			
Proficient	**5 year Survival**	Alive	275 (45.5)	52 (34)	327 (43.2)	2.6	-2.6	**6.629 (0.01)**
		Dead	329 (54.5)	101 (66)	430 (56.8)	-2.6	2.6	
		Total	604 (100)	153 (100)	757 (100)			
Deficient	**5 year Survival**	Alive	36 (45)	22 (32.4)	58 (39.2)	1.6	-1.6	2.467 (0.116)
		Dead	44 (55)	46 (67.6)	90 (60.8)	-1.6	1.6	
		Total	80 (100)	68 (100)	148 (100)			
Total	**5 year Survival**	Alive	316 (45.8)	75 (33.5)	391 (42.8)	3.2	-3.2	**10.47 (0.001)**

		Dead	374 (54.2)	149 (66.5)	523 (57.2)	-3.2	3.2	
		Total	690 (100)	224 (100)	914 (100)			
MMR Status	**Clinical variable**	**Condition**	**Low**	**High**	**Total**	**Adjusted residual**		**X^2 (p value)**
Proficient	**Synchronous metastasis**	No metastasis	521 (86.3)	143 (93.5)	664 (87.7)	-2.4	2.4	**5.882 (0.015)**
		With metastasis	83 (13.7)	10 (6.5)	93 (12.3)	2.4	-2.4	
		Total	604 (100)	153 (100)	757 (100)			
Deficient	**Synchronous metastasis**	No metastasis	74 (92.5)	66 (97.1)	140 (94.6)	-1.2	1.2	1.494 (0.222)
		With metastasis	6 (7.5)	2 (2.9)	8 (5.4)	1.2	-1.2	
		Total	80 (100)	68 (100)	148 (100)			
Total	**Synchronous metastasis**	No metastasis	601 (87.1)	211 (94.2)	812 (88.8)	-2.9	2.9	**8.586 (0.003)**
		With metastasis	89 (12.9)	13 (5.8)	102 (11.2)	2.9	-2.9	
		Total	690 (100)	224 (100)	914 (100)			
Proficient	**Recurrence**	No Recurrence	413 (68.4)	116 (75.8)	529 (69.9)	-1.8	1.8	3.210 (0.073)
		Recurrence	191 (31.6)	37 (24.2)	228 (30.1)	1.8	-1.8	
		Total	604 (100)	153 (100)	757 (100)			
Deficient	**Recurrence**	No Recurrence	65 (81.3)	55 (80.9)	120 (81.1)	0.1	-0.1	0.003 (0.955)
		Recurrence	15 (18.8)	13 (19.1)	28 (18.9)	-0.1	0.1	
		Total	80 (100)	68 (100)	148 (100)			
Total	**Recurrence**	No Recurrence	484 (70.1)	173 (77.2)	657 (71.9)	-2	2	**4.202 (0.04)**
		Recurrence	206 (29.9)	51 (22.8)	257 (28.1)	2	-2	
		Total	690 (100)	224 (100)	914 (100)			
Proficient	**Survival Status**	Alive	371 (61.4)	108 (70.6)	479 (63.3)	-2.1	2.1	**4.412 (0.036)**
		Dead	233 (38.6)	45 (29.4)	278 (36.7)	2.1	-2.1	
		Total	604 (100)	153 (100)	757 (100)			
Deficient	**Survival Status**	Alive	49 (61.3)	38 (55.9)	87 (58.8)	0.7	-0.7	0.437 (0.509)
		Dead	31 (38.8)	30 (44.1)	61 (41.2)	-0.7	0.7	
		Total	80 (100)	68 (100)	148 (100)			
Total	**Survival Status**	Alive	425 (61.6)	148 (66.1)	573 (62.7)	-1.2	1.2	1.449 (0.229)
		Dead	265 (38.4)	76 (33.9)	341 (37.3)	1.2	-1.2	
		Total	690 (100)	224 (100)	914 (100)			
Proficient	**Final stage**	1	90 (14.9)	45 (29.4)	135 (17.8)	-4.2	4.2	**20.40 (<0.001)**

		2	230 (38.1)	51 (33.3)	281 (37.1)	1.1	-1.1	
		3	202 (33.4)	47 (30.7)	249 (32.9)	0.6	-0.6	
		4	82 (13.6)	10 (6.5)	92 (12.2)	2.4	-2.4	
		Total	604 (100)	153 (100)	757 (100)			
Deficient	**Final stage**	1	4 (5)	7 (10.3)	11 (7.4)	-1.2	1.2	6.008 (0.111)
		2	54 (67.5)	37 (54.4)	91 (61.5)	1.6	-1.6	
		3	16 (20)	22 (32.4)	38 (25.7)	-1.7	1.7	
		4	6 (7.5)	2 (2.9)	8 (5.4)	1.2	-1.2	
		Total	80 (100)	68 (100)	148 (100)			
Total	**Final stage**	1	95 (13.8)	52 (23.2)	147 (16.1)	-3.3	3.3	**16.93 (0.001)**
		2	287 (41.6)	89 (39.7)	376 (41.1)	0.5	-0.5	
		3	220 (31.9)	70 (31.3)	290 (31.7)	0.2	-0.2	
		4	88 (12.8)	13 (5.8)	101 (11.1)	2.9	-2.9	
		Total	690 (100)	224 (100)	914 (100)			
MMR Status	**Clinical variable**	**Condition**	**Low**	**High**	**Total**	**Adjusted residual**		**X^2 (p value)**
Proficient	**Primary tumour**	T1	35 (5.8)	28 (18.3)	63 (8.3)	-5	5	**30.51 (<0.001)**
		T2	65 (10.8)	24 (15.7)	89 (11.8)	-1.7	1.7	
		T3	321 (53.1)	68 (44.4)	389 (51.4)	1.9	-1.9	
		T4	183 (30.3)	33 (21.6)	216 (28.5)	2.1	-2.1	
		Total	604 (100)	153 (100)	757 (100)			
Deficient	**Primary tumour**	T1	1 (1.3)	3 (4.4)	4 (2.7)	-1.2	1.2	1.575 (0.665)
		T2	4 (5)	4 (5.9)	8 (5.4)	-0.2	0.2	
		T3	51 (63.8)	43 (63.2)	94 (63.5)	0.1	-0.1	
		T4	24 (30)	18 (26.5)	42 (28.4)	0.5	-0.5	
		Total	80 (100)	68 (100)	148 (100)			
Total	**Primary tumour**	T1	36 (5.2)	31 (13.8)	67 (7.3)	-4.3	4.3	**21.15 (<0.001)**
		T2	70 (10.1)	28 (12.5)	98 (10.7)	-1	1	
		T3	376 (54.5)	112 (50)	488 (53.4)	1.2	-1.2	
		T4	208 (30.1)	53 (23.7)	261 (28.6)	1.9	-1.9	
		Total	690 (100)	224 (100)	914 (100)			
Proficient	**Metastasis**	No metastasis	521 (86.3)	143 (93.5)	664 (87.7)	-2.4	2.4	**5.882 (0.015)**

		With metastasis	83 (13.7)	10 (6.5)	93 (12.3)	2.4	-2.4	
		Total	604 (100)	153 (100)	757 (100)			
Deficient	**Metastasis**	No metastasis	74 (92.5)	66 (97.1)	140 (94.6)	-1.2	1.2	1.494 (0.222)
		With metastasis	6 (7.5)	2 (2.9)	8 (5.4)	1.2	-1.2	
		Total	80 (100)	68 (100)	148 (100)			
Total	**Metastasis**	No metastasis	601 (87.1)	211 (94.2)	812 (88.8)	-2.9	2.9	**8.586 (0.003)**
		With metastasis	89 (12.9)	13 (5.8)	102 (11.2)	2.9	-2.9	
		Total	690 (100)	224 (100)	914 (100)			
MMR Status	**Clinical variable**	**Condition**	**Low**	**High**	**Total**	**Adjusted residual**		**X^2 (p value)**
Proficient	**Vascular invasion**	Absent	280 (46.7)	90 (63.4)	370 (49.9)	-3.6	3.6	**12.83 (<0.001)**
		Present	320 (53.3)	52 (36.6)	372 (50.1)	3.6	-3.6	
		Total	600 (100)	142 (100)	742 (100)			
Deficient	**Vascular invasion**	Absent	44 (55)	45 (66.2)	89 (60.1)	-1.4	1.4	1.915 (0.166)
		Present	36 (45)	23 (33.8)	59 (39.9)	1.4	-1.4	
		Total	80 (100)	68 (100)	148 (100)			
Total	**Vascular invasion**	Absent	328 (47.8)	137 (64.3)	465 (51.7)	-4.2	4.2	**17.73 (<0.001)**
		Present	358 (52.2)	76 (35.7)	434 (48.3)	4.2	-4.2	
		Total	686 (100)	213 (100)	899 (100)			
Proficient	**Tumour edge**	Infiltrative	258 (58.2)	52 (45.6)	310 (55.7)	2.4	-2.4	**5.856 (0.016)**
		Pushing	185 (41.8)	62 (54.4)	247 (44.3)	-2.4	2.4	
		Total	443 (100)	114 (100)	557 (100)			
Deficient	**Tumour edge**	Infiltrative	21 (33.9)	15 (24.2)	36 (29)	1.2	-1.2	1.409 (0.235)
		Pushing	41 (66.1)	47 (75.8)	88 (71)	-1.2	1.2	
		Total	62 (100)	62 (100)	124 (100)			
Total	**Tumour edge**	Infiltrative	281 (55.1)	67 (37.6)	348 (50.6)	4	-4	**16.08 (<0.001)**
		Pushing	229 (44.9)	111 (62.4)	340 (49.4)	-4	4	
		Total	510 (100)	178 (100)	688 (100)			
Proficient	**Peritumoural lymphocytes**	Inconspicuous	336 (76.2)	80 (69)	416 (74.7)	1.6	-1.6	2.536 (0.111)
		Conspicuous	105 (23.8)	36 (31)	141 (25.3)	-1.6	1.6	
		Total	441 (100)	116 (100)	557 (100)			
Deficient	**Peritumoural lymphocytes**	Inconspicuous	38 (58.5)	36 (58.1)	74 (58.3)	0	0	0.002 (0.964)

		Conspicuous	27 (41.5)	26 (41.9)	53 (41.7)	0	0	
		Total	65 (100)	62 (100)	127 (100)			
Total	**Peritumoural lymphocytes**	Inconspicuous	376 (73.6)	118 (65.6)	494 (71.5)	2.1	-2.1	**4.207 (0.04)**
		Conspicuous	135 (26.4)	62 (34.4)	197 (28.5)	-2.1	2.1	
		Total	511 (100)	180 (100)	691 (100)			
MMR Status	**Clinical variable**	**Condition**	**Low**	**High**	**Total**	**Adjusted residual**		**X^2 (p value)**
Proficient	**Grade dichotomised**	Low	572 (94.7)	150 (98.7)	722 (95.5)	-2.1	2.1	**4.484 (0.034)**
		High	32 (5.3)	2 (1.3)	34 (4.5)	2.1	-2.1	
		Total	604 (100)	152 (100)	756 (100)			
Deficient	**Grade dichotomised**	Low	50 (62.5)	56 (82.4)	106 (71.6)	-2.7	2.7	**7.128 (0.008)**
		High	30 (37.5)	12 (17.6)	42 (28.4)	2.7	-2.7	
		Total	80 (100)	68 (100)	148 (100)			
Total	**Grade dichotomised**	Low	628 (91)	208 (93.3)	836 (91.6)	-1.1	1.1	1.114 (0.291)
		High	62 (9)	15 (6.7)	77 (8.4)	1.1	-1.1	
		Total	690 (100)	223 (100)	913 (100)			
Proficient	**Site dichotomised**	Right side	251 (41.6)	70 (45.8)	321 (42.4)	-0.9	0.9	0.880 (0.348)
		Left side	353 (58.4)	83 (54.2)	436 (57.6)	0.9	-0.9	
		Total	604 (100)	153 (100)	757 (100)			
Deficient	**Site dichotomised**	Right side	69 (86.3)	55 (80.9)	124 (83.8)	0.9	-0.9	0.779 (0.377)
		Left side	11 (13.8)	13 (19.1)	24 (16.2)	-0.9	0.9	
		Total	80 (100)	68 (100)	148 (100)			
Total	**Site dichotomised**	Right side	321 (46.5)	126 (56.3)	447 (48.9)	-2.5	2.5	**6.405 (0.011)**
		Left side	369 (53.5)	98 (43.8)	467 (51.1)	2.5	-2.5	
		Total	690 (100)	224 (100)	914 (100)			
**Significant results (p <0.05) highlighted in bold*								

Survival analysis of average CD44 membrane H-scores

Kaplan-Meier analysis was used to measure survival status at 3 and 5 years, and in the case of death relating to colorectal cancer, as seen in figure 15. The data was dichotomised by 75th percentile and stratified by MMR status. 3 year survival was not significantly different (p 0.101) between patients with low and high expression of membranous CD44, even considering MMR status. Median survival could not be calculated but 71.9% of patients survived to three years with low CD44, and 80.4% with high CD44 in the pMMR group. Approximately

72.5% of patients survived with low CD44 and 68.7% survived with high CD44 in the dMMR group. There was a significant difference in 5 year survival between high and low expressors of membranous CD44 (p 0.022). In pMMR patients, low expression of membranous CD44 led to improved 5 year survival (median 56.4 months, confidence intervals not computed) over high membranous CD44 expression (51.8 months 95% CI 49-55 months). In dMMR patients, 5 year survival was also higher in patients expressing low membranous CD44 (55.7 months 95% CI 48-64 months), than in patients with high membranous CD44 (51.3 months 95% CI 46-57 months). The 5 year survival curves in pMMR patients crossed over at approximately 40 months, casting some doubt on the reliability of true significance of the results. The statistical analysis was repeated using the Tarone-Ware method [85], instead of the log rank test used in previous analyses, and showed a non-significant relationship (p 0.102) between CD44 expression and survival. Death related to CRC was not significantly different between high and low CD44 membranous expression (p 0.615). Approximate 5 year survival for patients with low or high CD44 membrane expression in pMMR patients was 80.5% and 82.2% and in dMMR was 84.7 and 84.3 in dMMR.

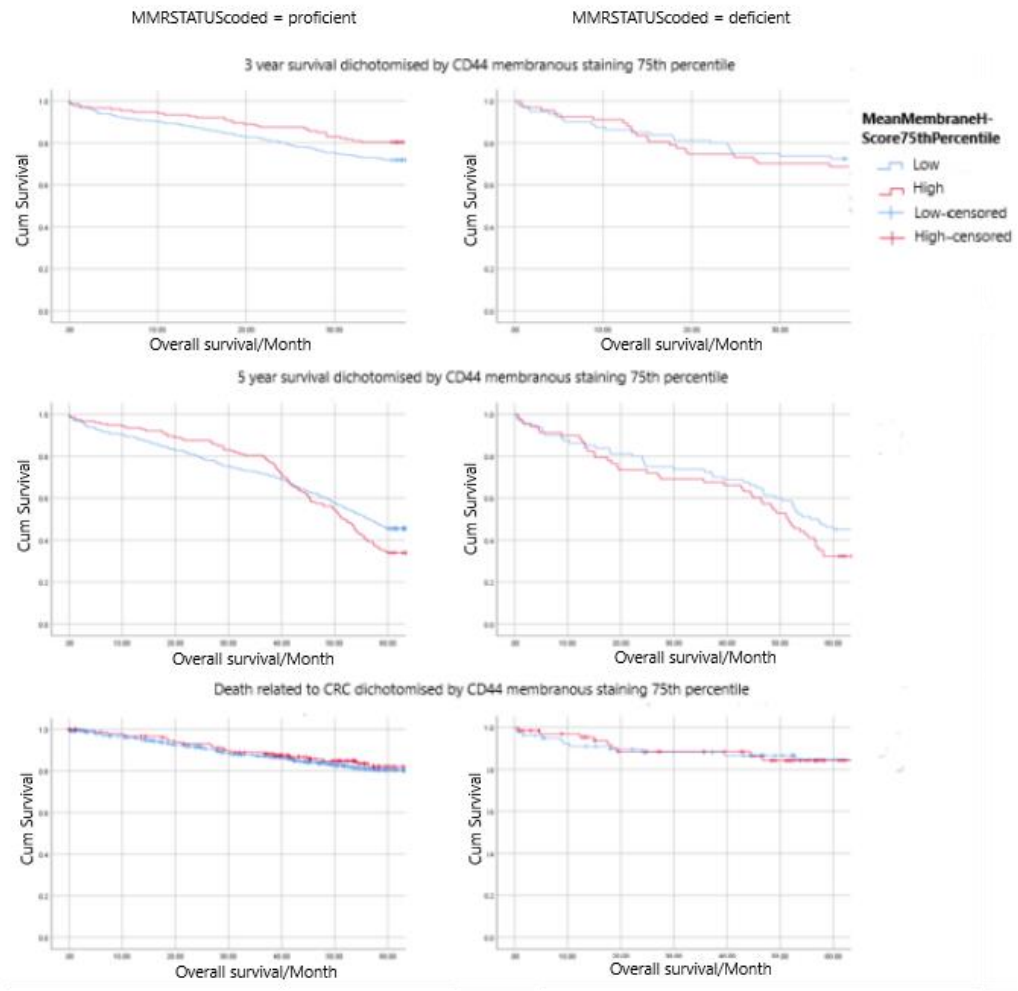


Figure 15: Kaplan-Meier survival curves for CD44 membrane staining split by 75th percentile and stratified by MMR status, with MMR proficient patients on the left and MMR deficient patients on the right.

Univariate and multivariate Cox regression analyses were performed on overall survival to assess whether average membrane CD44 staining was associated with survival or whether the effects on survival were due to other factors, as seen in table 12. All data was stratified by MMR status. Overall survival was not significantly related to increased membranous CD44 staining in univariate analysis (p 0.483) but was when included in multivariate analysis (p 0.035). Recurrence significantly increased the risk of death (p <0.001), as did increased primary tumour stage (p <0.001) and grade dichotomised (p <0.001). Pushing tumour edge led to a significantly lower hazard ratio (p 0.035) and better survival.

Table 12: Cox regression analysis of CD44 staining and significantly associated variables				
Variable	**Hazard ratio**	**95.0% CI**		**p value**
		Lower	**Upper**	
Mean Membrane H-Score 75th Percentile	0.91	0.7	1.184	0.483
	Hazard Ratio	**95.0% CI**		**p value**
		Lower	**Upper**	
Mean Membrane H-Score 75th Percentile*	1.387	1.023	1.88	**0.035**
Gender	1.174	0.904	1.524	0.228
Synchronous metastasis	1.602	0.454	5.652	0.464
Recurrence	2.715	1.982	3.718	**<0.001**
Final Stage				0.896
Final Stage (1)	0.561	0.07	4.525	0.587
Final Stage (2)	0.519	0.066	4.091	0.534
Final Stage (3)	0.452	0.026	7.986	0.588
Primary tumour T1				**<0.001**
Primary tumour T2	1.16	0.315	4.27	0.823
Primary tumour T3	2.513	0.24	26.28	0.442
Primary tumour T4	5.353	0.509	56.24	0.162
Metastasis	1.595	0.151	16.867	0.698
Vascular invasion	1.181	0.872	1.598	0.283
Tumour edge	0.717	0.526	0.978	**0.035**
Peritumoural lymphocytes	0.741	0.536	1.023	0.069
Grade Dichotomised	2.221	1.497	3.295	**<0.001**
Site Dichotomised	0.972	0.741	1.275	0.839
**Significant results (p <0.05) highlighted in bold*				

Independent tumour region expression: survival analysis

5 year survival was analysed for individual tumour regions split at the 75th percentile to see if the crossing of survival curves observed in figure 15 was due to differences in the tumour regions. As seen in figure 16 below, the curves for the individual tumour regions still cross over. 5 year survival was not significantly different in luminal tumour samples, when assessed by Tarone-Ware analysis (p 0.186). Advancing edge tumours showed an earlier crossover, around 46 months, and survival was not significantly different (p 0.543). In the central tumour region, high and low CD44 membrane expressors crossed over at multiple points (approximately 19 and 44 months) and survival was significantly different by Tarone-Ware analysis (p 0.034). Individual tumour regions did not show distinct survival patterns in Kaplan-Meier curves, reflective of the lack of heterogeneity shown in the correlation analysis.

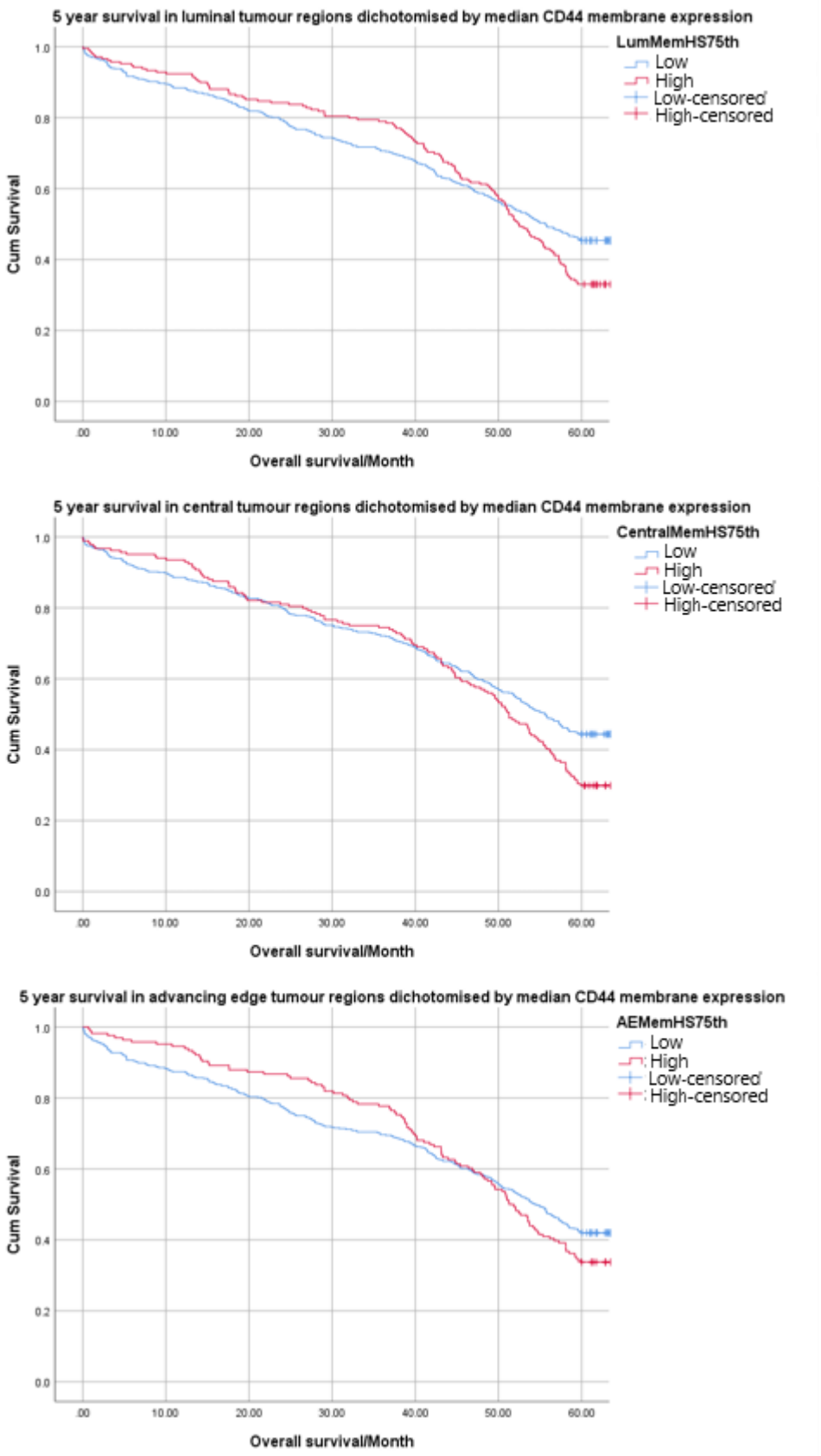


Figure 16: Survival analysis of individual tumour regions. Membrane-specific CD44 staining was assessed in 3 year survival, 5 year survival and death relating to colorectal cancer.

Cytoplasmic staining of CD44

Cytoplasmic staining was assessed by a H-score on a scale of 0-200, unlike other staining compartments, as the staining did not reach a high level of intensity. The distribution of the data was normal, and therefore the mean and standard deviation are presented in the table below. The data was split by median average cytoplasmic expression (58.33) in order to be consistent with other analyses and to ensure an equal split of positive and negative expressors.

Table 13: Details of cytoplasmic CD44 expression

	Luminal Cytoplasm H-Score	Central Cytoplasm H-Score	Advancing Edge Cytoplasm H-Score	Normal Cytoplasm H-score	Mean Cytoplasmic H-Score
Included	771	743	690	572	914
Excluded	229	257	310	428	86
Mean	50.26	45.76	43.01	30.07	61.88
Median	40	40	35	20	58.3333
Std. Deviation	43.15	40.09	38.71	33.82	43.63
Interquartile range	70	60	60	50	70
Pearson's correlation with average cytoplasmic H-Score (p value)	0.818 (<0.001)	0.799 (<0.001)	0.759 (<0.001)	0.51 (0.231)	-

Chi-squared analysis of cytoplasmic CD44

After the average cytoplasmic data was split by the median, it was compared with the clinical variables to determine if there was any association. The results are demonstrated in table 14 below. 3 year survival was significantly associated with positive cytoplasmic CD44 expression (p 0.002), as was the lack of metastasis (p 0.001), no recurrence (p <0.001) and overall survival (p 0.001). Higher final stage was associated with negative cytoplasmic CD44 (p <0.001), as was primary tumour stage (p <0.001) and lymph node metastasis (p <0.001). Absence of metastasis (p 0.001) and vascular invasion (p <0.001) were associated with positive CD44 expression. Presence of perineural invasion (p <0.001) and intratumoural lymphovascular invasion (p 0.005) were associated with negative cytoplasmic CD44. “Pushing” tumour edge (p 0.001) and dMMR (p 0.005) were both associated with positive cytoplasmic CD44 expression. Poorly differentiated tumours were associated with negative cytoplasmic CD44 expression (p <0.001).

Table 14: Chi-square analysis of cytoplasmic CD44 expression							
Clinical variable	**Condition**	**Negative**	**Positive**	**Total**	**Adjusted residual**		**X^2 (p value)**
Gender	Female	189 (41.8)	211 (45.7)	400 (43.8)	-1.2	1.2	1.381 (0.24)
	Male	263 (58.2)	251 (54.3)	514 (56.2)	1.2	-1.2	
	Total	452 (100)	462 (100)	914 (100)			
3 year Survival*	Alive	310 (68.6)	358 (77.7)	668 (73.2)	-3.1	3.1	**9.569 (0.002)**
	Dead	142 (31.4)	103 (22.3)	245 (26.8)	3.1	-3.1	
	Total	452 (100)	461 (100)	913 (100)			
5 year Survival	Alive	201 (44.5)	190 (41.1)	391 (42.8)	1	-1	1.043 (0.307)
	Dead	251 (55.5)	272 (58.9)	523 (57.2)	-1	1	
	Total	452 (100)	462 (100)	914 (100)			
Synchronous metastasis	1	386 (85.4)	426 (92.2)	812 (88.8)	-3.3	3.3	**10.68 (0.001)**
	2	66 (14.6)	36 (7.8)	102 (11.2)	3.3	-3.3	
	Total	452 (100)	462 (100)	914 (100)			
Recurrence	No Recurrence	301 (66.6)	356 (77.1)	657 (71.9)	-3.5	3.5	**12.37 (<0.001)**
	Recurrence	151 (33.4)	106 (22.9)	257 (28.1)	3.5	-3.5	
	Total	452 (100)	462 (100)	914 (100)			
Cause of death	Alive	259 (57.3)	314 (68)	573 (62.7)	-3.3	3.3	**12.88 (0.005)**
	CRC	95 (21)	63 (13.6)	158 (17.3)	3	-3	
	Other	56 (12.4)	52 (11.3)	108 (11.8)	0.5	-0.5	
	Unknown	42 (9.3)	33 (7.1)	75 (8.2)	1.2	-1.2	
	Total	452 (100)	462 (100)	914 (100)			
Survival Status	Alive	259 (57.3)	314 (68)	573 (62.7)	-3.3	3.3	**11.11 (0.001)**
	Dead	193 (42.7)	148 (32)	341 (37.3)	3.3	-3.3	
	Total	452 (100)	462 (100)	914 (100)			
Final stage	1	50 (11.1)	97 (21)	147 (16.1)	-4.1	4.1	**28.42 (<0.001)**
	2	177 (39.2)	199 (43.1)	376 (41.1)	-1.2	1.2	
	3	159 (35.2)	131 (28.4)	290 (31.7)	2.2	-2.2	
	4	66 (14.6)	35 (7.6)	101 (11.1)	3.4	-3.4	
	Total	452 (100)	462 (100)	914 (100)			
Primary tumour	T1	20 (4.4)	47 (10.2)	67 (7.3)	-3.3	3.3	**28.59 (<0.001)**
	T2	40 (8.8)	58 (12.6)	98 (10.7)	-1.8	1.8	
	T3	232 (51.3)	256 (55.4)	488 (53.4)	-1.2	1.2	
	T4	160 (35.4)	101 (21.9)	261 (28.6)	4.5	-4.5	
	Total	452 (100)	462 (100)	914 (100)			
Lymph node metastasis	N0	241 (53.7)	287 (65.4)	528 (59.5)	-3.6	3.6	**17.05 (<0.001)**
	N1	115 (25.6)	101 (23)	216 (24.3)	0.9	-0.9	
	N2	93 (20.7)	51 (11.6)	144 (16.2)	3.7	-3.7	
	Total	449 (100)	439 (100)	888 (100)			
Metastasis	No metastasis	386 (85.4)	426 (92.2)	812 (88.8)	-3.3	3.3	**10.68 (0.001)**

	With metastasis	66 (14.6)	36 (7.8)	102 (11.2)	3.3	-3.3	
	Total	452 (100)	462 (100)	914 (100)			
Clinical variable	**Condition**	**Negative**	**Positive**	**Total**	**Adjusted residual**		**X^2 (p value)**
Vascular invasion	Absent	197 (43.9)	268 (59.6)	465 (51.7)	-4.7	4.7	**22.13 (<0.001)**
	Present	252 (56.1)	182 (40.4)	434 (48.3)	4.7	-4.7	
	Total	449 (100)	450 (100)	899 (100)			
Perineural invasion	Absent	328 (74.2)	383 (87.2)	711 (80.7)	-4.9	4.9	**24.03 (<0.001)**
	Present	114 (25.8)	56 (12.8)	170 (19.3)	4.9	-4.9	
	Total	442 (100)	439 (100)	881 (100)			
Intramural lymphovascular invasion	Absent	266 (64.6)	320 (73.4)	586 (69.1)	-2.8	2.8	**7.738 (0.005)**
	Present	146 (35.4)	116 (26.6)	262 (30.9)	2.8	-2.8	
	Total	412 (100)	436 (100)	848 (100)			
Tumour edge	Infiltrative	189 (57.1)	159 (44.5)	348 (50.6)	3.3	-3.3	**10.84 (0.001)**
	Pushing	142 (42.9)	198 (55.5)	340 (49.4)	-3.3	3.3	
	Total	331 (100)	357 (100)	688 (100)			
Tumour budding score	Low	210 (64.8)	246 (69.5)	456 (67.3)	-1.3	1.3	1.680 (0.195)
	High	114 (35.2)	108 (30.5)	222 (32.7)	1.3	-1.3	
	Total	324 (100)	354 (100)	678 (100)			
Peritumoural lymphocytes	Inconspicuous	246 (74.5)	248 (68.7)	494 (71.5)	1.7	-1.7	2.892 (0.089)
	Conspicuous	84 (25.5)	113 (31.3)	197 (28.5)	-1.7	1.7	
	Total	330 (100)	361 (100)	691 (100)			
MMR Status	Proficient	393 (86.9)	364 (78.8)	757 (83.6)	3.1	-3.1	**9.814 (0.002)**
	Deficient	56 (12.4)	92 (19.9)	148 (16.4)	-3.1	3.1	
	Total	452 (100)	462 (100)	905 (100)			
Age dichotomised	Young	34 (7.5)	30 (6.5)	64 (7)	0.6	-0.6	0.371 (0.542)
	Old	418 (92.5)	432 (93.5)	850 (93)	-0.6	0.6	
	Total	452 (100)	462 (100)	914 (100)			
Grade dichotomised	Low	399 (88.3)	437 (94.8)	836 (91.6)	-3.5	3.5	**12.56 (<0.001)**
	High	53 (11.7)	24 (5.2)	77 (8.4)	3.5	-3.5	
	Total	452 (100)	461 (100)	913 (100)			
Site dichotomised	Right side	204 (45.1)	243 (52.6)	447 (48.9)	-2.3	2.3	**5.095 (0.024)**
	Left side	248 (54.9)	219 (47.4)	467 (51.1)	2.3	-2.3	
	Total	452 (100)	462 (100)	914 (100)			
**Significant associations (p<0.05) highlighted in bold*							

Survival analysis of cytoplasmic CD44

Survival relating to cytoplasmic CD44 staining was assessed using Kaplan-Meier curves and Cox regression analyses. The average H-scores were split by tertiles, in order to show the difference in the wide distribution of H-scores. 3 year

survival (p 0.001) was significantly associated with positive CD44 cytoplasmic staining status. Death relating to colorectal cancer (p 0.005) was negatively significantly associated with cytoplasmic CD44 expression, as shown in figure 17, with more patients surviving CRC if they were positive for cytoplasmic CD44. Median survival was incalculable as more than 50% of patients survived beyond 3 years, and only a small proportion of patients died directly as a result of CRC. 66.9% of patients survived to 3 years with low cytoplasmic CD44, whereas 72.6% survived with medium cytoplasmic CD44 and 81.4% survived if they showed high expression. Death related to colorectal cancer showed a similar upward trend relating to increase in cytoplasmic CD44, with 77% of low expressing patients surviving to 5 years, 81.3% of medium expressing patients surviving to 5 years and 87% surviving with high cytoplasmic expression. 43.5% of low and medium expressors survived to 5 years, and 40.9% of high expressing cytoplasmic CD44 patients survived to 5 years. This is similar to what was observed in membrane-specific expression of CD44. 5 year survival was not significantly related to cytoplasmic CD44 (p 0.910), and the highest tertile crossed the other two at 50 months, similar to membranous CD44.

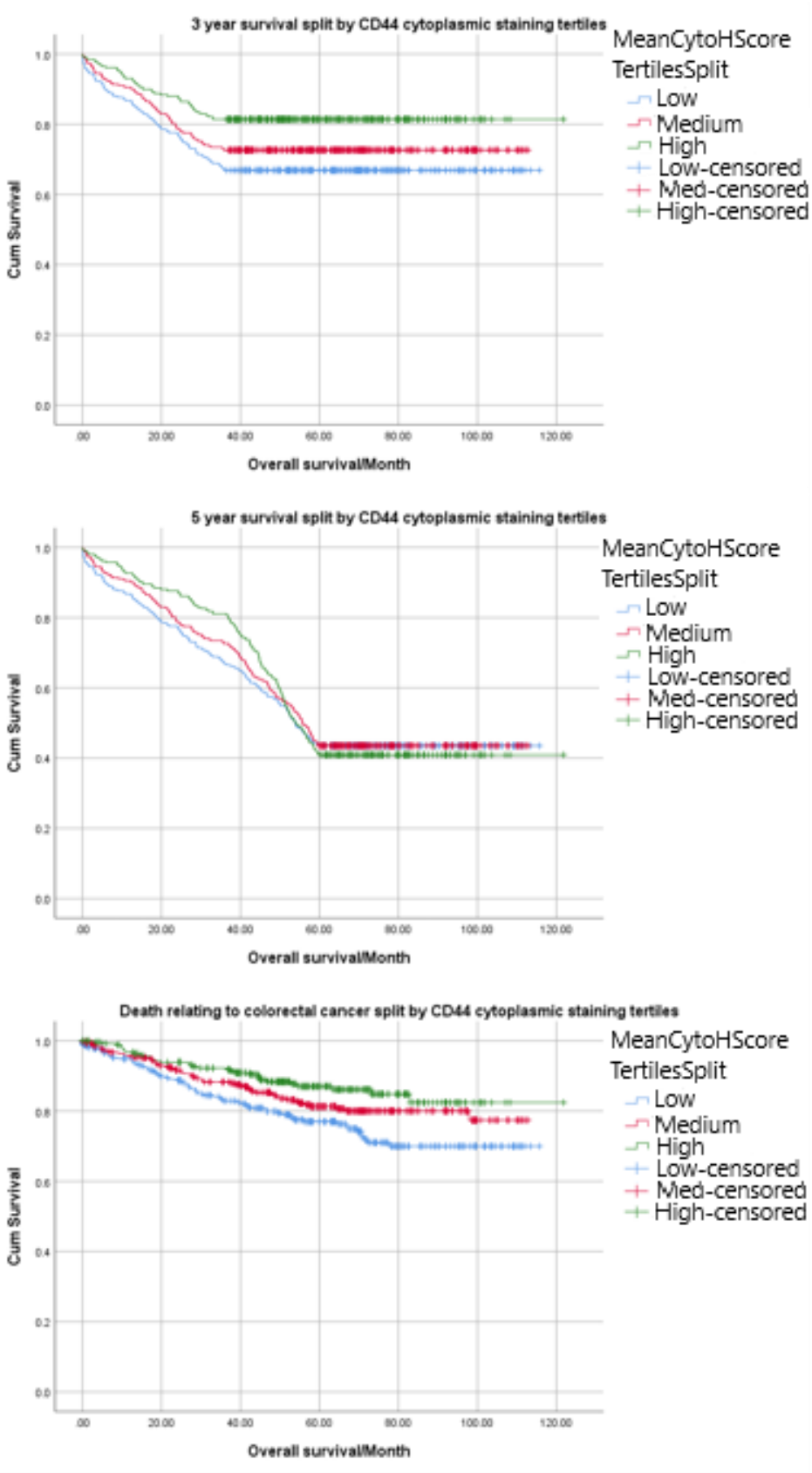


Figure 17: Kaplan-Meier survival analysis of low, medium and high expressors of cytoplasmic CD44 in relation to 3 year survival, 5 year survival and death relating to colorectal cancer.

Univariate Cox regression analysis revealed a significant association between increasing cytoplasmic H-score and survival (p 0.01). However, this trend was not present in multivariate analysis (p 0.945). Increased recurrence (p <0.001), primary tumour stage (p <0.001), lymph node metastasis (p 0.017) and grade dichotomised (p 0.013) were all associated with increased hazard ratios. Patients with MMR deficiency showed a worse survival than MMR proficient patients (p 0.01).

Table 15: Cox regression analysis of cytoplasmic CD44 staining with significantly associated variables

Variable	Hazard Ratio	95.0% CI		p value
		Lower	Upper	
Mean Cyto H-Score Tertiles Split*	0.836	0.729	0.959	**0.01**
Variables	**Hazard Ratio**	**95.0% CI**		**p value**
		Lower	**Upper**	
Synchronous metastasis	1.405	0.41	4.813	0.588
Recurrence	2.844	2.07	3.906	**<0.001**
Final Stage 1				0.239
Final Stage 2	0.815	0.101	6.588	0.848
Final Stage 3	0.491	0.062	3.916	0.502
Final Stage 4	0.519	0.029	9.218	0.655
Primary tumour T1				**<0.001**
T2	1.129	0.307	4.152	0.855
T3	1.762	0.17	18.304	0.635
T4	3.568	0.341	37.314	0.288
Lymph node metastasis	1.393	1.061	1.829	**0.017**
Metastasis	1.644	0.153	17.659	0.681
Perineural invasion	1.37	0.991	1.894	0.057
Intramural lymphovascular invasion	0.864	0.648	1.151	0.318
Tumour edge	0.759	0.554	1.039	0.086
MMR status	1.665	1.132	2.449	**0.01**
Grade Dichotomised	1.721	1.122	2.638	**0.013**
Site Dichotomised	0.946	0.72	1.244	0.693
Mean Cyto H-Score Tertiles Split	1.006	0.845	1.199	0.945

**Significant results (p <0.05) highlighted in bold*

Visual observations

Different types of staining could be observed for cytoplasmic CD44. A blanket-like staining was observed in some tumours, but in others a more granular pattern could be seen. Examples of the staining patterns are shown in figure 18. Granular staining was typically observed close to the basolateral edge of the cells, where the cells interact with the stroma, as opposed to the apical edge, facing the lumen. Given that CD44 acts as a mediator between stromal and epithelial factors, this seems to be true staining. When the staining was observed in the granular pattern, it was difficult at times to discern if it was

membranous or cytoplasmic. For the purposes of this study all granular staining was counted as cytoplasmic.

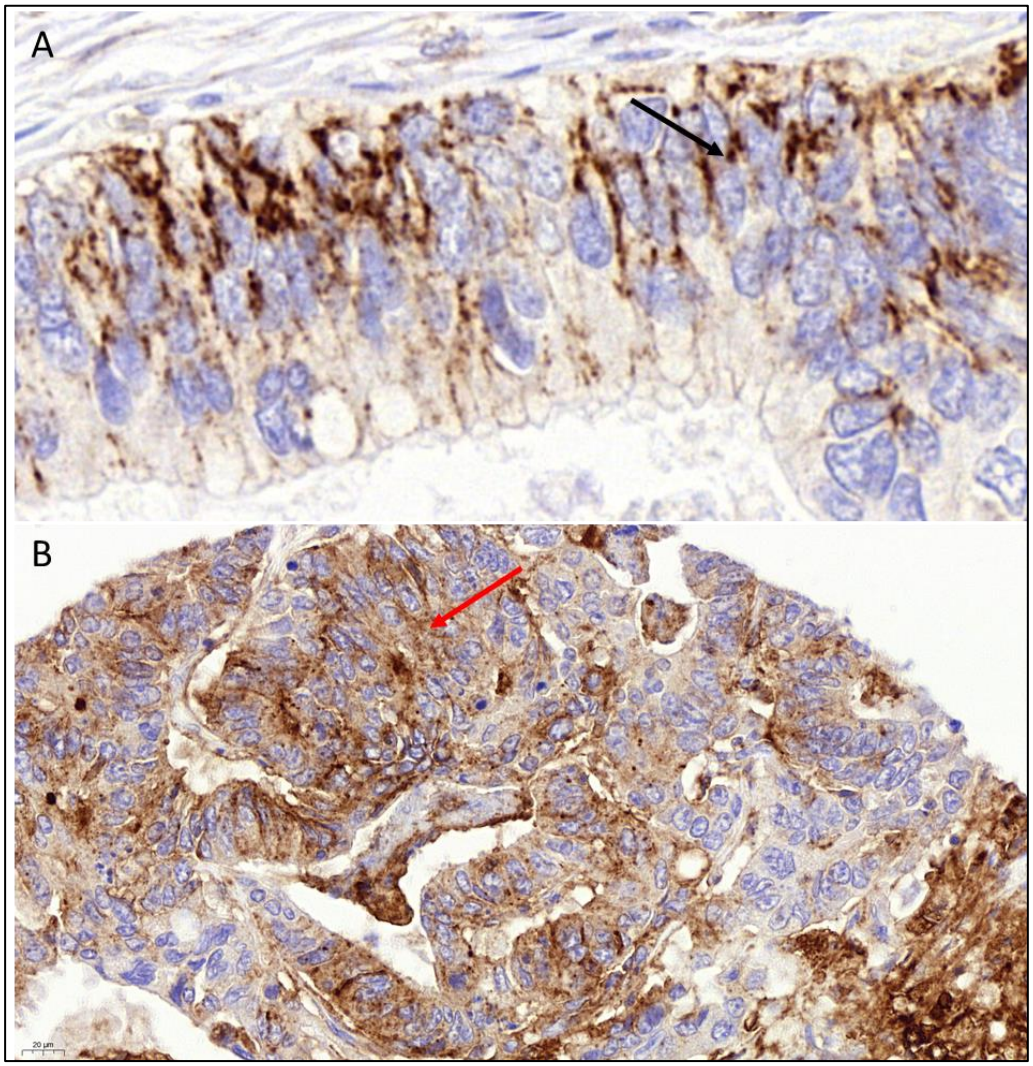


Figure 18: Cytoplasmic CD44 staining in colorectal cancer. The staining may be granular (A) or diffuse (B). Granular staining was often found near the basolateral edge of the cells near the stroma.

Nuclear staining of CD44

Nuclear CD44 staining was highly unusual and only occurred in four patients. No patients had nuclear staining in the normal-adjacent region. Because the number of patients with nuclear expression was so low, any statistical analyses would have been inconclusive and therefore were not performed.

Stromal staining of CD44

Stromal staining of CD44 was normally distributed. The mean average stromal H-score was 49.39 (standard deviation 27.72). The luminal, central and advancing edge samples were all strongly positively correlated with the average H-score (Pearson's correlation above 0.75) as seen in table 16 below.

Table 16: Distribution of CD44 stromal H-score data				
	Luminal Stroma H-Score	**Central Stroma H-Score**	**Advancing Edge Stroma H-Score**	**Mean Stroma H-Score**
Included	770	743	690	914
Excluded	230	257	310	86
Mean	49.85	49.64	47.73	49.39
Std. Deviation	33.55	34.34	34.15	27.72
Range	170	200	190	165
Median	45	45	40	47.5
Interquartile range	48	50	50	38.42
Pearson's correlation with mean stromal H-score (p value)	0.796 (<0.001)	0.786 (<0.001)	0.791 (<0.001)	-

Chi-square analysis of stromal CD44 with clinical variables

To perform Chi-squared analysis, the data were initially split by median expression, as in previous analyses. However this returned no significant associations with any of the clinical variables and therefore the average H-score data was split by tertiles (low 0-35, medium 35-60, high 60-300). The results are shown in the table below and demonstrate that 5 year survival increases with decreasing expression of CD44 (p 0.005). This was not the case for 3 year survival (p 0.367) or survival status (p 0.523). The presence of peritumoural lymphocytes was associated with increased stromal CD44 (p 0.001), confirming the expression of CD44 in lymphocytes.

Table 17: Chi-square analysis of stromal CD44 expression with clinical variables									
Clinical variable	**Condition**	**Low (%)**	**Medium (%)**	**High (%)**	**Total (%)**	**Adjusted residual**			**X^2 (p value)**
Gender	Female	145 (46)	122 (39.9)	133 (45.4)	400 (43.8)	1	-1.7	0.7	2.860 (0.239)
	Male	170 (54)	184 (60.1)	160 (54.6)	514 (56.2)	-1	1.7	-0.7	
	Total	315 (100)	306 (100)	293 (100)	914 (100)				
3 year Survival	Alive	236 (74.9)	215 (70.3)	217 (74.3)	668 (73.2)	0.9	-1.4	0.5	2.005 (0.367)
	Dead	79 (25.1)	91 (29.7)	75 (25.7)	245 (26.8)	-0.9	1.4	-0.5	
	Total	315 (100)	306 (100)	292 (100)	913 (100)				
5 year Survival*	Alive	158 (50.2)	119 (38.9)	114 (38.9)	391 (42.8)	3.3	-1.7	-1.6	**10.69 (0.005)**
	Dead	157 (49.8)	187 (61.1)	179 (61.1)	523 (57.2)	-3.3	1.7	1.6	
	Total	315 (100)	306 (100)	293 (100)	914 (100)				
Synchronous metastasis	No metastasis	280 (88.9)	263 (85.9)	269 (91.8)	812 (88.8)	0	-2	2	5.188 (0.075)
	With metastasis	35 (11.1)	43 (14.1)	24 (8.2)	102 (11.2)	0	2	-2	
	Total	315 (100)	306 (100)	293 (100)	914 (100)				
Recurrence	No Recurrence	224 (71.1)	208 (68)	225 (76.8)	657 (71.9)	-0.4	-1.9	2.3	5.900 (0.052)

	Recurrence	91 (28.9)	98 (32)	68 (23.2)	257 (28.1)	0.4	1.9	-2.3	
	Total	315 (100)	306 (100)	293 (100)	914 (100)				
Clinical variable	**Condition**	**Low (%)**	**Medium (%)**	**High (%)**	**Total (%)**	**Adjusted residual**			**X^2 (p value)**
Cause of death	Alive	202 (64.1)	184 (60.1)	187 (63.8)	573 (62.7)	0.7	-1.1	0.5	3.329 (0.767)
	CRC	53 (16.8)	57 (18.6)	48 (16.4)	158 (17.3)	-0.3	0.8	-0.5	
	Other	31 (9.8)	41 (13.4)	36 (12.3)	108 (11.8)	-1.3	1.1	0.3	
	Unknown	29 (9.2)	24 (7.8)	22 (7.5)	75 (8.2)	0.8	-0.3	-0.5	
	Total	315 (100)	306 (100)	293 (100)	914 (100)				
Survival Status	Alive	202 (64.1)	184 (60.1)	187 (63.8)	573 (62.7)	0.7	-1.1	0.5	1.296 (0.523)
	Dead	113 (35.9)	122 (39.9)	106 (36.2)	341 (37.3)	-0.7	1.1	-0.5	
	Total	315 (100)	306 (100)	293 (100)	914 (100)				
Final stage	1	45 (14.3)	47 (15.4)	55 (18.8)	147 (16.1)	-1.1	-0.4	1.5	9.509 (0.147)
	2	128 (40.6)	118 (38.6)	130 (44.4)	376 (41.1)	-0.2	-1.1	1.4	
	3	108 (34.3)	98 (32)	84 (28.7)	290 (31.7)	1.2	0.1	-1.4	
	4	34 (10.8)	43 (14.1)	24 (8.2)	101 (11.1)	-0.2	2.1	-1.9	
	Total	315 (100)	306 (100)	293 (100)	914 (100)				
Primary tumour	T1	19 (6)	22 (7.2)	26 (8.9)	67 (7.3)	-1.1	-0.1	1.2	5.764 (0.45)
	T2	33 (10.5)	31 (10.1)	34 (11.6)	98 (10.7)	-0.2	-0.4	0.6	
	T3	180 (57.1)	154 (50.3)	154 (52.6)	488 (53.4)	1.6	-1.3	-0.3	
	T4	83 (26.3)	99 (32.4)	79 (27)	261 (28.6)	-1.1	1.8	-0.7	
	Total	315 (100)	306 (100)	293 (100)	914 (100)				
Lymph node metastasis	N0	172 (56.8)	172 (57.3)	184 (64.6)	528 (59.5)	-1.2	-0.9	2.1	4.953 (0.292)
	N1	76 (25.1)	79 (26.3)	61 (21.4)	216 (24.3)	0.4	1	-1.4	
	N2	55 (18.2)	49 (16.3)	40 (14)	144 (16.2)	1.1	0.1	-1.2	
	Total	303 (100)	300 (100)	285 (100)	888 (100)				
Metastasis	No metastasis	280 (88.9)	263 (85.9)	269 (91.8)	812 (88.8)	0	-2	2	5.188 (0.075)
	With metastasis	35 (11.1)	43 (14.1)	24 (8.2)	102 (11.2)	0	2	-2	
	Total	315 (100)	306 (100)	293 (100)	914 (100)				
Vascular invasion	Absent	158 (51)	151 (50)	156 (54.4)	465 (51.7)	-0.3	-0.7	1.1	1.226 (0.542)
	Present	152 (49)	151 (50)	131 (45.6)	434 (48.3)	0.3	0.7	-1.1	
	Total	310 (100)	302 (100)	287 (100)	899 (100)				

Clinical variable	Condition	Low (%)	Medium (%)	High (%)	Total (%)	Adjusted residual			X^2 (p value)
Perineural invasion	Absent	250 (82.5)	237 (80.3)	224 (79.2)	711 (80.7)	1	-0.2	-0.8	1.096 (0.578)
	Present	53 (17.5)	58 (19.7)	59 (20.8)	170 (19.3)	-1	0.2	0.8	
	Total	303 (100)	295 (100)	283 (100)	881 (100)				
Intramural lymphovascular invasion	Absent	192 (66.4)	197 (70.6)	197 (70.4)	586 (69.1)	-1.2	0.7	0.6	1.466 (0.481)
	Present	97 (33.6)	82 (29.4)	83 (29.6)	262 (30.9)	1.2	-0.7	-0.6	
	Total	289 (100)	279 (100)	280 (100)	848 (100)				
Tumour edge	Infiltrative	115 (51.1)	115 (48.3)	118 (52.4)	348 (50.6)	0.2	-0.9	0.7	0.825 (0.662)
	Pushing	110 (48.9)	123 (51.7)	107 (47.6)	340 (49.4)	-0.2	0.9	-0.7	
	Total	225 (100)	238 (100)	225 (100)	688 (100)				
Tumour budding score	Low	154 (69.7)	153 (65.4)	149 (66.8)	456 (67.3)	0.9	-0.8	-0.2	0.983 (0.612)
	High	67 (30.3)	81 (34.6)	74 (33.2)	222 (32.7)	-0.9	0.8	0.2	
	Total	221 (100)	234 (100)	223 (100)	678 (100)				
Peritumoural lymphocytes	Inconspicuous	176 (78.2)	178 (73.9)	140 (62.2)	494 (71.5)	2.7	1	-3.7	**15.14 (0.001)**
	Conspicuous	49 (21.8)	63 (26.1)	85 (37.8)	197 (28.5)	-2.7	-1	3.7	
	Total	225 (100)	241 (100)	225 (100)	691 (100)				
MMR Status	Proficient	262 (84.2)	266 (87.8)	229 (78.7)	757 (83.6)	0.4	2.4	-2.8	**9.099 l0.011**
	Deficient	49 (15.8)	37 (12.2)	62 (21.3)	148 (16.4)	-0.4	-2.4	2.8	
	Total	315 (100)	306 (100)	293 (100)	905 (100)				
Age dichotomised	Young	20 (6.3)	24 (7.8)	20 (6.8)	64 (7)	-0.6	0.7	-0.1	0.553 (0.759)
	Old	295 (93.7)	282 (92.2)	273 (93.2)	850 (93)	0.6	-0.7	0.1	
	Total	315 (100)	306 (100)	293 (100)	914 (100)				
Grade dichotomised	Low	288 (91.4)	281 (92.1)	267 (91.1)	836 (91.6)	-0.1	0.4	-0.3	0.207 (0.902)
	High	27 (8.6)	24 (7.9)	26 (8.9)	77 (8.4)	0.1	-0.4	0.3	
	Total	315 (100)	305 (100)	293 (100)	913 (100)				
Site dichotomised	Right side	157 (49.8)	149 (48.7)	141 (48.1)	447 (48.9)	0.4	-0.1	-0.3	0.188 (0.91)
	Left side	158 (50.2)	157 (51.3)	152 (51.9)	467 (51.1)	-0.4	0.1	0.3	
	Total	315 (100)	306 (100)	293 (100)	914 (100)				

**Significant associations (p <005) highlighted in bold*

Survival analysis of stromal CD44

Kaplan-Meier and Cox regression analyses were performed to further understand the relationship between stromal CD44 staining and survival. Kaplan-Meier analysis was performed with the CD44 stromal staining split into tertiles, looking at 3 year survival, 5 year survival and death related to colorectal cancer, as seen in figure 19. 3 year survival was not significantly associated with stromal CD44 expression (p 0.443), and neither was death relating to colorectal cancer (p 0.672). 5 year survival could not be calculated for patients expressing low levels of stromal CD44, and did not differ much between medium (53.9 months) and high expression (51.8 months) but was significantly different (p 0.022). Univariate Cox regression analysis did not show a significant association between stromal CD44 and overall survival (p 0.475). Multivariate analysis revealed a slightly stronger but still not significant association (p 0.363). MMR status was not significantly associated with survival (p 0.102) but the presence of peritumoural lymphocytes led to worse outcome (hazard ratio 0.569, p <0.001).

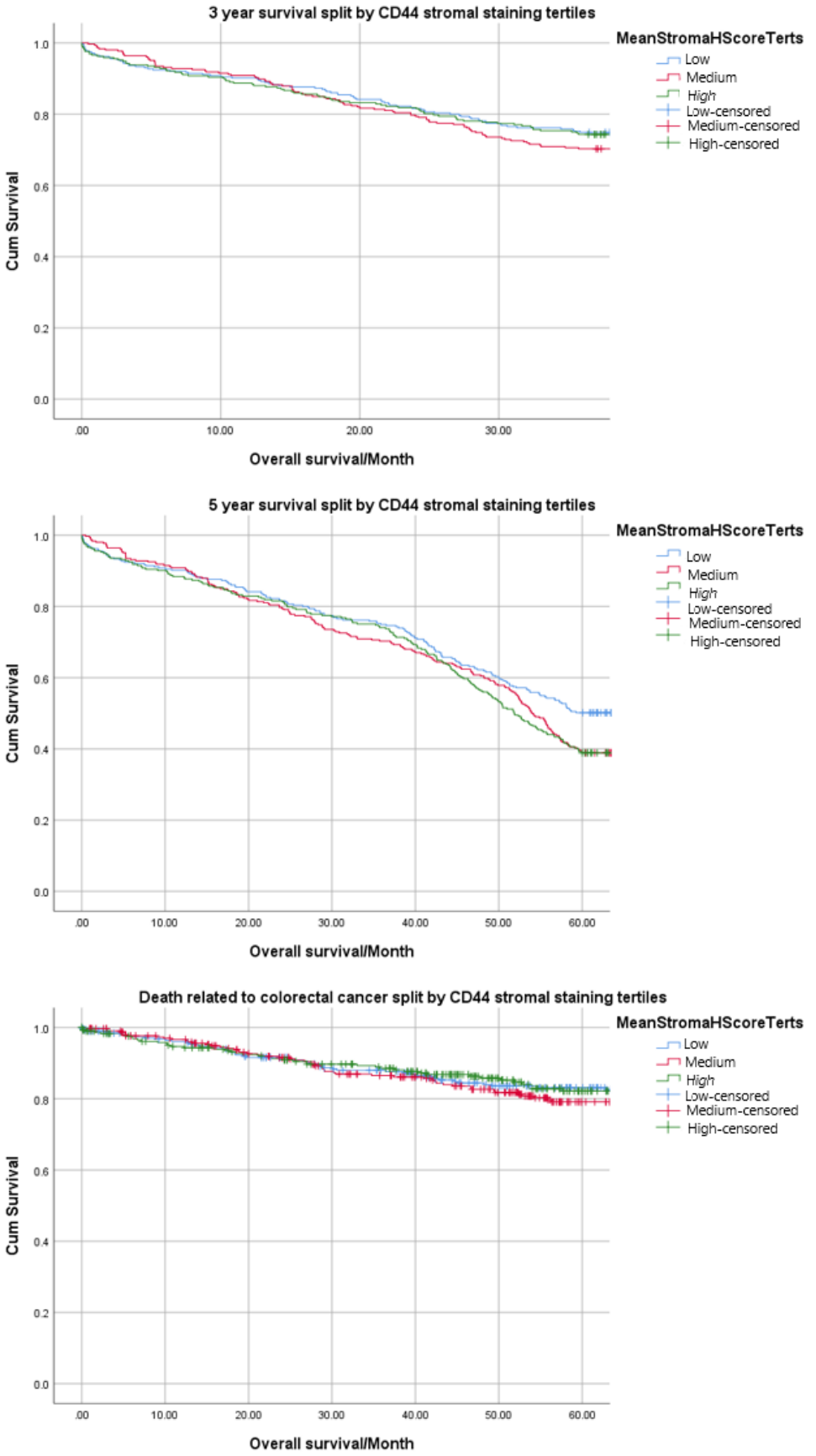


Figure 19: Kaplan-Meier analysis of stromal CD44 expression revealed a significant association with 5 year survival, but not with 3 year survival or death relating to CRC.

CD44 and c-Myc interaction

c-Myc and CD44 are known to share common pathways and interactions in colorectal cancer. Chi-square analysis showed that the patients with the highest expression of CD44 are associated with the highest expression of c-Myc ($p < 0.001$). The data of both were then combined (see the method used to combine c-Myc nuclear and cytoplasmic staining), using the median cut-off for nuclear c-Myc expression and the 75th percentile for CD44 expression. Good 5 year survival was associated with c-Myc-/CD44- patients ($p < 0.001$), whereas the lowest primary tumour stage was associated with c-Myc+/CD44+ patients. A lack of metastasis was associated with high CD44 expression, regardless of c-Myc expression (p 0.018). The presence of vascular invasion was most strongly associated with patients who were c-Myc-/CD44- ($p < 0.001$). Right sided tumours showed a high affiliation with cMyc-/CD44+, whereas left-sided tumours were associated with c-Myc+/CD44- expression (p 0.002). These associations reveal that although CD44 and c-Myc do share some common pathways, they also act independently, leading to different clinical outcomes.

Table 18: Nuclear c-Myc expression and Membranous CD44 expression combined with clinical variables in Chi-square analysis

Clinical variable	Condition	c-Myc-/CD44-	c-Myc-/CD44+	c-Myc+/CD44-	c-Myc+/CD44+	Total	Adjusted residuals				X^2 (p value)
5 year survival*	Alive	217 (51.5)	35 (50)	126 (40.9)	40 (26)	418 (43.9)	4.3	1.1	-1.3	-4.9	**32.26 (<0.001)**
	Dead	204 (48.5)	35 (50)	182 (59.1)	114 (74)	535 (56.1)	-4.3	-1.1	1.3	4.9	
	Total	421 (100)	70 (100)	308 (100)	154 (100)	953 (100)					
Synchronous metastasis	No metastasis	366 (86.9)	68 (97.1)	267 (86.7)	143 (92.9)	844 (88.6)	-1.4	2.3	-1.3	1.8	**10.06 (0.018)**
	With metastasis	55 (13.1)	2 (2.9)	41 (13.3)	11 (7.1)	109 (11.4)	1.4	-2.3	1.3	-1.8	
	Total	421 (100)	70 (100)	308 (100)	154 (100)	953 (100)					
Final stage	1	56 (13.3)	16 (22.9)	47 (15.3)	36 (23.4)	155 (16.3)	-2.2	1.6	-0.6	2.6	**20.15 (0.017)**
	2	178 (42.3)	33 (47.1)	122 (39.6)	56 (36.4)	389 (40.8)	0.8	1.1	-0.5	-1.2	
	3	132 (31.4)	19 (27.1)	99 (32.1)	51 (33.1)	301 (31.6)	-0.1	-0.8	0.3	0.4	
	4	55 (13.1)	2 (2.9)	40 (13)	11 (7.1)	108 (11.3)	1.5	-2.3	1.1	-1.8	
	Total	421 (100)	70 (100)	308 (100)	154 (100)	953 (100)					

Clinical variable	Condition	c-Myc-/CD44-	c-Myc-/CD44+	c-Myc+/CD44-	c-Myc+/CD44+	Total	Adjusted residuals				X^2 (p value)
Primary tumour	T1	19 (4.5)	7 (10)	20 (6.5)	24 (15.6)	70 (7.3)	-3	0.9	-0.7	4.3	**26.27 (0.002)**
	T2	45 (10.7)	11 (15.7)	30 (9.7)	17 (11)	103 (10.8)	-0.1	1.4	-0.7	0.1	
	T3	222 (52.7)	35 (50)	172 (55.8)	77 (50)	506 (53.1)	-0.2	-0.5	1.2	-0.8	
	T4	135 (32.1)	17 (24.3)	86 (27.9)	36 (23.4)	274 (28.8)	2	-0.9	-0.4	-1.6	
	Total	421 (100)	70 (100)	308 (100)	154 (100)	953 (100)					
Metastasis	No metastasis	366 (86.9)	68 (97.1)	267 (86.7)	143 (92.9)	844 (88.6)	-1.4	2.3	-1.3	1.8	**10.06 (0.018)**
	With metastasis	55 (13.1)	2 (2.9)	41 (13.3)	11 (7.1)	109 (11.4)	1.4	-2.3	1.3	-1.8	
	Total	421 (100)	70 (100)	308 (100)	154 (100)	953 (100)					
Vascular invasion	Absent	191 (45.6)	48 (70.6)	155 (50.7)	89 (61.4)	483 (51.5)	-3.3	3.3	-0.4	2.6	**21.54 (<0.001)**
	Present	228 (54.4)	20 (29.4)	151 (49.3)	56 (38.6)	455 (48.5)	3.3	-3.3	0.4	-2.6	
	Total	419 (100)	68 (100)	306 (100)	145 (100)	938 (100)					
Tumour edge	Infiltrative	144 (55.6)	18 (37.5)	161 (55.9)	49 (37.7)	372 (51.3)	1.7	-2	2	-3.4	**17.65 (0.001)**
	Pushing	115 (44.4)	30 (62.5)	127 (44.1)	81 (62.3)	353 (48.7)	-1.7	2	-2	3.4	
	Total	259 (100)	48 (100)	288 (100)	130 (100)	725 (100)					
Peritumoural lymphocytes	Inconspicuous	191 (73.5)	38 (77.6)	210 (72.7)	80 (61.1)	519 (71.2)	1	1	0.7	-2.8	**8.471 (0.037)**
	Conspicuous	69 (26.5)	11 (22.4)	79 (27.3)	51 (38.9)	210 (28.8)	-1	-1	-0.7	2.8	
	Total	260 (100)	49 (100)	289 (100)	131 (100)	729 (100)					
MMR Status	Proficient	372 (89.4)	46 (66.7)	266 (86.6)	107 (70.4)	791 (83.8)	4.2	-4	1.7	-4.9	**46.54 (<0.001)**
	Deficient	44 (10.6)	23 (33.3)	41 (13.4)	45 (29.6)	153 (16.2)	-4.2	4	-1.7	4.9	
	Total	416 (100)	69 (100)	307 (100)	152 (100)	944 (100)					
Site dichotomised	Right side	210 (49.9)	45 (64.3)	127 (41.2)	81 (52.6)	463 (48.6)	0.7	2.7	-3.1	1.1	**14.84 (0.002)**

Clinical variable	Condition	c-Myc-/CD44-	c-Myc-/CD44+	c-Myc+/CD44-	c-Myc+/CD44+	Total	Adjusted residuals				X^2 (p value)
	Left side	211 (50.1)	25 (35.7)	181 (58.8)	73 (47.4)	490 (51.4)	-0.7	-2.7	3.1	-1.1	
	Total	421 (100)	70 (100)	308 (100)	154 (100)	953 (100)					
**Significant results (p <0.05) highlighted in bold*											

Tumour stroma content as a biomarker

Tumour stromal content is considered an important biomarker in colorectal cancer. Stromal content was determined as the percentage of the core occupied by stromal cells rather than epithelial cells. The consensus molecular subtype CMS4 tumours are characterised by a high stromal content [26]. Tumour epithelium content was assessed prior to evaluating CD44 and c-Myc staining, and stromal content was calculated by taking the percentage of epithelium away from 100. The average tumour stromal content was normally distributed as seen in figure 20, with a mean of 49.39 (standard deviation 27.7). The data was split by median expression (47.5) in keeping with the previous analyses of c-Myc and CD44 staining, and the association with clinical variables assessed by Chi-square analysis.

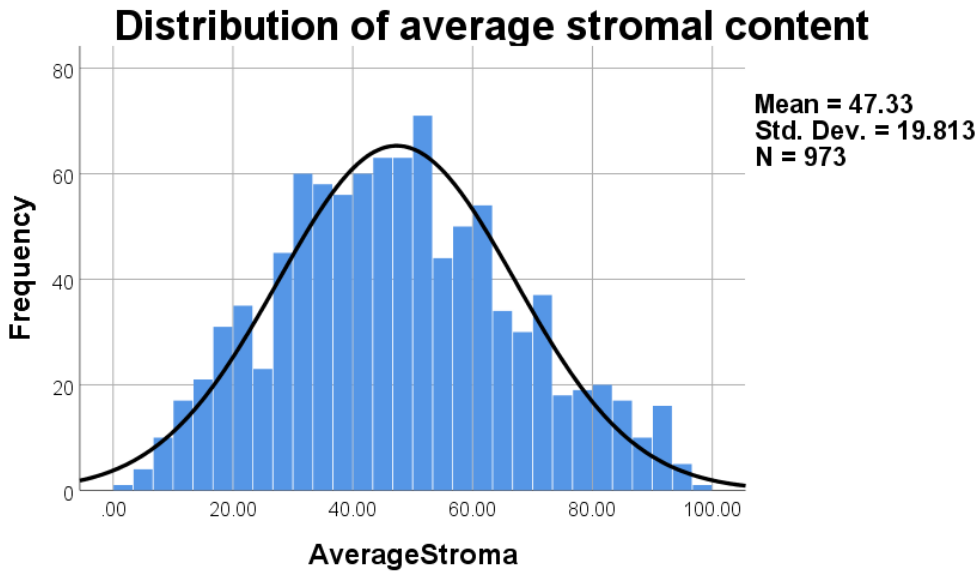


Figure 20: Distribution of average stromal content in CD44 and c-Myc stained slides. Average stromal content was calculated from average epithelial content and shows a normal distribution.

Chi-square analysis of tumour stroma content

Tumour stromal content was compared to the clinical data using a median cut-off. High tumour stromal content was associated with increased metastasis (p 0.013), increased primary tumour stage (p <0.001) and increased lymph node metastasis (p 0.016). Increased final stage (p 0.002) was associated with high stromal content, and so it is unsurprising that alongside these associations,

recurrence (p <0.001), poor 3 year survival (p 0.015) and poor survival status (p 0.005) are also observed. Death related to colorectal cancer was associated with high stromal content (p 0.013). Vascular invasion (p 0.003), perineural invasion (p 0.004) and an infiltrative tumour edge (p 0.002) were all associated with high stromal content, as seen in table 19.

Table 19: Chi-square analysis of tumour stromal content with clinical variables

Clinical variable	Condition	Low	High	Total	Adjusted residual		X^2 (p value)
Gender*	Female	239 (48.4)	182 (38)	421 (43.3)	3.3	-3.3	**10.68 (0.001)**
	Male	255 (51.6)	297 (62)	552 (56.7)	-3.3	3.3	
	Total	494 (100)	479 (100)	973 (100)			
3 year Survival	Alive	377 (76.5)	333 (69.5)	710 (73)	2.4	-2.4	**5.961 (0.015)**
	Dead	116 (23.5)	146 (30.5)	262 (27)	-2.4	2.4	
	Total	493 (100)	479 (100)	972 (100)			
5 year Survival	Alive	225 (45.5)	200 (41.8)	425 (43.7)	1.2	-1.2	1.422 (0.233)
	Dead	269 (54.5)	279 (58.2)	548 (56.3)	-1.2	1.2	
	Total	494 (100)	479 (100)	973 (100)			
Synchronous metastasis	No metastais	448 (90.7)	412 (86)	860 (88.4)	2.3	-2.3	**5.180 (0.023)**
	With metastasis	46 (9.3)	67 (14)	113 (11.6)	-2.3	2.3	
	Total	494 (100)	479 (100)	973 (100)			
Recurrence	No Recurrence	380 (76.9)	315 (65.8)	695 (71.4)	3.9	-3.9	**14.84 (<0.001)**
	Recurrence	114 (23.1)	164 (34.2)	278 (28.6)	-3.9	3.9	
	Total	494 (100)	479 (100)	973 (100)			
Cause of death	Alive	330 (66.8)	278 (58)	608 (62.5)	2.8	-2.8	**10.77 (0.013)**
	CRC	72 (14.6)	101 (21.1)	173 (17.8)	-2.7	2.7	
	Other	60 (12.1)	57 (11.9)	117 (12)	0.1	-0.1	
	Unknown	32 (6.5)	43 (9)	75 (7.7)	-1.5	1.5	
	Total	494 (100)	479 (100)	973 (100)			
Survival Status	Alive	330 (66.8)	278 (58)	608 (62.5)	2.8	-2.8	**7.969 (0.005)**
	Dead	164 (33.2)	201 (42)	365 (37.5)	-2.8	2.8	
	Total	494 (100)	479 (100)	973 (100)			
Final stage	1	95 (19.2)	63 (13.2)	158 (16.2)	2.6	-2.6	**14.57 (0.002)**
	2	212 (42.9)	183 (38.2)	395 (40.6)	1.5	-1.5	

	3	142 (28.7)	166 (34.7)	308 (31.7)	-2	2	
	4	45 (9.1)	67 (14)	112 (11.5)	-2.4	2.4	
	Total	494 (100)	479 (100)	973 (100)			
Clinical variable	**Condition**	**Low**	**High**	**Total**	**Adjusted residual**		**X^2 (p value)**
Primary tumour	T1	47 (9.5)	25 (5.2)	72 (7.4)	2.6	-2.6	**19.58 (<0.001)**
	T2	58 (11.7)	46 (9.6)	104 (10.7)	1.1	-1.1	
	T3	274 (55.5)	241 (50.3)	515 (52.9)	1.6	-1.6	
	T4	115 (23.3)	167 (34.9)	282 (29)	-4	4	
	Total	494 (100)	479 (100)	973 (100)			
Lymph node metastasis	N0	303 (63.5)	257 (54.7)	560 (59.1)	2.8	-2.8	**8.298 (0.016)**
	N1	100 (21)	131 (27.9)	231 (24.4)	-2.5	2.5	
	N2	74 (15.5)	82 (17.4)	156 (16.5)	-0.8	0.8	
	Total	477 (100)	470 (100)	947 (100)			
Metastasis	No metastasis	449 (90.9)	411 (85.8)	860 (88.4)	2.5	-2.5	**6.131 (0.013)**
	With metastasis	45 (9.1)	68 (14.2)	113 (11.6)	-2.5	2.5	
	Total	494 (100)	479 (100)	973 (100)			
Vascular invasion	Absent	272 (56.3)	222 (46.7)	494 (51.6)	3	-3	**8.796 (0.003)**
	Present	211 (43.7)	253 (53.3)	464 (48.4)	-3	3	
	Total	483 (100)	475 (100)	958 (100)			
Perineural invasion	Absent	398 (84.5)	361 (77.1)	759 (80.8)	2.9	-2.9	**8.216 (0.004)**
	Present	73 (15.5)	107 (22.9)	180 (19.2)	-2.9	2.9	
	Total	471 (100)	468 (100)	939 (100)			
Intramural lymphovascular invasion	Absent	322 (69.2)	302 (69.1)	624 (69.2)	0	0	0.002 (0.964)
	Present	143 (30.8)	135 (30.9)	278 (30.8)	0	0	
	Total	465 (100)	437 (100)	902 (100)			
Tumour edge	Infiltrative	170 (45.9)	212 (57.3)	382 (51.6)	-3.1	3.1	**9.545 (0.002)**
	Pushing	200 (54.1)	158 (42.7)	358 (48.4)	3.1	-3.1	
	Total	370 (100)	370 (100)	740 (100)			
Tumour budding score	Low	253 (68.9)	240 (66.3)	493 (67.6)	0.8	-0.8	0.580 (0.446)
	High	114 (31.1)	122 (33.7)	236 (32.4)	-0.8	0.8	
	Total	367 (100)	362 (100)	729 (100)			

Clinical variable	Condition	Low	High	Total	Adjusted residual	X^2 (p value)	Clinical variable
Peritumoural lymphocytes	Inconspicuous	270 (72.4)	260 (70.1)	530 (71.2)	0.7	-0.7	0.482 (0.487)
	Conspicuous	103 (27.6)	111 (29.9)	214 (28.8)	-0.7	0.7	
	Total	373 (100)	371 (100)	744 (100)			
MMR Status		5 (1)	12 (2.5)	17 (1.7)	-1.8	1.8	**7.006 (0.03)**
	Proficient	398 (80.6)	402 (83.9)	800 (82.2)	-1.4	1.4	
	Deficient	91 (18.4)	65 (13.6)	156 (16)	2.1	-2.1	
	Total	494 (100)	479 (100)	973 (100)			
Age dichotomised	Young	31 (6.3)	32 (6.7)	63 (6.5)	-0.3	0.3	0.066 (0.797)
	Old	463 (93.7)	447 (93.3)	910 (93.5)	0.3	-0.3	
	Total	494 (100)	479 (100)	973 (100)			
Grade dichotomised	Low	448 (90.7)	437 (91.4)	885 (91)	-0.4	0.4	0.161 (0.688)
	High	46 (9.3)	41 (8.6)	87 (9)	0.4	-0.4	
	Total	494 (100)	478 (100)	972 (100)			
Site dichotomised	Right side	253 (51.2)	220 (45.9)	473 (48.6)	1.6	-1.6	2.720 (0.099)
	Left side	241 (48.8)	259 (54.1)	500 (51.4)	-1.6	1.6	
	Total	494 (100)	479 (100)	973 (100)			
**Significant associations (p <0.05) highlighted in bold*							

Survival analysis of tumour stroma content

Survival of patients relating to tumour stromal content was assessed using the median cut-off. 3 year survival was significantly higher (p 0.017) in patients with low stromal content, with 76.5% of patients surviving to 3 years and 69.5% of patients with high stromal content surviving to the same time. Deaths relating to CRC were significantly lower in patients with low stromal content (p 0.005), with an 84.7% chance of surviving versus a 77% chance of survival in patients with high stromal content. 5 year survival was not significantly associated with tumour stroma content (p 0.110) but showed the same trend, as shown in figure 21.

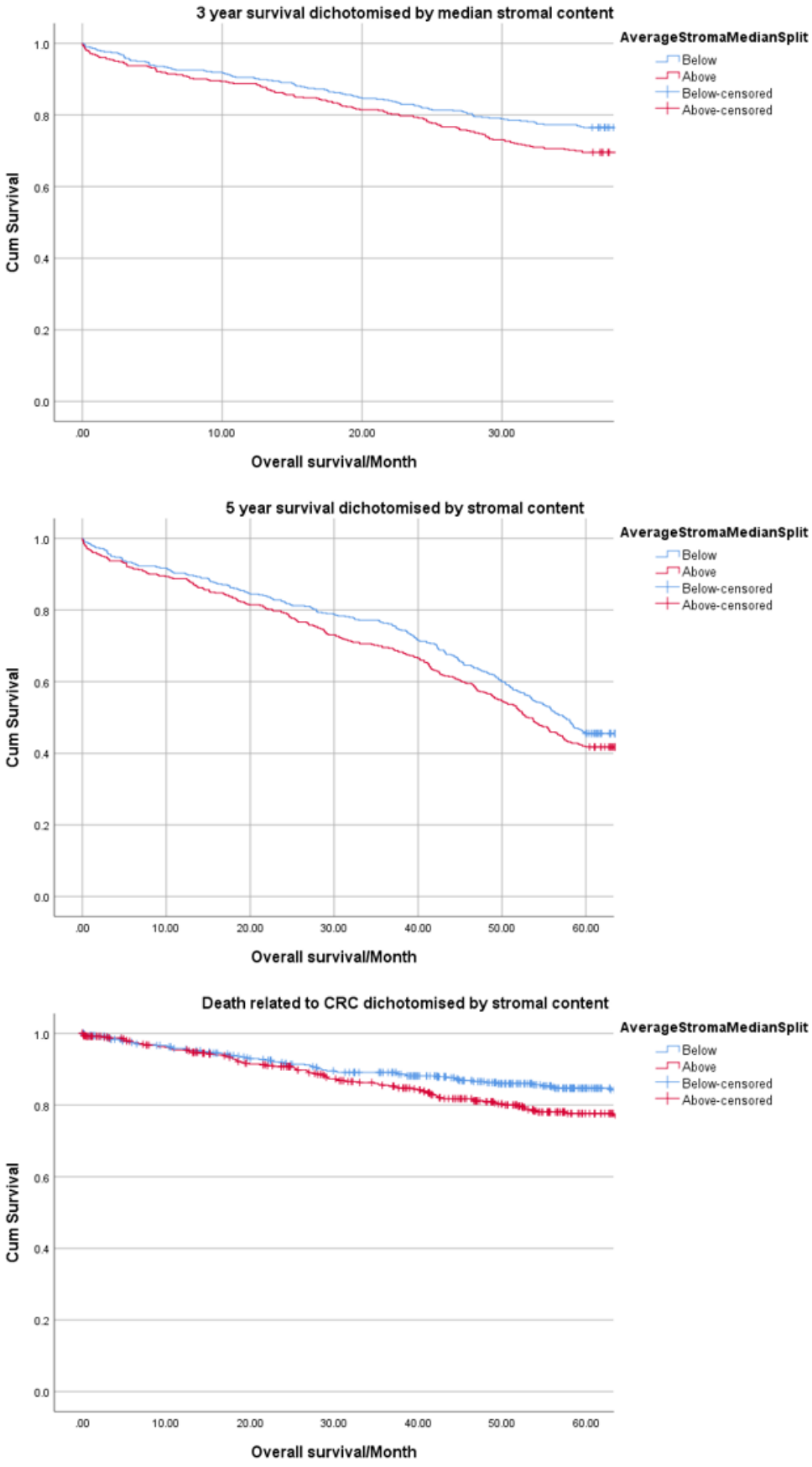


Figure 21: Kaplan-Meier analysis of tumour stroma with 3 year survival, 5 year survival and death relating to CRC. Only 5 year survival was not significantly associated with survival.

Association with biomarkers

Mann-Whitney U analysis was performed to assess the relationship between tumour stroma content and c-Myc and CD44 staining. c-Myc showed no association with stromal content (p 0.342), and the mean stromal content was very similar between positive (62.22, 95% CI 57-67) and negative (64.92, 95% CI 60-70) nuclear expression groups. CD44 however showed a strong relationship with stromal content (p <0.001), with low stromal content associated with high CD44 membranous staining. The mean average stromal content for low CD44 staining was 42.26% (95% CI 38-47), whereas the mean stromal content in CD44 high tumours was 28.69 (95% CI 25-33). When compared via multivariate Cox regression analysis it was clear that CD44 was not significantly associated with survival (p 0.897) whereas low stromal content was (p 0.020).

Bioinformatic analysis of c-Myc and CD44 in the Cancer Genome Atlas

c-Myc analysis

In order to further understand the importance of c-Myc expression in colorectal cancer, TCGA colorectal cancer data was assessed for differential expression of genes associated with c-Myc. The log2 fold change of all genes was assessed in conjunction with c-Myc. Of the 60488 genes included, 9984 had a log2 fold change of ±1. 8247 genes were significantly changed ($p<0.05$), and of those, 7872 were changed with an adjusted p value <0.05. This equates to approximately 13% of the genome, and is similar to published figures concerning c-Myc regulation. 11 genes were significantly downregulated, and the other 7861 were upregulated, demonstrating the involvement of c-Myc in stimulating gene expression rather than repressing it. Pathways significantly associated with c-Myc expression are shown in figure 22.

Pathway analysis was performed in WebGestalt using KEGG, Panther and Reactome databases, as described in figure 6. 169 unique genes were shown to be significantly upregulated as part of pathways across KEGG, Panther and Reactome databases. 34 histone genes were found to be upregulated in the presence of c-Myc, and were seen across multiple pathways including Reactome pathways “pre-NOTCH transcription and translation”, and “formation of the β-catenin:TCF transactivating complex”, or KEGG pathway “Systemic lupus erythematosus”. This is in accordance with the β-catenin:TCF interaction regulating c-Myc overexpression in colorectal cancer [39]. The Panther “Wnt signalling” pathway was also upregulated. In addition, multiple pathways associated with metabolic processes including the “Ionotropic glutamate receptor pathway” and “vitamin D metabolism” and pathway were associated with c-Myc expression, demonstrating the role of c-Myc in metabolism and glycolysis in colorectal cancer. The “Cadherin” pathway, which includes TCF/LEF, EGFR and Wnt, was also significantly associated with c-Myc expression. 25 olfactory receptor genes were upregulated, such as OR51B4 and

OR2C3, which have been associated with cancer in the past [86, 87]. Of the 11 genes which were downregulated, EYA1 is known to be repressed by Wnt signalling [88], and SFTPC downregulation has been linked to increased proliferation and progression in lung cancer [89].

CD44 analysis

TCGA colorectal cancer data was assessed for differential expression of genes associated with CD44 in the same way as for c-Myc. All genes were assessed for the log2 fold change of their expression in conjunction with CD44. Of the 60488 genes included, 2268 had a log2 fold change of ±1. 1851 genes were significantly changed ($p<0.05$), and of those, 1693 showed differential expression with an adjusted p value <0.05. 74 genes showed downregulation and the other 1619 were upregulated. The upregulated genes were analysed using WebGestalt to determine the number of pathways significantly upregulated.

Out of the top 10 KEGG, Panther and Reactome pathways, 87 unique genes were found to be upregulated. No Panther pathways had a significant FDR score (<0.05), and only 4 KEGG pathways did. The pathways that were enriched were associated with cholesterol metabolism, bile secretion, and fat digestion, among others. These pathways are associated with general gut processes [90], but are also upregulated in colorectal cancer stem cells [91, 92]. The pathway "Transcriptional regulation of pluripotent stem cells" specifically was upregulated, demonstrating this. All of these pathways are reminiscent of CD44 in stem cell initiation and maintenance [71, 73, 93]. The "Melanogenesis" pathway was also upregulated, and genes such as PRKACG and WNT7A, known to be upregulated in CRC, were overexpressed [94, 95]. Additional genes such as CALML5 and CREB3L3 were also upregulated in this pathway, but have not been associated with colorectal cancer before.

Analysis of differential expression of genes in the presence of CD44 and c-Myc in colorectal cancer in this way allowed an additional opportunity to understand their functions in this disease, and how they can act as possible targets for therapeutic agents. For example, disrupting the proliferation of cancer cells via suppression of c-Myc.

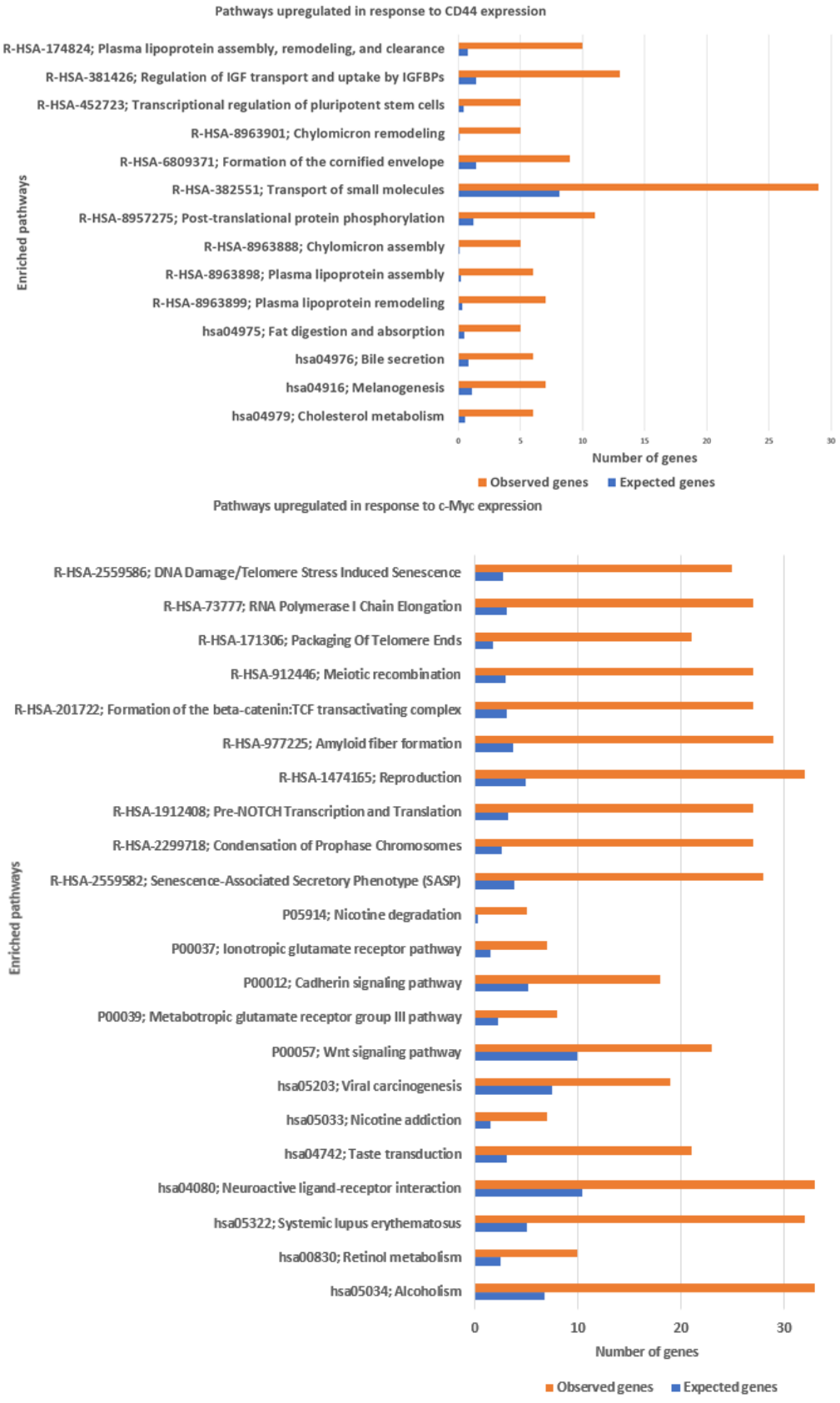


Figure 22: Pathways upregulated with regard to CD44 (upper) and c-Myc (lower) expression. Expected numbers of genes are in blue, observed number of genes in orange. No upregulated pathways were shared between CD44 and c-Myc. A higher number of pathways upregulated in the presence of c-Myc than in CD44.

Discussion

This project was an opportunity to assess a large cohort of patients for c-Myc and CD44 expression, as well as tumour stroma content. As the samples were in TMA form, it was possible to assess heterogeneity of CD44 and c-Myc in colorectal cancer. High correlation of both proteins between the tumour regions showed that they were homogenously expressed. Histological analysis is highly subjective, but performing inter and intra-observer comparisons demonstrated that the data were robust. CD44 and c-Myc are most commonly expressed in the tumour epithelial membrane and nucleus respectively, but this study did not stop at these cellular compartments. By interrogating analysis in all cellular compartments where staining was present, including in stromal cells, further understanding of the clinical significance of these biomarkers in colorectal cancer could be obtained. An additional limitation of this study was the confounding effects of MMR status. By separating the mismatch repair proficient and deficient patients, it was possible to see the clinical significance of each protein without interference.

c-Myc tumoural expression

c-Myc is a well-studied protein in colorectal cancer, with functions in multiple pathways that lead to increased pathogenicity. It is associated with the CMS2 molecular subtype and overexpression of the Wnt signalling pathway. The staining was assessed using the Y69 clone antibody, which is specific to the N-terminal domain of c-Myc and should not react with the paralogues n-Myc and l-Myc (alignments checked using Uniprot, 15th September 2020) [96].

As c-Myc is involved in transcriptional regulation, its expression is expected in the nucleus. When the nuclear expression of c-Myc was split into positive and negative expression, positive c-Myc expression was associated with poor 5 year survival, which was emphasised in MMR proficient patients. Death relating to colorectal cancer was significantly higher in c-Myc positive patients, even accounting for clinical variables. Cytoplasmic staining was generally weak and only associated with low lymph node status, absence of perineural invasion and left-sided tumours. It is possible that much of the c-Myc cytoplasmic staining was artefact, as the associations found were not strongly significant and were consistent with the nuclear staining trends.

Some studies relating to c-Myc and survival show a tendency to disagree with the consensus molecular subtypes, and show worse survival with increased c-Myc expression [97]. Other factors in this analysis, such as c-Myc being highly expressed in left-sided tumours, and primary and lymph node stages, show an agreement with the CMS subtyping in MMR proficient patients, but not regarding survival [26]. It may be that the multiple functions c-Myc plays in colorectal cancer have contradictory clinical results that ultimately lead to poor outcomes. An example is the role c-Myc plays in cancer stem cell maintenance, which would lead to increased risk of recurrence and overall poor survival [98],

but is not necessarily associated with metastasis [99]. Another possibility is that c-Myc expression was interpreted differently in other studies which using different antibodies and methodologies. The association between MLH1 deficiency and c-Myc expression is supported by the literature [48] and could be driving proliferation and giving c-Myc positive cells a selective advantage.

Stromal c-Myc expression was observed in multiple patients and staining was localised to the cytoplasm of plasma cells. The strong association with conspicuous peritumoural lymphocytes supports the possibility that, as plasma cells tend to be associated more with normal colon stroma than tumour stroma, this shows the presence of immune response [100, 101]. Indeed, high stromal c-Myc was associated with lower stage, primary tumour stage and lymph node metastasis, and was indicative of improved survival. Increased c-Myc staining in the stroma may be informative of an active immune response and improved outcome.

This study of 1000 primary tumours from CRC patients has demonstrated an interesting association with clinical outcomes. Further work could be pursued in order to demonstrate precisely which pathways are responsible for the influence of c-Myc expression on survival. This could include immunohistochemical comparison with other markers such as β-catenin and other proteins involved in Wnt signalling, or markers of proliferation and metastasis. As increased c-Myc expression is associated with chemoresistance [102, 103], comparing c-Myc expression in matched pre and post treatment patient samples may help to devise better treatment strategies for patients.

CD44 tumoural expression

CD44 proteins play multiple roles in the normal and cancerous colon. Its function in the interaction between tumour epithelium and stroma make it an interesting protein and potential biomarker. The immunohistochemical staining of the TMAs was performed with the antibody clone DF1485, which is described as reacting to the whole protein. CD44 however, is not one single protein, but multiple variants which are predictive of different clinical outcomes [56, 104]. CD44 staining is expected in multiple cellular compartments, including the membrane, cytoplasm and nucleus. Very little nuclear staining was observed in this study and could represent a rare population of colorectal cancer patients. CD44 interacts with its main ligand hyaluronic acid on the cell surface, and this is the main cellular compartment where staining is expected [55]. The association between CD44 expression and survival was inconclusive due to the crossing of the Kaplan-Meier curves. High membrane-specific staining was associated with low stage, lack of invasiveness and tumours originating in the right side of the colon, especially in MMR proficient patients. CD44 positive tumours could represent a CMS2-like population, given the relationship between CD44 and the Wnt-signalling pathway. Bioinformatic analysis revealed an association between increased

CD44 expression and loss of expression at ch18q21, a recognised feature of the canonical adenoma-carcinoma pathway [6], enriched in CMS2 tumours [26].

Expression of CD44 in the cytoplasm was represented by granular and diffuse staining, and in future analyses it would be interesting to split the two groups to see if there were differences in clinical outcomes. As the granular staining was generally localised to the basolateral edge of cells, it could be that this represents an intracellular form of CD44, involved in detaching tumour epithelial cells from the extracellular matrix and leading to invasion and metastasis [63]. Studying this interaction could demonstrate the method by which CD44 regulates metastasis. High cytoplasmic CD44 was associated with low primary tumour stage, lack of metastasis and invasiveness, as was membranous staining. This is surprising as previous studies have suggested that CD44 is implicated in metastasis, due to its interaction with the extracellular matrix. Intracellular forms of CD44 are associated with detachment from the extracellular matrix and maintaining a stem cell state [63, 65]. Studies have shown that while increased CD44v6 is associated with worse survival and metastasis [105], increased CD44s is associated with better survival and lower primary tumour stage [106], as seen in this study. If the antibody used here detects multiple isoforms of CD44, it may be that the effect of one isoform is confounding the effect of the other. Use of antibodies specifically targeting the different isoforms could improve the reliability of the results.

Lymphocytes are known to express CD44 [107] and so it is unsurprising that CD44 staining was found in the tumour stroma. Increased stromal CD44 staining was associated with peritumoural lymphocytes and worse 5 year survival, although Cox regression analysis showed that this effect was confounded by the presence of peritumoural lymphocytes. Given that CD44 is expressed in these lymphocytes, it is possible that CD44 could be used in colorectal cancer as both a marker of cancer stem cells and as a marker of tumour infiltrating lymphocytes.

Both c-Myc and CD44 have been studied extensively for their roles in colorectal cancer. By looking at their joint expression in tumours, it has provided an opportunity to understand which clinical variables are influenced by both proteins, and their shared pathways. CD44 and c-Myc feature in initiating and maintaining a stem-like state in colorectal cancer cells [108, 109]. This study demonstrates that although this may be true, there appears to be no impact on tumour stage or overall survival when the two are expressed together. Furthermore, there is an association between the expression of both biomarkers with low primary tumour stage and metastasis, suggesting that this population of c-Myc+/CD44+ tumours may have a good outcome.

Tumour stroma content

Tumour stroma content has a prognostic role in colorectal cancer. Increased stromal content is associated with CMS4, a mesenchymal molecular subtype

associated with EMT and the worst overall survival of the consensus molecular subtypes [26]. A median cut-off of 46.6% was used to determine high and low stromal content, consistent with other studies using a 50% cut-off [110]. Clinical variables associated with metastasis, including vascular invasion, and increased primary tumour stage were associated with stromal content. High stromal content was associated with poor survival, consistent with previous analyses [111, 112]. c-Myc expression did not show any significant association with tumour stroma content, but CD44 did. There was a higher level of CD44 expression in the patients with low stroma, suggesting that the presence of CD44 is indicative of better outcomes, in agreement with the independent biomarker analysis. When the two were analysed in a multivariate Cox-regression analysis, it became clear that the tumour stroma content was confounding the effects of CD44 on survival. Improving the stromal analysis could include measuring the stroma directly, either using digital image analysis or using antibodies to assess the content of fibroblasts or lymphocytes.

Bioinformatic analyses

This project represented an opportunity to investigate the expression of CD44 and c-Myc in archived colorectal cancer samples. The samples used here do not have RNA data and access to other biomarker data, to supplement this work, was limited. The use of TCGA data therefore represented a chance to extend this work and understand which genes these proteins interact with in CRC. The upregulation of genes associated with proliferation and stem-state in the c-Myc analysis confirms what was seen in the literature regarding its functions in CRC. This analysis also offered an opportunity to look for previously unobserved connections, such as that between CD44 and CALM5 and CREB3L3, which both belong to the Melanogenesis pathway. This could have significant repercussions when selecting therapies for colorectal cancer, as the Melanogenesis pathway has been associated with resistance to oxaliplatin [113]. Time constraints meant that further analyses could not be performed to interrogate these data, but this is something that could be pursued as a future work to enable enhanced understanding of the roles of c-Myc and CD44 in colorectal cancer.

Conclusion

This study presented an opportunity to study three different biomarkers of colorectal cancer. By analysing the expression of CD44 and c-Myc in stroma as well as epithelial cells, additional information about patient outcomes could be inferred. Future analyses could include assessing the relationships of other biomarkers with CD44 and c-Myc, to find the combination with the most prognostic value. Looking at both c-Myc and CD44 expression in matched metastatic and lymph node samples could enable better understanding of how these proteins are involved in the metastatic process. The findings here regarding CD44 contrasted with the existing literature. This demonstrates the importance of using variant-specific antibodies to study CD44 expression. c-Myc positive tumours were found to have worse survival that c-Myc negative

tumours, knowledge which could be used to tailor treatment for patients of CRC. Finally, the analysis of tumour stroma content demonstrated the negative impact of stroma-rich tumours on clinical outcomes.

Acknowledgements

I would like to express my gratitude to those people without whose help I would not have been able to complete this research project. First, I would like to thank my supervisors Prof Ilyas and Dr Mukherjee for everything they have done throughout my Masters degree. Second, I would like to thank Zaki and Declan for their support in the short time I spent in the lab. I would also like to acknowledge Dr Venn for her brilliant teaching of statistics, and Dr Mongan for his bioinformatics support. Thanks to my friends and family for their support and encouragement. Finally I would like to thank Nick for his relentless cheerleading and IT support for the last 6 months.

References

1. CRUK. *Cancer Research UK*. 2020 [cited 2020 01 April]; Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer#heading-One.
2. Taieb, J., T. André, and E. Auclin, *Refining adjuvant therapy for non-metastatic colon cancer, new standards and perspectives.* Cancer Treatment Reviews, 2019. **75**: p. 1-11.
3. Pino, M.S. and D.C. Chung, *The Chromosomal Instability Pathway in Colon Cancer.* Gastroenterology, 2010. **138**(6): p. 2059-2072.
4. Fearon, E.R. and B. Vogelstein, *A genetic model for colorectal tumorigenesis.* Cell, 1990. **61**(5): p. 759-767.
5. Komiya, Y. and R. Habas, *Wnt signal transduction pathways.* Organogenesis, 2008. **4**(2): p. 68-75.
6. Armaghany, T., et al., *Genetic alterations in colorectal cancer.* Gastrointestinal cancer research : GCR, 2012. **5**(1): p. 19-27.
7. Lièvre, A., et al., *KRAS Mutations As an Independent Prognostic Factor in Patients With Advanced Colorectal Cancer Treated With Cetuximab.* Journal of Clinical Oncology, 2008. **26**(3): p. 374-379.
8. Miyaki, M., et al., *Higher frequency of Smad4 gene mutation in human colorectal cancer with distant metastasis.* Oncogene, 1999. **18**(20): p. 3098-3103.
9. Grady, W.M., *CIMP and colon cancer gets more complicated.* Gut, 2007. **56**(11): p. 1498-1500.
10. Tapial, S., et al., *Cimp-Positive Status is More Representative in Multiple Colorectal Cancers than in Unique Primary Colorectal Cancers.* Scientific Reports, 2019. **9**(1): p. 10516.
11. Ionov, Y., et al., *Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis.* Nature, 1993. **363**(6429): p. 558-561.
12. Sahnane, N., et al., *Aberrant DNA methylation profiles of inherited and sporadic colorectal cancer.* Clin Epigenetics, 2015. **7**: p. 131.

13. Fricke, F., et al., *TGFBR2-dependent alterations of microRNA profiles in extracellular vesicles and parental colorectal cancer cells.* Int J Oncol, 2019. **55**(4): p. 925-937.
14. Mucci, L.A., et al., *Familial Risk and Heritability of Cancer Among Twins in Nordic Countries.* Jama, 2016. **315**(1): p. 68-76.
15. Lynch, H.T., et al., *Milestones of Lynch syndrome: 1895–2015.* Nature Reviews Cancer, 2015. **15**(3): p. 181-194.
16. Sargent, D.J., et al., *Defective Mismatch Repair As a Predictive Marker for Lack of Efficacy of Fluorouracil-Based Adjuvant Therapy in Colon Cancer.* Journal of Clinical Oncology, 2010. **28**(20): p. 3219-3226.
17. Zaanan, A., et al., *Role of Deficient DNA Mismatch Repair Status in Patients With Stage III Colon Cancer Treated With FOLFOX Adjuvant Chemotherapy: A Pooled Analysis From 2 Randomized Clinical Trials.* JAMA Oncology, 2018. **4**(3): p. 379-383.
18. Clark, S.K., *Management of genetically determined colorectal cancer.* The Surgeon, 2019. **17**(3): p. 165-171.
19. Budinska, E., et al., *Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer.* The Journal of pathology, 2013. **231**(1): p. 63-76.
20. Schlicker, A., et al., *Subtypes of primary colorectal tumors correlate with response to targeted treatment in colorectal cell lines.* BMC Medical Genomics, 2012. **5**(1): p. 66.
21. Roepman, P., et al., *Colorectal cancer intrinsic subtypes predict chemotherapy benefit, deficient mismatch repair and epithelial-to-mesenchymal transition.* International Journal of Cancer, 2014. **134**(3): p. 552-562.
22. Sadanandam, A., et al., *A colorectal cancer classification system that associates cellular phenotype and responses to therapy.* Nature Medicine, 2013. **19**(5): p. 619-625.
23. Marisa, L., et al., *Gene Expression Classification of Colon Cancer into Molecular Subtypes: Characterization, Validation, and Prognostic Value.* PLOS Medicine, 2013. **10**(5): p. e1001453.
24. Perez Villamil, B., et al., *Colon cancer molecular subtypes identified by expression profiling and associatedto stroma, mucinous type and different clinical behavior.* BMC Cancer, 2012. **12**(1): p. 260.
25. Muzny, D.M., et al., *Comprehensive molecular characterization of human colon and rectal cancer.* Nature, 2012. **487**(7407): p. 330-337.
26. Guinney, J., et al., *The consensus molecular subtypes of colorectal cancer.* Nature Medicine, 2015. **21**(11): p. 1350-1356.
27. Lei, Z., et al., *Identification of Molecular Subtypes of Gastric Cancer With Different Responses to PI3-Kinase Inhibitors and 5-Fluorouracil.* Gastroenterology, 2013. **145**(3): p. 554-565.
28. Vu, T. and P.K. Datta, *Regulation of EMT in Colorectal Cancer: A Culprit in Metastasis.* Cancers (Basel), 2017. **9**(12).
29. Elbadawy, M., et al., *Emerging Roles of C-Myc in Cancer Stem Cell-Related Signaling and Resistance to Cancer Chemotherapy: A Potential Therapeutic Target Against Colorectal Cancer.* International Journal of Molecular Sciences, 2019. **20**(9): p. 2340.
30. Taub, R., et al., *Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells.* Proc Natl Acad Sci U S A, 1982. **79**(24): p. 7837-41.

31. Watt, R., et al., *The structure and nucleotide sequence of the 5' end of the human c-myc oncogene.* Proceedings of the National Academy of Sciences, 1983. **80**(20): p. 6307.
32. Grandori, C., et al., *The Myc/Max/Mad Network and the Transcriptional Control of Cell Behavior.* Annual Review of Cell and Developmental Biology, 2000. **16**(1): p. 653-699.
33. Allevato, M., et al., *Sequence-specific DNA binding by MYC/MAX to low-affinity non-E-box motifs.* PLOS ONE, 2017. **12**(7): p. e0180147.
34. Nair, S.K. and S.K. Burley, *X-Ray Structures of Myc-Max and Mad-Max Recognizing DNA: Molecular Bases of Regulation by Proto-Oncogenic Transcription Factors.* Cell, 2003. **112**(2): p. 193-205.
35. Murphy, D.J., et al., *Distinct Thresholds Govern Myc's Biological Output In Vivo.* Cancer Cell, 2008. **14**(6): p. 447-457.
36. Imperial, R., et al., *Comparative proteogenomic analysis of right-sided colon cancer, left-sided colon cancer and rectal cancer reveals distinct mutational profiles.* Molecular Cancer, 2018. **17**(1): p. 177.
37. Deng, M., et al., *Down-regulation of SLC35C1 induces colon cancer through over-activating Wnt pathway.* Journal of Cellular and Molecular Medicine, 2020. **24**(5): p. 3079-3090.
38. He, T.-C., et al., *Identification of c-MYC as a Target of the APC Pathway.* Science, 1998. **281**(5382): p. 1509.
39. Hao, Y.-H., et al., *Induction of LEF1 by MYC activates the WNT pathway and maintains cell proliferation.* Cell Communication and Signaling, 2019. **17**(1): p. 129.
40. Poli, V., et al., *MYC-driven epigenetic reprogramming favors the onset of tumorigenesis by inducing a stem cell-like state.* Nature Communications, 2018. **9**(1): p. 1024.
41. Cho, J.H., M. Dimri, and G.P. Dimri, *A positive feedback loop regulates the expression of polycomb group protein BMI1 via WNT signaling pathway.* J Biol Chem, 2013. **288**(5): p. 3406-18.
42. Hishida, T., et al., *Indefinite Self-Renewal of ESCs through Myc/Max Transcriptional Complex-Independent Mechanisms.* Cell Stem Cell, 2011. **9**(1): p. 37-49.
43. Takahashi, K. and S. Yamanaka, *Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors.* Cell, 2006. **126**(4): p. 663-676.
44. Nakagawa, M., et al., *Promotion of direct reprogramming by transformation-deficient Myc.* Proceedings of the National Academy of Sciences, 2010. **107**(32): p. 14152.
45. Satoh, K., et al., *Global metabolic reprogramming of colorectal cancer occurs at adenoma stage and is induced by MYC.* Proceedings of the National Academy of Sciences, 2017. **114**(37): p. E7697.
46. Tamada, M., M. Suematsu, and H. Saya, *Pyruvate Kinase M2: Multiple Faces for Conferring Benefits on Cancer Cells.* Clinical Cancer Research, 2012. **18**(20): p. 5554.
47. Yae, T., et al., *Alternative splicing of CD44 mRNA by ESRP1 enhances lung colonization of metastatic cancer cell.* Nat Commun, 2012. **3**: p. 883.
48. Partlin, M.M., et al., *Interactions of the DNA mismatch repair proteins MLH1 and MSH2 with c-MYC and MAX.* Oncogene, 2003. **22**(6): p. 819-825.

49. Bindra, R.S. and P.M. Glazer, *Co-repression of mismatch repair gene expression by hypoxia in cancer cells: Role of the Myc/Max network.* Cancer Letters, 2007. **252**(1): p. 93-103.
50. Zhai, D., et al., *Sterol regulatory element-binding protein 1 cooperates with c-Myc to promote epithelial-mesenchymal transition in colorectal cancer.* Oncol Lett, 2018. **15**(4): p. 5959-5965.
51. Li, H., et al., *The integrated pathway of TGFβ/Snail with TNFα/NFκB may facilitate the tumor-stroma interaction in the EMT process and colorectal cancer prognosis.* Scientific Reports, 2017. **7**(1): p. 4915.
52. Allen-Petersen, B.L. and R.C. Sears, *Mission Possible: Advances in MYC Therapeutic Targeting in Cancer.* BioDrugs, 2019. **33**(5): p. 539-553.
53. Beaulieu, M.-E., et al., *Intrinsic cell-penetrating activity propels Omomyc from proof of concept to viable anti-MYC therapy.* Science Translational Medicine, 2019. **11**(484): p. eaar5012.
54. Tögel, L., et al., *Dual Targeting of Bromodomain and Extraterminal Domain Proteins, and WNT or MAPK Signaling, Inhibits c-MYC Expression and Proliferation of Colorectal Cancer Cells.* Molecular Cancer Therapeutics, 2016. **15**(6): p. 1217.
55. Ponta, H., L. Sherman, and P.A. Herrlich, *CD44: From adhesion molecules to signalling regulators.* Nature Reviews Molecular Cell Biology, 2003. **4**(1): p. 33-45.
56. Prochazka, L., R. Tesarik, and J. Turanek, *Regulation of alternative splicing of CD44 in cancer.* Cellular Signalling, 2014. **26**(10): p. 2234-2239.
57. Wang, Z., et al., *CD44/CD44v6 a Reliable Companion in Cancer-Initiating Cell Maintenance and Tumor Progression.* Frontiers in cell and developmental biology, 2018. **6**: p. 97-97.
58. Lokeshwar, V.B. and L.Y. Bourguignon, *Post-translational protein modification and expression of ankyrin-binding site(s) in GP85 (Pgp-1/CD44) and its biosynthetic precursors during T-lymphoma membrane biosynthesis.* Journal of Biological Chemistry, 1991. **266**(27): p. 17983-17989.
59. Gao, T., et al., *Disruption of Core 1-mediated O-glycosylation oppositely regulates CD44 expression in human colon cancer cells and tumor-derived exosomes.* Biochemical and Biophysical Research Communications, 2020. **521**(2): p. 514-520.
60. Dzwonek, J. and G.M. Wilczynski, *CD44: molecular interactions, signaling and functions in the nervous system.* Frontiers in cellular neuroscience, 2015. **9**: p. 175-175.
61. Borland, G., J.A. Ross, and K. Guy, *Forms and functions of CD44.* Immunology, 1998. **93**(2): p. 139-148.
62. Hanley, W.D., et al., *CD44 on LS174T colon carcinoma cells possesses E-selectin ligand activity.* Cancer Res, 2005. **65**(13): p. 5812-7.
63. Okamoto, I., et al., *CD44 cleavage induced by a membrane-associated metalloprotease plays a critical role in tumor cell migration.* Oncogene, 1999. **18**(7): p. 1435-1446.
64. Okamoto, I., et al., *Proteolytic cleavage of the CD44 adhesion molecule in multiple human tumors.* The American journal of pathology, 2002. **160**(2): p. 441-447.
65. Nagano, O. and H. Saya, *Mechanism and biological significance of CD44 cleavage.* Cancer Science, 2004. **95**(12): p. 930-935.

66. Wang, Y., et al., *Cytoskeletal Regulation of CD44 Membrane Organization and Interactions with E-selectin.* Journal of Biological Chemistry, 2014. **289**(51): p. 35159-35171.
67. Ma, L., L. Dong, and P. Chang, *CD44v6 engages in colorectal cancer progression.* Cell Death & Disease, 2019. **10**(1): p. 30.
68. Zeilstra, J., et al., *Stem cell CD44v isoforms promote intestinal cancer formation in Apc(min) mice downstream of Wnt signaling.* Oncogene, 2014. **33**(5): p. 665-670.
69. Brabletz, T., Jung, A., Spaderna, S. et al., *Migrating cancer stem cells — an integrated concept of malignant tumour progression.* Nat Rev Cancer, 2005. **5**: p. 744-749.
70. Afify, A., et al., *The expression of CD44v6 in colon: from normal to malignant.* Annals of Diagnostic Pathology, 2016. **20**: p. 19-23.
71. Leng, Z., et al., *Lgr5+CD44+EpCAM+ Strictly Defines Cancer Stem Cells in Human Colorectal Cancer.* Cellular Physiology and Biochemistry, 2018. **46**(2): p. 860-872.
72. Mashita, N., et al., *Epithelial to mesenchymal transition might be induced via CD44 isoform switching in colorectal cancer.* Journal of Surgical Oncology, 2014. **110**(6): p. 745-751.
73. Todaro, M., et al., *CD44v6 Is a Marker of Constitutive and Reprogrammed Cancer Stem Cells Driving Colon Cancer Metastasis.* Cell Stem Cell, 2014. **14**(3): p. 342-356.
74. Weber, G.F., *The metastasis gene osteopontin: a candidate target for cancer therapy.* Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, 2001. **1552**(2): p. 61-85.
75. Wang, J.L., et al., *CD44v6 overexpression related to metastasis and poor prognosis of colorectal cancer: A meta-analysis.* Oncotarget, 2017. **8**(8): p. 12866-12876.
76. Mulder, J.W., et al., *Expression of mutant p53 protein and CD44 variant proteins in colorectal tumorigenesis.* Gut, 1995. **36**(1): p. 76-80.
77. Misra, S., S. Ghatak, and B.P. Toole, *Regulation of MDR1 expression and drug resistance by a positive feedback loop involving hyaluronan, phosphoinositide 3-kinase, and ErbB2.* J Biol Chem, 2005. **280**(21): p. 20310-5.
78. Ilyas, M., et al., *Guidelines and considerations for conducting experiments using tissue microarrays.* Histopathology, 2013. **62**(6): p. 827-39.
79. Molinari, C., et al., *Heterogeneity in Colorectal Cancer: A Challenge for Personalized Medicine?* International Journal of Molecular Sciences, 2018. **19**(12).
80. Bankhead, P., et al., *QuPath: Open source software for digital pathology image analysis.* Scientific Reports, 2017. **7**(1): p. 16878.
81. Liao, Y., et al., *WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs.* Nucleic Acids Research, 2019. **47**(W1): p. W199-W205.
82. Support, M.C. *Staging and grading of bowel cancer*. 07/09/2020]; Available from: https://www.macmillan.org.uk/cancer-information-and-support/bowel-cancer/staging-and-grading-of-bowel-cancer.
83. Koelzer, V.H. and A. Lugli, *The tumor border configuration of colorectal cancer as a histomorphological prognostic indicator.* Frontiers in oncology, 2014. **4**: p. 29-29.
84. Overman, M.J., M.S. Ernstoff, and M.A. Morse, *Where We Stand With Immunotherapy in Colorectal Cancer: Deficient Mismatch Repair, Proficient*

Mismatch Repair, and Toxicity Management. American Society of Clinical Oncology Educational Book, 2018(38): p. 239-247.
85. Tarone, R.E. and J. Ware, *On distribution-free tests for equality of survival distributions.* Biometrika, 1977. **64**(1): p. 156-160.
86. Weber, L., et al., *Activation of odorant receptor in colorectal cancer cells leads to inhibition of cell proliferation and apoptosis.* PloS one, 2017. **12**(3): p. e0172491-e0172491.
87. Ranzani, M., et al., *Revisiting olfactory receptors as putative drivers of cancer.* Wellcome open research, 2017. **2**: p. 9-9.
88. Freyer, L. and B.E. Morrow, *Canonical Wnt signaling modulates Tbx1, Eya1, and Six1 expression, restricting neurogenesis in the otic vesicle.* Dev Dyn, 2010. **239**(6): p. 1708-22.
89. Li, B., et al., *MiR-629-3p-induced downregulation of SFTPC promotes cell proliferation and predicts poor survival in lung adenocarcinoma.* Artif Cells Nanomed Biotechnol, 2019. **47**(1): p. 3286-3296.
90. Kanehisa, M., et al., *New approach for understanding genome variations in KEGG.* Nucleic Acids Res, 2019. **47**(D1): p. D590-d595.
91. Wang, C., et al., *Cholesterol Enhances Colorectal Cancer Progression via ROS Elevation and MAPK Signaling Pathway Activation.* Cellular Physiology and Biochemistry, 2017. **42**(2): p. 729-742.
92. Fu, T., et al., *FXR Regulates Intestinal Cancer Stem Cell Proliferation.* Cell, 2019. **176**(5): p. 1098-1112.e18.
93. Zhou, Y., et al., *Cancer stem cells in progression of colorectal cancer.* Oncotarget, 2017. **9**(70).
94. Wang, Y., et al., *Overexpression of Wnt7α protein predicts poor survival in patients with colorectal carcinoma.* Tumor Biology, 2015. **36**(11): p. 8781-8787.
95. Kit, O.I., et al., *A Proteomics Analysis Reveals 9 Up-Regulated Proteins Associated with Altered Cell Signaling in Colon Cancer Patients.* The Protein Journal, 2017. **36**(6): p. 513-522.
96. The UniProt, C., *UniProt: a worldwide hub of protein knowledge.* Nucleic Acids Research, 2019. **47**(D1): p. D506-D515.
97. Dong, S., et al., *Identification of Prognostic Biomarkers and Drugs Targeting Them in Colon Adenocarcinoma: A Bioinformatic Analysis.* Integrative Cancer Therapies, 2019. **18**: p. 1534735419864434.
98. Gupta, R., et al., *Colon cancer stem cells: Potential target for the treatment of colorectal cancer.* Cancer Biology & Therapy, 2019. **20**(8): p. 1068-1082.
99. Zalata, K.R., et al., *Genetic dissimilarity between primary colorectal carcinomas and their lymph node metastases: ploidy, p53, bcl-2, and c-myc expression—a pilot study.* Tumor Biology, 2015. **36**(8): p. 6579-6584.
100. Weisz-Carrington, P., M.E. Poger, and M.E. Lamm, *Secretory immunoglobulins in colonic neoplasms.* Am J Pathol, 1976. **85**(2): p. 303-16.
101. Banner, B.F., et al., *Characterization of the inflammatory cell populations in normal colon and colonic carcinomas.* Virchows Archiv B, 1993. **64**(1): p. 213-220.
102. Reyes-González, J.M., et al., *Targeting c-MYC in Platinum-Resistant Ovarian Cancer.* Molecular Cancer Therapeutics, 2015. **14**(10): p. 2260.
103. Walker, T.L., et al., *Tumour cells surviving in vivo cisplatin chemotherapy display elevated c-myc expression.* British Journal of Cancer, 1996. **73**(5): p. 610-614.

104. Li, X.-D., et al., *Clinical Significance of CD44 Variants Expression in Colorectal Cancer.* Tumori Journal, 2013. **99**(1): p. 88-92.
105. Saito, S., et al., *CD44v6 expression is related to mesenchymal phenotype and poor prognosis in patients with colorectal cancer.* Oncol Rep, 2013. **29**(4): p. 1570-1578.
106. Lugli, A., et al., *Prognostic impact of the expression of putative cancer stem cell markers CD133, CD166, CD44s, EpCAM, and ALDH1 in colorectal cancer.* British Journal of Cancer, 2010. **103**(3): p. 382-390.
107. Weimann, T.K., et al., *CD44 variant isoform v10 is expressed on tumor-infiltrating lymphocytes and mediates hyaluronan-independent heterotypic cell–cell adhesion to melanoma cells.* Experimental Dermatology, 2003. **12**(2): p. 204-212.
108. Morath, I., T.N. Hartmann, and V. Orian-Rousseau, *CD44: More than a mere stem cell marker.* The International Journal of Biochemistry & Cell Biology, 2016. **81**: p. 166-173.
109. Yoshida, G.J., *Emerging roles of Myc in stem cell biology and novel tumor therapies.* Journal of Experimental & Clinical Cancer Research, 2018. **37**(1): p. 173.
110. van Pelt, G.W., et al., *The tumour–stroma ratio in colon cancer: the biological role and its prognostic impact.* Histopathology, 2018. **73**(2): p. 197-206.
111. Huijbers, A., et al., *The proportion of tumor-stroma as a strong prognosticator for stage II and III colon cancer patients: validation in the VICTOR trial.* Annals of Oncology, 2013. **24**(1): p. 179-185.
112. Park, J.H., et al., *The relationship between tumour stroma percentage, the tumour microenvironment and survival in patients with primary operable colorectal cancer.* Annals of Oncology, 2014. **25**(3): p. 644-651.
113. Klahan, S., et al., *Bioinformatic analyses revealed underlying biological functions correlated with oxaliplatin responsiveness.* Tumor Biology, 2016. **37**(1): p. 583-590.