
This is the author’s final accepted version.

There may be differences between this version and the published version. You are advised to consult the publisher’s version if you wish to cite from it.

[http://eprints.gla.ac.uk/140271/](http://eprints.gla.ac.uk/140271/)

Deposited on: 25 April 2017
Colorectal cancer subtypes: Translation to routine clinical pathology

Antonia K Roseweir¹,², Donald C McMillan¹, Paul G Horgan¹, Joanne Edwards²

¹ Academic Unit of Surgery, School of Medicine, University of Glasgow, Royal Infirmary, Glasgow, United Kingdom

² Unit of Experimental Therapeutics, Institute of Cancer Sciences, University of Glasgow, Gascube Estate, Glasgow, United Kingdom

Corresponding Author: Dr Antonia Roseweir, Institute of Cancer Sciences, University of Glasgow, Wolfson Wohl Cancer Research Centre, Glasgow, United Kingdom, antonia.roseweir@glasgow.ac.uk, 0141 330 8607
Abstract

Colorectal cancer (CRC) is the second most common cause of cancer death in Europe. Although outcomes have improved, it is clear that from a genomic standpoint CRC is not one disease, but a heterogeneous group of malignancies that arise within one organ. Given that different subtypes have different outcomes, the ability to subtype tumours in the clinic would be highly favourable, enabling optimal treatment for individual patients. In 2015, a consortium proposed four consensus subtypes for CRC (MSI immune, canonical, metabolic, and mesenchymal) based on six classifications systems reported to have prognostic value. However, genomic assessment of tumours is not readily translated into routine pathology with a need for standardisation and reproducibility of assessment. Immunohistochemistry is widely used in routine pathology, and would present a more readily translatable method for subtyping CRC tumours. Therefore, the literature was reviewed to characterise the genomic and phenotypic features associated with each subtype, with the aim of enabling subtyping of CRC to be taken forward into routine clinical practice.

Keywords: Colorectal Cancer, molecular subtypes, genomic, phenotypic, translational
Introduction

Colorectal cancer (CRC) is the second most common cause of cancer death in Europe [1]. Although outcomes have improved over the past decades, predominantly as a result of improvements in surgical technique and adjuvant/neo-adjuvant therapies, survival remains poor, with an overall 5-year survival of approximately 50% [2]. Therefore, there remains a need to identify characteristics pertaining to both the tumour and microenvironment which may not only guide prognosis, but also provide novel targets for adjuvant therapies. It is now recognised that from a genomic standpoint, as with many cancer types, CRC is not just one disease, but a heterogeneous group of malignancies that arise within the same organ. Therefore, given their prognostic value the ability to subtype patient tumours in the routine pathology setting would enable optimal treatment allocation.

The concept of subtyping tumours based on genomic alterations is not new. In breast cancer, subtypes have been recognised since the early 2000s, mainly using large gene expression profiling on arrays of over 400 genes [3-5]. Since then breast cancer subtypes have been refined to the four classifications used in the clinic today (Luminal A, luminal B, HER2- type and triple negative). These subtypes can now be routinely tested on patient tumours using only four immunohistochemical (IHC) markers (ER, PR, HER2, and Ki67) [6, 7]. These subtypes are employed to determine the appropriate treatment for patients. For example, luminal tumours receive Tamoxifen and/or aromatase inhibitors and HER2+ tumours receive Pertuzimab and chemotherapy.

In CRC, heterogeneity of tumours has also been recognised with microsatellite instability (MSI) and CpG island methylation (CIMP) well documented. MSI is genetic hyper-mutability that results from impaired DNA mismatch repair (MMR). MSI usually leads to a
better prognosis for patients and is often paired with high CIMP (CIMP-H) [8]. More recently, mutational difference in key pathways have also begun to emerge, with BRAF and KRAS mutational status being linked to poorer patient outcome [9, 10]. However, it has been reported that in MSI tumours the detrimental effects of BRAF mutations are mitigated [11]. Effects of the immune system, stromal invasion and tumour proliferation rates are now also being recognised as important features for tumour progression [9, 12, 13]. These genomic and phenotypic characteristics may be employed to subtype CRC tumours, improving personalised treatment for patients.

Guinney and co-workers [14] recently proposed, based on the consolidation of six different large genomic subtyping studies shown to have prognostic value, four subtypes for CRC (MSI Immune, Canonical, Metabolic and Mesenchymal). However, these four subtypes are dependent on specialised genetic testing and therefore not readily translated into the routine clinical pathology laboratory. This present work therefore, reviews the literature on these four subtypes to examine whether there is a simple subtyping protocol for patients with CRC.
Methods

A review of the published literature was undertaken. The outcome of interest was subtype components for CRC. The review was carried out using the US national library of medicine (PubMed) database with the following search terms: colorectal cancer and subtyping. To be eligible for inclusion, studies had to meet the following criteria: (a) pertain to colorectal cancer, (b) have defined genomic or phenotypic characteristics associated with the four consensus subtypes. This resulted in 43 articles to be included in the study.
The Consensus Subtypes of Colorectal Cancer

In 2015, a large scale consortium of the leading scientists within the colorectal field, reported a consensus of four subtypes for CRC (CMS1: MSI Immune, CMS2: Canonical, CMS3: Metabolic and CMS4: Mesenchymal) [14]. The consortium undertook large scale data sharing and analysis of the current research within the field and found interconnectivity between six proposed classification systems with prognostic value (Table 1) [15-20]. These classifications were based on mRNA profiling data of 18 CRC databases from public and proprietary sources including The Cancer Genome Atlas (TCGA). Each classification system employed different expression platforms, sample types and study designs, and therefore normalisation of the data was undertaken and the consortium applied each classification system to all 18 datasets, to test their validity. Network based transcriptomic analysis was then performed to study associations between the different classification systems and the four robust CMS groups were identified.

These subtypes were then further analysed to delineate their individual molecular characteristic including genomic features such as CIMP, somatic copy number analysis (SCNA), mutational profile, specifically KRAS and BRAF mutational status, as well as proteomic analysis of oncogenic signalling cascades and phenotypic characteristics such as stromal invasion and immune infiltrate (Table 2). Finally, they assessed the clinical implications of each CMS and CMS4 was shown to have the worst relapse free survival (5yr survival – 60%), with the other three subtypes showing better survival rates (5yr survival CMS1–75%, CMS2/3-73%; CMS1vCMS4- HR 1.74 95% CI 1.29-2.33).

However, transcriptomic, genomic and proteomic approaches do not currently meet the standardisation and reproducibility requirements needed to translate these subtypes into
routine pathological testing. However, the phenotypic characteristics associated with these subtypes may be more readily incorporated into routine pathology. Therefore, analysis of each molecular component is required to assess its relevance to patient outcome and therefore its role in the subtyping of this disease. This may enable the creation of a simplified subtyping method that may be translated to routine pathology.

**CMS1: MSI Immune**

*Genomic Alterations - Microsatellite Instability (MSI), CpG Island Methylation Phenotype (CIMP) and BRAF mutational status*

Overall microsatellite instability (MSI) occurs in 15-20% of CRC tumours and occurs due to mutations in mismatch repair (MMR) genes (MLH1, MSH2, MSH6), resulting in failure to repair errors in the microsatellite DNA repeats [21, 22]. These MMR mutations can arise in two ways; firstly, through inherited mutations giving rise to Lynch syndrome and secondly via hyper-methylation of the MLH1 promoter (sporadic). The majority of MSI tumours are sporadic, which have a worse prognosis than for patients with Lynch syndrome, therefore there is a need distinguish sporadic patients from those with Lynch syndrome [8]. One way to distinguish between sporadic MSI and Lynch syndrome is to assess the mutational burden for BRAF (V600E) in the tumours as Lynch syndrome tumours do not carry BRAF mutations, but most sporadic MSI tumours do [8]. Therefore, combined assessment of MSI with BRAF mutational analysis, can be used to distinguish sporadic MSI patients.

Although survival is worse than for Lynch syndrome, sporadic CRC patients with MSI-H tumours have still been shown to have a better prognosis that those with microsatellite stable (MSS) tumours (HR 0.73 95% CI 0.59-0.91), with relative 5yr survival rates of 72.4% and
63% respectively [23]. This better prognosis is also seen in the presence of BRAF mutations (HR 0.44 95% CI 0.26-0.75), with MSI/BRAF mutant patients having a 5yr survival rate of 65% compared to 46% for MSS/BRAF mutant patients. This suggests that MSI can mitigate the oncogenic effects of the BRAF pathway resulting in a better outcome for patients. Conversely, in MSS patients, BRAF mutations convey a poor prognosis and their prognosis declines further in BRAF-mutant metastatic disease [24].

There are two common methods of detection for MSI; the first is to assess the length of five specific microsatellite amplicons by PCR (BAT25, BAT26, D2S123, D5S346 and D17S250). Using this method, patients are classified as MSI-H if two or more of these amplicons are mutated and MSI-L if one is mutated. The main limitation with this method is the associated variability and time consuming nature [21]. The second method is based on the MMR proteins. It assesses loss of expression of the three MMR proteins via immunohistochemistry (IHC) in combination with PMS2. PMS2 forms a heterodimer with MLH1, so loss of MLH1 will also cause loss of PMS2, increasing the sensitivity of the assay [22]. For this method patients are classified as MSI-H if expression of any of these proteins is lost, this is a more reliable and less time consuming method. When the two methods are compared for sensitivity, MSI by PCR has a sensitivity of 90.2% and MMR by IHC has a sensitivity of 88.8%, suggesting that IHC is an equivalent approach for the detection of MSI and is readily translated into routine pathology [21].

Another feature of CMS1 tumours that also provides a better patient prognosis is CpG island methylation phenotype (CIMP). CIMP-H patients have a 5yr survival rate of 86% compared to 80% for CIMP-L patients (HR 0.88 95% CI 0.57-1.38), however it is 74% and 45%, respectively when BRAF mutations are present (HR 0.45 95% CI 0.26-0.79) [10]. CIMP is the hyper-methylation of DNA promoter-associated CpG islands, mainly within tumour
suppressor and DNA repair genes, and leads to transcriptional silencing of these genes. It is thought that CIMP is the main method by which MLH1 is hyper-methylated leading to sporadic MSI [8]. Therefore, most CIMP-H tumours are also MSI-H and contain BRAF (V600E) mutations [25]. This has been shown in patient studies, where CIMP-H was shown to associate with older female patients, smoking, proximal location, MSI-H tumours and BRAF mutations [10]. This study also noted a proximal location of CIMP-H tumours, and this has been reported in other studies where it was suggested that CIMP-H and MSI-H increase from the rectum to the caecum [26]. This is interesting as it is thought there is a similar distribution of local inflammation across CRC tumours, and CIMP-H/MSI-H/BRAF mutant tumours are also associated with an increase in intra-tumour inflammation characterised by a lymphocytic infiltration [8].

Given the above relationship between MSI, CIMP and BRAF it raises the question of whether subtyping of CRC tumours can be carried out more simply. In particular, MSI appears to be the most important genomic component of CMS1. Therefore, if the main role of CIMP in CRC is to hyper-methylate MLH1 leading to MSI, should MSI be simply assessed using an IHC approach? One potential obstacle to using MSI alone is the observation of MSI tumours within other CMS groups; therefore assessment of MSI may not sufficient to separate CMS1 tumours from the other subtypes and another component would be needed to fully differentiate CMS1 patients.

**Phenotypic Characteristics – Immune Infiltration**

There is good evidence that the association of better prognosis and MSI tumours is due to the high local inflammatory infiltration [27]. De Smedt et al showed that MSI tumours have increased cytotoxic T-lymphocytes in the tumour and peritumoural areas, where they also
reported increased macrophages compared to MSS tumours [28]. Furthermore studies have shown that MSI patients with high cytotoxic T-lymphocytes have better overall survival and they proposed that this is due to increased intratumoural inflammation [27, 29].

Local inflammatory responses have been widely shown to play an active role in tumour development across a wide range of cancers including CRC [30-32]. Local inflammation is characterised by an increase in tumour infiltrating lymphocytes (TILs) within the tumour microenvironment. TILs are comprised of two main types, CD4+ helper T-lymphocytes and cytotoxic CD8+ T-lymphocytes. CD4+ helper T-lymphocytes are essential to promote the proliferation of the cytotoxic CD8+ T-lymphocytes that can then kill the tumour cells. TILs at both the invasive margin of the tumour and within the cancer cell nests slow tumour growth and invasion and are associated with a better patient prognosis [27]. 5yr survival rates for cytotoxic CD8+ T-lymphocytes at both locations are 37.3% when both low, 61.1% when one high and 73.8% when both high, therefore TILS would appear to be a stronger prognostic factor than MSI for CMS1 [33, 34].

Indeed immunotherapy, directed to stimulate TILs, has been proposed to control CRC progression and is the subject of intense clinical research [35, 36]. The tumour inflammatory cell infiltrate have also been proposed to improve the staging of tumours, for example, the Galon immunoscore [37] and Klintrup Makinen (KM) grade [12]. Both of these have prognostic value independent of current TNM staging used for CRC [38] and may therefore also be useful in subtyping CMS1 tumours.

Galon’s immunoscore is based on IHC detection of the two main TILs, CD4+ and CD8+ T-lymphocytes, at the invasive margin and within cancer cell nests, giving a five point scoring system [37]. The KM grade, assesses the general local inflammatory infiltrate at the invasive margin using an H&E section, and marks it on a four point score [12]. When Park and co-
workers compared the two scores in 246 patients with CRC, the immunoscore stratified patients survival from 93% to 61% (HR 0.66 95% CI 0.56-0.88) and the KM grade from 88% to 66% (HR 0.44 95% CI 0.25-0.76), similar ranges to those seen for TILs [39]. Since the prognostic value of these scores was similar either approach provides a simple, effective method to subtype CMS1 tumours within routine pathology. Recently, Becht et al showed that both CMS1 and CMS4 associated with immune signatures. CMS1 was associated with cytotoxic lymphocyte genes recognised to be important in intratumoural inflammation and CMS4 was associated with monocyte genes recognised to be associated with the systemic inflammatory response [40]. Therefore, as the immunoscore assesses the ratio of cytotoxic T-lymphocytes within the tumour and KM grade assesses intratumoural inflammation, they should only detect CMS1 patients and therefore be suitable markers for this subtype.

**CMS2: Canonical**

*Genomic Alterations - Somatic copy number analysis (SCNA), WNT and MAPK*

Somatic copy number analysis (SCNA) are changes that cause a loss or gain of DNA sections within the genome during meiosis [41]. SCNAs are detected via whole genome sequencing, which is a variable and time consuming method. In CRC, the main gains are found in chromosome regions 20q, 13q, 8q and 7, with losses mainly found in 4, 8p, 18q and 17p. These specific SCNAs then drive cancer progression by affecting the WNT and MAPK pathways [42, 43]. The dysregulation of these pathways are associated with differences in patient survival. Therefore, given that the analysis of SCNAs is variable and time consuming, it may be that analysis of the affected pathways at the protein level would give better prognostic information.
WNT signalling is a critical pathway in the initiation of CRC, due to driver mutations in APC, which lead to the development of non-invasive polyps, the precursors to CRC. Following this event, upregulation of nuclear β-catenin, is crucial for progression to CRC. This upregulation can occur via mutations in β-catenin itself or due to mutations in KRAS that promote nuclear localisation of β-catenin and convey a worse prognosis to patients [44, 45]. In a study of 75 CRC patients, 5yr survival for patients with high nuclear β-catenin expression was 0% compared to 73% for patients with low expression [46]. However, in MSI tumours Wnt2 and Wnt5 can inhibit the nuclear localisation of β-catenin and this may be another reason that MSI tumours have a better prognosis [44].

One factor regulating β-catenin nuclear localisation is KRAS, which is also known to be a driver mutation for CRC. KRAS is a member of the mitogen activated protein kinase (MAPK) pathway, which is associated poor prognosis in many solid tumours including CRC. In a study of 1989 CRC patients, 5yr survival for KRAS WT patients was 78% compared to 73% for patients with KRAS mutations (HR 1.37 95% CI 1.13-1.66) [47]. When KRAS is mutated at G12/13 to valine in early stage disease, this causes constitutive activation of the MAPK cascade, which can regulate β-catenin, via the p90rsk pathway [47, 48]. Therefore, as one of the roles of MAPK in CRC is to regulate β-catenin and as β-catenin is associated with poorer prognosis, β-catenin may be used as the main genomic marker for this subtype, and it can be routinely analysed using IHC.

**Phenotypic Characteristics – Proliferation**

One of the phenotypic characteristics regulated by β-catenin is proliferation rate and high proliferation rate was a common phenotype of some of the subtypes used to compile CMS2 [17, 18]. Roepman et al showed B-type patents, which align with CMS2 had an increased proliferation rate measured by Ki67 [17]. Similarly, Budinska et al showed that their lower
crypt-like patients, where the only MSS patients to have an upregulation of proliferation genes, and again these patients align with CMS2 [18]. Proliferation rate is known to be a critical component of cancer progression, and has been associated with a decrease in CRC patient survival [9, 49]. In a study of 41 CRC patients with liver metastasis, 3yr survival rate for low proliferation was 68% compared to 16% for high proliferation (HR 3.04 95% CI 1.08-8.58), similar to β-catenin [49]. Proliferation can be measured in a variety of ways, from flow cytometry, to measurements of DNA synthesis or nuclear Ki67 staining. Ki67 is involved in proliferation at all stages of the cell cycle but it is absent from resting cells, making it a good marker of cellular proliferation. The advantages of Ki67 are that it can be detected in tissue by IHC, its scoring can be automated, and it is currently measured in the routine pathology of breast cancer. However, measurement of Ki67 does have limitations, it is known to change on fixation and the thresholds for low and high expression remain to be established. In spite of these limitations, it is still a widely used marker of proliferation. Hence, as β-catenin is widely known to increase cell proliferation, a high Ki67 proliferation rate may be used as a surrogate marker for β-catenin activation. Therefore, Ki67 proliferation rate may be a useful characteristic to determine CMS2 in patients and is already established within routine clinical pathology.

**CMS3: Metabolic**

*Genomic Alterations – RAS mutations*

In CRC, KRAS is mutated at G12/13 to valine in early stage disease; this is thought to be a driver mutation causing constitutive activation of the MAPK cascade leading to a poorer prognosis [47]. In 1989 CRC patients, 5yr survival for KRAS WT patients was 78%
compared to 73% for patients with KRAS mutations (HR 1.37 95% CI 1.13-1.66) [47]. In more advanced disease, KRAS mutations are associated with the development of lung metastasis, possibly due to direct effects of MAPK on invasion [48]. As well as KRAS, mutations in NRAS have now been established as prevalent in ~5% of KRAS WT CRC patients and also convey a poor prognosis to patients [50, 51]. Both KRAS and NRAS are established predictive biomarkers for response to EGFR inhibitors, such as cetuximab [52].

RAS (KRAS or NRAS) mutational analysis is performed using RT-PCR and sequencing. This genomic test is already employed within routine clinical pathology to determine patients suitable for cetuximab therapy; therefore it is readily translated to CRC subtyping. However, although the majority of CMS3 tumours have KRAS mutations (~75%), they are not exclusive to this subtype. Furthermore as CMS3 is relatively small, most KRAS mutant cancers will align with one of the other three CMS. Therefore a better marker is needed to distinguish CMS3 patients.

**Phenotypic Characteristics – Metabolism**

Recently, KRAS mutations have been associated with metabolism in cancer cells [53]. In particular, it is well established that cancer cells undergo metabolic reprogramming to a glycolytic state to survive and maintain proliferation in a hostile environment and there is increasing evidence this warburg effect is promoted by oncogenes, including KRAS [53]. In CRC, leptin production has been associated with adverse outcome in male patients, associating with advanced disease, nodal status and venous invasion [54]. In another study, a metabolic gene signature able to predict patient’s risk of relapse was shown to be associated with lower disease-free survival (HR 4.2 95% CI 1.79-9.88) [55]. The 3 year survival rate for patients with a low-risk metabolic gene signature was 84% compared to 46% for those with high-risk metabolic gene signature. However, such deregulation of metabolism is difficult to
measure within routine clinical pathology. Another phenotypic feature of these tumours is that they are the only subtype to have low proliferation, low stromal invasion and a low immune infiltration potentially as a result of this metabolic deregulation. Therefore, the combination of low proliferation and low stromal invasion and a low immune infiltration may be a better classification system for these tumours and may be readily translated into routine clinical pathology.

**CMS4: Mesenchymal**

*Genetic Alterations – Somatic Copy Number Alterations (SCNA)*

Like CMS2, this subtype also has a high number of SCNAs, and as mentioned above as the prognostic significance of SCNAs is due to the dysregulation of the affected signalling pathways, which for CMS4 appears to be TGFβ [42, 43]. Therefore, as the analysis of SCNAs is variable and time consuming, it may be that analysis of the TGFβ pathway at the protein level would give better prognostic information.

*Phenotypic Characteristics – Stromal Invasion*

It has been shown that the main source of TGFβ within CRC is the stromal cells; therefore an increase in stromal invasion leads to increased TGFβ production and in turn increased metastasis [56]. This increase in stromal invasion has been shown to be an important prognostic characteristic of CRC tumour progression. Parks and coworkers observed that an increase in the tumour stroma as measured by the tumour-stroma percentage (TSP) was a poor prognostic factor for CRC patients (HR 2.46 95% CI 1.56-3.89). TSP stratified 5 year
patient survival from 80% (low) to 57% (high) [39]. The TSP measures the percentage of the tumour area infiltrated by stroma, with 50% being a commonly used cut off between low and high. Other studies have also shown TSP to be an independent prognostic factor for CRC, having association with poorer cancer specific and overall survival, as well as increased stage and venous invasion [13, 57]. In a study of 710 CRC patients, 5-year survival again decreased from 77% in low stromal tumours to 59% in high stromal tumours (HR 2.15 95% CI 1.61-2.86) [57]. It may be hypothesised that an increase in stroma, inhibits the immune infiltration into the tumour. This may be why TSP appears to stratify survival in patients with a low KM grade [36]. These observations are consistent with a CMS4 subtype, where immune infiltrate is low but stromal invasion high. Therefore, high TSP may be a good marker to determine CMS4 subtype in the routine clinical pathology of patients with CRC.

Translation to Routine Clinical Pathology

This is not the first attempt to simplify the CMS classifications; Dunne and co-workers reported that a 30-gene transcriptomic subtyping method could be used to stratify outcome in patients with colorectal cancer. However, this approach was limited as sampling from stromal-rich areas appeared to alter their results [58]. This would suggest that an IHC derived subtyping method would have an advantage as tumour-rich area can readily be identified, avoiding stromal-rich areas, by utilising full block sections for staining and analysis to obviate, as much as possible, the issues around tissue sampling. Trinh et al have previously produced an IHC-based subtyping for CRC utilising 5 genes (CDX2, FRMD6, HTR2B, ZED1 and KER) which showed an 87% concordance with the CMS classifications. However they were only successful in distinguishing epithelial (CMS1-3) and mesenchymal (CMS4) tumours, and had to incorporate MSI to determine CMS1. They could not separate CMS2
and CMS3 [59]. Similarly, Fessler et al could only distinguish CMS4 using miR200 [60]. From this and other previous work in patients with breast cancer it is clear that for subtyping to be taken into routine pathology, there needs to be development of a simple IHC based approach.

From the present review it is clear that the genomic alterations associated with each subtype result in specific phenotypic characteristics. For CMS1 increased immune infiltration, CMS2 increased proliferation, CMS3 metabolic deregulation and CMS4 increased stromal invasion are key phenotypic characteristics. Also, given the prognostic value of such phenotypic characteristics they have the potential to have equivalent or better prognostic value than their genomic equivalent. Moreover, the phenotypic characteristics are more readily translated to routine pathology. To test this hypothesis, we characterised our phenotypic subtypes (Table 3) in a pilot cohort of 237 CRC patients and were able to successfully group patients into the four groups. When assessing the effect of the subtypes of cancer-specific survival, as strong association was observed (p<0.001), with CMS1 and CMS2 having a good prognosis, CMS3 and intermediate prognosis and CMS4 having the worst prognosis (Figure 1). No associations were seen with non-cancer related survival (p=0.192) and only a weak association with overall survival (p=0.043), suggesting these are cancer specific subtypes.

Conclusions

Since most cancers are now recognised as heterogeneous groups of malignancies within a specific organ, subtyping of tumours is becoming a standard approach in the field of cancer research. Since the advent of subtypes for breast cancer and with the rise in whole-genome analysis and RNA-seq, subtypes are now being proposed for stratification of other malignancies, such as pancreatic and lung cancer. For pancreatic ductal adenocarcinoma
(PDAC), RNA-seq analysis of 232 tumours has revealed four subtypes [61], and for lung cancer, whole genome sequencing of 230 tumours also revealed four distinct subtypes [62].

Similarly, there is a plethora of studies suggesting potential classification systems in patients with colorectal cancer. In 2015, a consortium combined information from these studies to produce four consensus subtypes; however, the subtyping focussed around genomic testing which is too costly and unreliable to be moved into routine pathology. After a comprehensive review of the literature, we propose a simple phenotypic subtyping method based on immune infiltrate, stromal invasion and proliferation rate (Table 3), all of which can be readily translated into routine pathology (Figure 2).

If this simple phenotypic method of subtyping proves transferable and reliable upon thorough testing, routine subtyping for CRC tumours could revolutionise the treatment of patients, with each subtype receiving a different therapeutic regime. For example, it is already known that the EGFR monoclonal antibody, cetuximab may not be useful for patients with the CMS1 or CMS3 subtypes [16]. However, CMS1 appears to be the only subtype sensitive to Src family kinase inhibitors such as Dasatinib [19]. It has also been shown that CMS4 is resistant to chemotherapy [17]. Therefore, subtyping of CRC tumours may be the first step toward precision medicine for these patients.

**Funding:** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
References


<table>
<thead>
<tr>
<th>Patients</th>
<th>Components</th>
<th>Subtypes</th>
<th>Prognosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marisa et al.</td>
<td>453 frozen 128 fresh 1058 database CIN, MSI, CIMP, BRAF, KRAS, Immune, stroma,</td>
<td>C1: CIN \text{immune down}</td>
<td>Better prognosis</td>
</tr>
<tr>
<td></td>
<td>proliferation, WNT</td>
<td>C2: dMMR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C3: KRASm</td>
<td>C4: CSC</td>
<td>Worse prognosis</td>
</tr>
<tr>
<td></td>
<td>C5: CIN wnt up</td>
<td>C6: CIN noel.</td>
<td></td>
</tr>
<tr>
<td>Sadanandam et al.</td>
<td>1290 database 387 fresh MSI, WNT, mesenchymal, chemokines, interferon,</td>
<td>Globlet-like</td>
<td>Best prognosis</td>
</tr>
<tr>
<td></td>
<td>enterocyte, stem cell, differentiation, globlet-specific</td>
<td>Enterocyte</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stem-like</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inflammatory</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Transit-amplifying</td>
<td></td>
</tr>
<tr>
<td>Budinska et al.</td>
<td>1833 database MSI, CIMP, BRAF, p53, immune, proliferation, stroma, CSC, side</td>
<td>Surface crypt-like</td>
<td>Best prognosis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lower crypt-like</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CIMP-H</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mesenchymal</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mixed</td>
<td></td>
</tr>
<tr>
<td>Schlicker et al.</td>
<td>62 fresh 1600 database MSI, Immune, Stroma</td>
<td>1.1 Mesenchymal</td>
<td>Worse prognosis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.1 Epithelial</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.2</td>
<td></td>
</tr>
<tr>
<td>Roepman et al.</td>
<td>731 fresh MSI, BRAF, stroma, proliferation, cell type</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Worst prognosis</td>
</tr>
<tr>
<td>De Sousa et al.</td>
<td>1164 database CIN, MSI, CIMP, BRAF, KRAS, stroma, side</td>
<td>CCS1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCS2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCS3</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. The Consensus Molecular Subtypes of CRC.

<table>
<thead>
<tr>
<th></th>
<th>CMS1 MSI Immune</th>
<th>CMS2 Canonical</th>
<th>CMS3 Metabolic</th>
<th>CMS4 Mesenchymal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Microsatellite Instability</strong></td>
<td>MSI</td>
<td>MSS</td>
<td>MSS</td>
<td>MSS</td>
</tr>
<tr>
<td><strong>CIMP</strong></td>
<td>High</td>
<td>Negative</td>
<td>Low</td>
<td>Negative</td>
</tr>
<tr>
<td><strong>SCNA</strong></td>
<td>Normal</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td><strong>Mutations</strong></td>
<td>BRAF</td>
<td>-</td>
<td>KRAS</td>
<td>-</td>
</tr>
<tr>
<td><strong>Immune infiltrate</strong></td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Normal</td>
</tr>
<tr>
<td><strong>Stromal Invasion</strong></td>
<td>Normal</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td><strong>Epithelial/Mesenchymal</strong></td>
<td>-</td>
<td>Epithelial</td>
<td>Epithelial</td>
<td>Mesenchymal</td>
</tr>
<tr>
<td><strong>Pathways</strong></td>
<td>High JAK/STAT</td>
<td>High WNT</td>
<td>High metabolic</td>
<td>High TGFβ/VEGF</td>
</tr>
<tr>
<td><strong>Survival</strong></td>
<td>Good Prognosis</td>
<td>-</td>
<td>-</td>
<td>Poor prognosis</td>
</tr>
<tr>
<td><strong>Subtypes based on</strong></td>
<td>C2[15]</td>
<td>C1&amp;C5[15]</td>
<td>C3[15]</td>
<td>C4&amp;C6[15]</td>
</tr>
<tr>
<td></td>
<td>inflammatory[16]</td>
<td>stem-like[16]</td>
<td>enterocyte &amp; TA[16]</td>
<td>stem-like[16]</td>
</tr>
<tr>
<td></td>
<td>1.2[10]</td>
<td>2.1&amp;2.2[16]</td>
<td>1.3[19]</td>
<td>1.1[19]</td>
</tr>
<tr>
<td></td>
<td>A[17]</td>
<td>B[27]</td>
<td>A/B[17]</td>
<td>C[27]</td>
</tr>
<tr>
<td></td>
<td>CSS2[20]</td>
<td>CSS1[20]</td>
<td>CSS1[20]</td>
<td>CSS3[20]</td>
</tr>
</tbody>
</table>
Table 3. Simple IHC-based testing for Molecular Subtypes of CRC.

<table>
<thead>
<tr>
<th>Intra-tumour immune infiltrate (KM grade: 0-1/2-3)</th>
<th>CMS1 MSI Immune</th>
<th>CMS2 Canonical</th>
<th>CMS3 Metabolic</th>
<th>CMS4 Mesenchymal</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stromal invasion (TSP: ≤50%/&gt;50%)</th>
<th>CMS1 MSI Immune</th>
<th>CMS2 Canonical</th>
<th>CMS3 Metabolic</th>
<th>CMS4 Mesenchymal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proliferation rate (Ki76; ≤50%/&gt;50%)</th>
<th>CMS1 MSI Immune</th>
<th>CMS2 Canonical</th>
<th>CMS3 Metabolic</th>
<th>CMS4 Mesenchymal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any</td>
<td>High</td>
<td>Low</td>
<td>Any</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cancer Specific Survival</th>
<th>CMS1 MSI Immune</th>
<th>CMS2 Canonical</th>
<th>CMS3 Metabolic</th>
<th>CMS4 Mesenchymal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best Prognosis</td>
<td>Good Prognosis</td>
<td>Poor Prognosis</td>
<td>Worst Prognosis</td>
<td></td>
</tr>
</tbody>
</table>


Figure 1. Phenotypic subtypes predict cancer-specific survival in CRC patients. Kaplan Meier curves showing the association between phenotypic subtypes (MSI Immune (solid black), Canonical (dotted black), Metabolic (solid grey) and Mesenchymal (dotted grey)) and (A) cancer-specific survival, (B) Non-cancer related survival and (C) overall survival in 237 CRC patients.

Figure 2. Flow chart showing biomarker testing that can be utilised in the clinic.