

# Molecular pathological classification of colorectal cancer

Mike F. Müller<sup>1</sup> · Ashraf E. K. Ibrahim<sup>2,3</sup> · Mark J. Arends<sup>1</sup>

Received: 30 March 2016 / Revised: 4 May 2016 / Accepted: 9 May 2016 / Published online: 20 June 2016  
© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** Colorectal cancer (CRC) shows variable underlying molecular changes with two major mechanisms of genetic instability: chromosomal instability and microsatellite instability. This review aims to delineate the different pathways of colorectal carcinogenesis and provide an overview of the most recent advances in molecular pathological classification systems for colorectal cancer. Two molecular pathological classification systems for CRC have recently been proposed. Integrated molecular analysis by The Cancer Genome Atlas project is based on a wide-ranging genomic and transcriptomic characterisation study of CRC using array-based and sequencing technologies. This approach classified CRC into two major groups consistent with previous classification systems: (1) ~16 % hypermutated cancers with either microsatellite instability (MSI) due to defective mismatch repair (~13 %) or ultramutated cancers with DNA polymerase epsilon proofreading mutations (~3 %); and (2) ~84 % non-hypermutated, microsatellite stable (MSS) cancers with a high frequency of DNA somatic copy number alterations, which showed common mutations in *APC*, *TP53*, *KRAS*, *SMAD4*, and *PIK3CA*. The recent Consensus Molecular Subtypes (CMS) Consortium analysing CRC expression profiling data from multiple studies described four CMS groups: almost all hypermutated MSI cancers fell into

the first category CMS1 (MSI-immune, 14 %) with the remaining MSS cancers subcategorised into three groups of CMS2 (canonical, 37 %), CMS3 (metabolic, 13 %) and CMS4 (mesenchymal, 23 %), with a residual unclassified group (mixed features, 13 %). Although further research is required to validate these two systems, they may be useful for clinical trial designs and future post-surgical adjuvant treatment decisions, particularly for tumours with aggressive features or predicted responsiveness to immune checkpoint blockade.

**Keywords** Colorectal · Cancer · Polymerase epsilon · Ultramutant · Hypermutant · Defective mismatch repair · Microsatellite instability · Chromosomal instability · Mutation · Somatic copy number alterations · Consensus molecular subtypes · The Cancer Genome Atlas · Serrated pathway

## Introduction

Colorectal cancer (CRC) is the third most common cancer in men and the second most common cancer in women, accounting for about 700,000 deaths per year [1]. The majority of 70–80 % of CRC are sporadic, while around 20–30 % of CRC have a hereditary component, due to either uncommon or rare, high-risk, susceptibility syndromes, such as Lynch Syndrome (LS) (3–4 %) and familial adenomatous polyposis (FAP) (~1 %) [2], or more common but low-risk alleles. Some of the latter, such as *Shroom2*, have been identified by genome-wide association studies (GWAS) [3]. A small subset of about 1–2 % of CRC cases arises as a consequence of inflammatory bowel diseases [4].

CRC is not a homogenous disease, but can be classified into different subtypes, which are characterised by specific molecular and morphological alterations. A major feature of

✉ Mark J. Arends  
M.Arends@ed.ac.uk

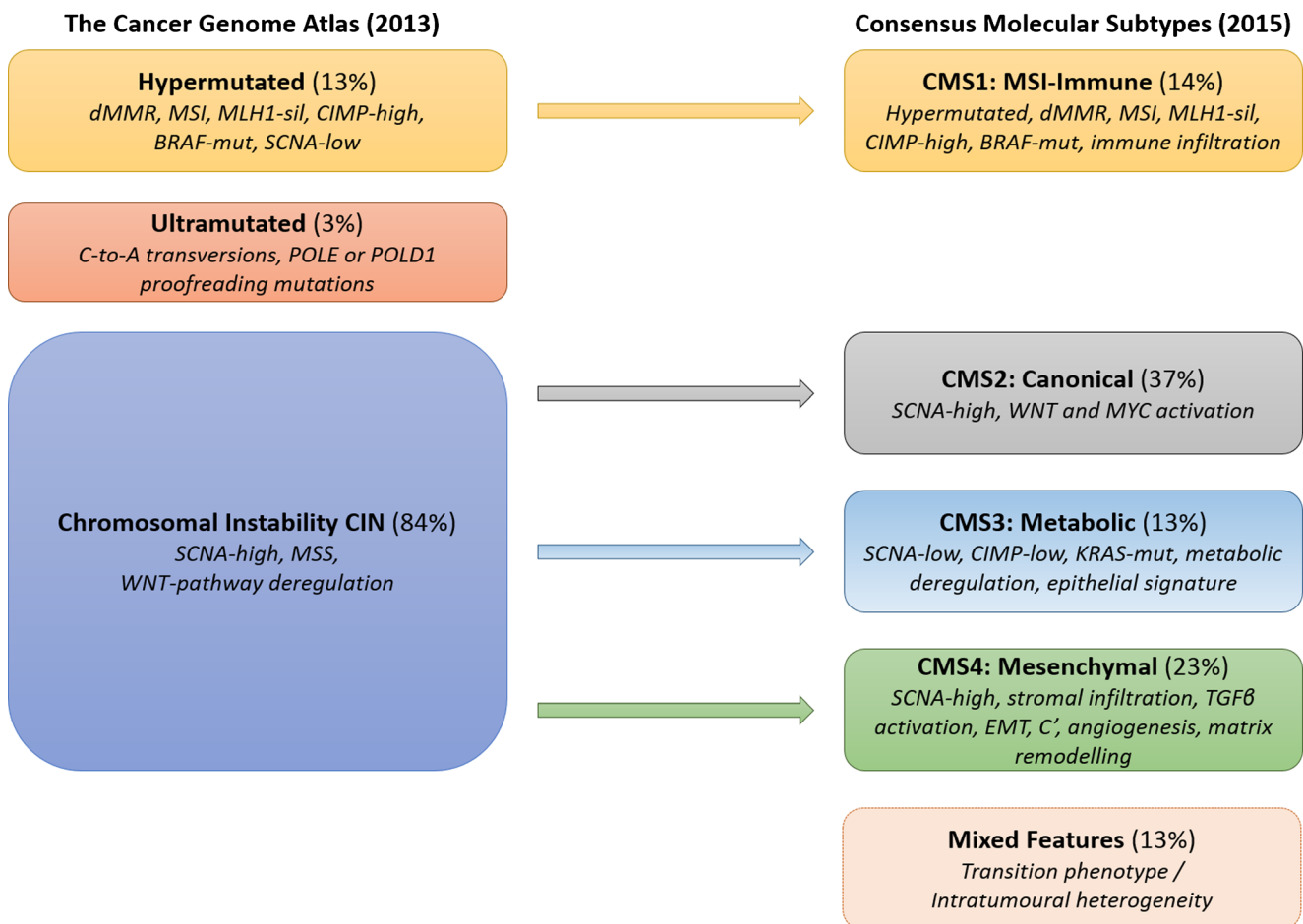
<sup>1</sup> Division of Pathology, Centre for Comparative Pathology, Edinburgh Cancer Research Centre, Institute of Genetics & Molecular Medicine, Western General Hospital, University of Edinburgh, Crewe Road South, Edinburgh EH4 2XR, UK

<sup>2</sup> Department of Pathology, Addenbrooke's Hospital, University of Cambridge, Hills Road, Cambridge CB2 0QQ, UK

<sup>3</sup> Bedford Hospital NHS Trust, Viapath Cellular Pathology, Kempston Road, Bedford MK42 9DJ, UK

CRC is genetic instability that can arise by at least two different mechanisms. The most common (around ~84 % of sporadic CRC) is characterised by chromosomal instability (CIN), with gross changes in chromosome number and structure including deletions, gains, translocations and other chromosomal rearrangements. These are often detectable as a high frequency of DNA somatic copy number alterations (SCNA), which are a hallmark of most tumours that arise by the adenoma-carcinoma sequence [5]. Previous molecular genetic studies have associated CIN with inactivating mutations or losses in the Adenomatous Polyposis Coli (*APC*) tumour suppressor gene, which occur as an early event in the development of neoplasia of the colorectum in this sequence. The

second group (around ~13–16 % of sporadic CRC) are hypermutated and show microsatellite instability (MSI) due to defective DNA mismatch repair (MMR), often associated with wild-type *TP53* and a near-diploid pattern of chromosomal instability (Fig. 1) [6]. Furthermore, CpG island methylation phenotype (CIMP) is a feature that induces epigenetic instability by promoter hypermethylation and silencing of a range of tumour suppressor genes, including *MLH1*, one of the MMR genes [7]. This review provides an overview of the integrated molecular and transcriptomic patterns in CRC, including new insights from The Cancer Genome Atlas (TCGA) project [8] and the Consensus Molecular Subtype (CMS) Consortium [9].



**Fig. 1** Molecular classification systems for colorectal cancers. *On the left* is a representation of The Cancer Genome Atlas integrated molecular classification of colorectal cancers into three groups: (1) ~13 % hypermutated tumours with microsatellite instability due to defective mismatch repair, usually caused by *MLH1* silencing via promoter hypermethylation, with the dMMR pathway causing a hypermutated phenotype resulting from failure to recognise and repair DNA mismatches or insertions/deletions; 80–90 % of sporadic hypermutated cancers have *BRAF* V600E (or similar) mutations; (2) ~3 % ultramutated tumours with DNA Polymerase Epsilon or Delta 1 (*POLE* or *POLD1*) exonuclease domain (proofreading) mutations (EDM), with the malfunctioning enzyme introducing incorrect nucleotides during DNA replication, resulting in an ultramutated phenotype; (3) ~84 % CIN tumours with a high frequency of

DNA SCNAs, a low mutation rate (<8/Mb), microsatellite stability (MSS) and deregulation of the WNT pathway most frequently by *APC* mutation. *On the right* is a representation of the consensus molecular subtypes (CMS) expression signature-based classification with four CMS groups—CMS1 (MSI-immune, 14 %), CMS2 (canonical, 37 %), CMS3 (metabolic, 13 %) and CMS4 (mesenchymal, 23 %), with a residual unclassified group (mixed features, 13 %). Molecular attributes and expression signatures for each CMS group are indicated. (*CIMP* CpG Island methylator phenotype, *CIN* chromosomal instability, *C'* complement activation signature, *CMS* consensus molecular subtypes, *dMMR* defective mismatch repair, *MLH1-sil* *MLH1* silencing by promoter hypermethylation, *MSI* microsatellite instability, *MSS* microsatellite stability, *SCNA* somatic copy number alteration, *POLE* DNA polymerase epsilon (or *DI*, Delta 1)).

## Chromosomal instability is linked to abnormalities of the WNT signalling pathway

CIN tumours usually arise as a consequence of a combination of oncogene activation (e.g. *KRAS*, *PIK3CA*) and tumour suppressor gene inactivation (e.g. *APC*, *SMAD4* and *TP53*) by allelic loss and mutation, which go along with changes in tumour characteristics in the adenoma to carcinoma sequence, as first described by Fearon and Vogelstein in 1990 [10]. A key early event in this pathway is hyperactivation of the WNT signalling pathway, usually arising from mutations of the *APC* gene. Abnormalities of the WNT pathway characterise the majority of sporadic colorectal cancers, as well as tumours that arise in FAP patients [11]. Over 80 % of adenomas and CRC exhibit *APC* mutations and a further 5–10 % are showing mutations or epigenetic changes in other WNT signalling components (e.g.  $\beta$ -catenin) that equally result in hyperactivation of the WNT pathway [12–14]. *APC* is an important negative regulator of the WNT pathway, being a component of the Axin-*APC* degradosome complex that promotes the proteasomal degradation of the WNT effector  $\beta$ -catenin. If this complex is defective as a consequence of mutational inactivation of *APC*, excess  $\beta$ -catenin accumulates within the cytoplasm and translocates into the nucleus where it operates a transcriptional switch leading to activation of *MYC* and many other genes [15]. Perturbation of the WNT pathway leads to a dysregulation of proliferation and differentiation with the development of dysplastic crypts, which progress to adenomas with increasing grade of dysplasia owing to loss of other tumour suppressor genes. The transition from adenoma to invasive carcinoma is usually associated with mutation and/or loss of the *TP53* tumour suppressor gene.

## Defective DNA mismatch repair leads to microsatellite instability in sporadic hypermutated cancers and Lynch syndrome cancers

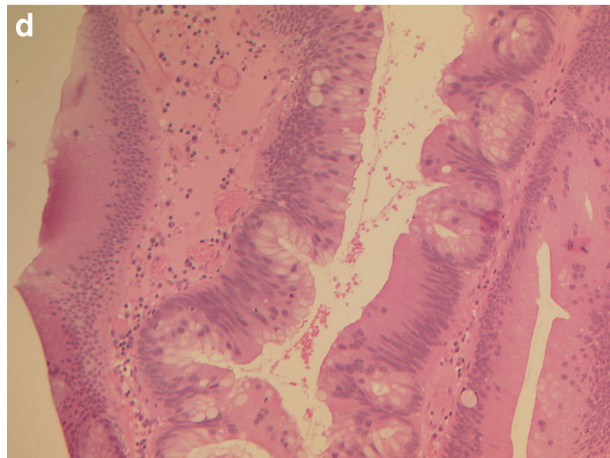
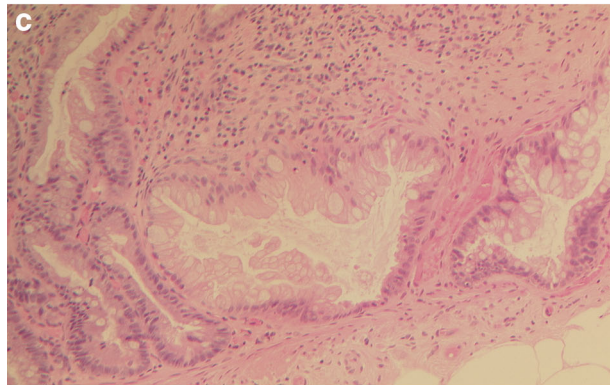
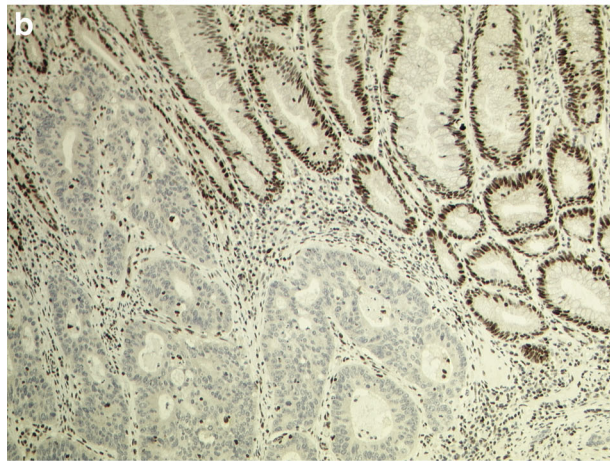
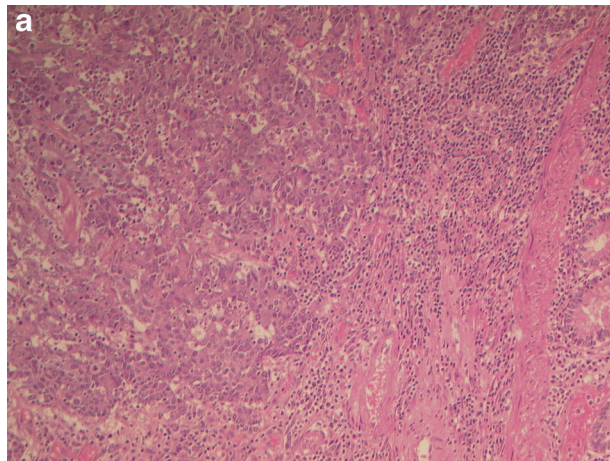
Lynch syndrome (LS), also previously known as hereditary non-polyposis colorectal cancer syndrome (HNPCC), is a syndrome of inherited susceptibility to cancers of several organs, primarily the large bowel, with the next most frequently affected being the endometrium. Moreover, there is also an increased risk of adenocarcinomas of the ovary, stomach, small intestine, transitional cell tumours of ureter and renal pelvis, skin neoplasms (sebaceous tumours and keratoacanthomas), and brain gliomas, amongst others. Development of a neoplasm involves inheriting and acquiring defects in the DNA MMR system in the neoplastic cells. The syndrome is caused by dominant inheritance of a mutant MMR gene (mostly either *MSH2* or *MLH1*), with all somatic cells containing one mutated and one wild-type MMR allele. During tumour formation, there is inactivation of the second MMR allele, by

mutation, deletion or promoter methylation (in the case of the *MLH1* gene), such that the neoplastic cell has inactivated both MMR alleles. In contrast, in sporadic colorectal cancers with defective mismatch repair, the mechanism is almost always (>95 %) promoter hypermethylation of both alleles of the *MLH1* gene, thus silencing *MLH1* expression and crippling the MMR pathway [16–20]. The selective pressure for defective mismatch repair within a neoplasm appears to be due to the reduced susceptibility to apoptosis induced by mismatch-related DNA damage [21–23].

LS colorectal cancers are adenocarcinomas in type, often poorly differentiated or sometimes undifferentiated, occasionally with a dyscohesive appearance. They have prominent tumour-infiltrating lymphocytes and peritumoural Crohns-like lymphoid cell aggregates (Fig. 2) and arise more often in the proximal than in the distal bowel. The major affected genes in LS are *MSH2* and *MLH1*, accounting for 40–45 % LS families each, with the others being mostly due to *MSH6* and *PMS2* mutations (~5–10 % LS families each), with rare LS families having other affected genes [18]. The resulting failure to repair DNA replication-associated mismatch errors in these tumour cells produces a high frequency of mutations, either as single base mismatches or in regions of short tandem DNA repeats (the repeat units often being 1–4 bp in length), known as microsatellites. Thus, DNA extracted from such LS tumours shows variation in length (longer and shorter) of a significant proportion of microsatellites, often more than 30 % of those microsatellite markers tested, a phenomenon known as microsatellite instability at high frequency (MSI-H).

Following DNA damage or most commonly following DNA replication-associated mismatch errors, MMR proteins normally recognise both base mismatches and the insertion/deletion loops (IDLs) that occur in repetitive sequences. Recognition of mismatches and single base IDLs involves the heterodimeric complexes of MutS-related proteins *MSH2* and *MSH6* (known as hMutS-Alpha), whereas IDLs of 2–8 nucleotides are recognised by the complex of *MSH2* and *MSH3* (known as hMutS-Beta). There is overlap in the specificities of these two complexes and hence some redundancy in their activity. A second type of heterodimeric complex, involving two MutL-related proteins, such as either *MLH1* and *PMS2* (hMutL-Alpha), or *MLH1* and *PMS1* (hMutL-Beta), binds to the hMutS complex along with other protein components, so that excision of the recently synthesised error-containing DNA strand occurs and resynthesis of the correct sequence of nucleotides can take place, thus repairing the error [20].

Loss or abnormal expression of the MMR proteins *MLH1*, *MSH2*, *MSH6* and *PMS2*, assessed by immunohistochemistry, is standard practice in many pathology laboratories and is used to help identify LS cancers along with MSI typing of tumour DNA [24–26] (Fig. 2). Distinguishing LS colorectal



◀ **Fig. 2** Integration of morphological and molecular features of colorectal cancer, including the serrated precursors sessile serrated adenoma/polyp and traditional serrated adenoma. **a** Poorly differentiated colorectal cancer (*on the left*) of CMS1 (MSI-immune) with prominent tumour-infiltrating lymphocytes (TILs) and underlying lymphocytes within the submucosa with adjacent muscularis mucosae and crypt bases (*on the right*). **b** Immunohistochemical stain for MLH1 showing loss of expression of MLH1 protein in the adenocarcinoma (*bottom left*) with positive staining for MLH1 in the overlying adenoma (*top right*) and adjacent lymphoid and stromal cells. **c** Sessile serrated adenoma/polyp showing a high-power view of the bases of dilated and serrated crypts with boot-shaped architecture and horizontal growth along the top of the muscularis mucosae, with mild nuclear enlargement but no dysplasia. **d** Traditional serrated adenoma showing a high-power view of an elongated dysplastic crypt with small lateral ectopic budding crypts, projecting at 90° to the main axis of the long crypt. The nuclei are elongated, displaying a pencillate pattern of low-grade dysplasia. (All photomicrographs taken at  $\times 100$  magnification)

cancers that show loss of MLH1 expression from sporadic MMR-deficient cancers is currently most appropriately performed by detection of the specific mutation *BRAF* V600E, which is found in around 80–90 % of sporadic MSI-H colorectal cancers, but rarely—if ever—in colorectal cancers due to Lynch syndrome [6, 27–31]. The presence of *MLH1* promoter hypermethylation may be used to distinguish sporadic CRC from Lynch syndrome-associated CRC, but there are interpretative problems as constitutive *MLH1* promoter methylation may occur, as well as technical challenges of performing this test [19]. In addition to *MLH1*, there are a number of other genes displaying DNA promoter hypermethylation changes, sometimes referred to as CIMP-genes, but there is some disagreement regarding which are the most reliable CIMP-genes and which tests to use for identification of CIMP tumours [7, 14, 32].

### Correlation of molecular pathways with serrated morphology

In addition to CRC development via the well-described adenoma-carcinoma sequence, it is estimated that about 10–20 % of carcinomas may develop via a different sequence of morphological changes, known as the serrated pathway. While the majority of serrated polyps (80–90 %) can be characterised as hyperplastic polyps, which are considered benign bystander lesions, a subset of serrated lesions can progress to colorectal carcinoma. The two premalignant precursor lesions are traditional serrated adenomas (TSA) and sessile serrated adenomas/polyps (SSA/P) (termed sessile serrated adenomas or alternatively sessile serrated polyps, previous European recommendations have also suggested the term sessile serrated lesions) [33, 34] (Fig. 2).

Cancers arising via the two serrated pathways are heterogeneous in terms of molecular patterns and cannot easily be classified based on characteristic mutations, but rather by

specific morphologic changes. A common feature of the serrated pathways is mutations in *KRAS* or *BRAF*, leading to hyperactivation of the MAPK pathway. Furthermore EphB2 can be downregulated by genomic loss or promoter methylation, also resulting in MAPK hyperactivation [33, 35, 36]. The characteristic morphological features of the traditional serrated pathway such as architectural dysplasia with ectopic budding crypt formation and epithelial serrations are likely to be linked with these molecular alterations that result in hyperproliferation and inhibition of apoptosis [33, 37–39].

TSAs are more often diagnosed in the left colon. They frequently (~80 %) have *KRAS* mutations or less often (20–30 %) *BRAF* mutations and are microsatellite stable (MSS) or MSI-L. They are diagnosed based on characteristic cytology (eosinophilic cytoplasm, central, elongated hyperchromatic nuclei) and slit-like epithelial serrations with ectopic crypt formation and may progress to adenocarcinoma (traditional serrated pathway) [35, 40].

SSA/P frequently occur in the right colon, and they tend to have *BRAF* mutations (~80 %). CIMP is an early feature of SSA/P and often leads to MSI, related to *MLH1* promoter hypermethylation. Also, *MTMG* can be silenced by promoter methylation, which on its own results in an MSI-L phenotype. SSA/P are characterised by abnormally shaped (boot, inverted-anchor, J, L or inverted T) crypts or horizontal growth along the muscularis mucosae, with crypt dilatation and serration extending down to the crypt base [41]. These architectural changes (without genuine dysplasia) are the hallmark of SSA/P and are believed to result from a displacement of the maturation zone [33, 41, 42]. SSA/P may progress to serrated or mucinous adenocarcinomas (sessile serrated pathway).

Colorectal cancers arising via the serrated pathways have been recognised as a distinct subtype overlapping with CIN and MSI tumours by molecular profiling, and are strongly associated with poor prognosis and therapy resistance. Since EMT and matrix remodelling proteins are upregulated in these lesions, it was hypothesised that this predisposes CRC developing via the serrated pathways to invasiveness and metastasis at an early stage [43]. Subsequent analysis revealed that MSI, which often develops within SSA/P, resulted in a more favourable prognosis, whereas MSS in carcinomas derived from SSA/P, and more often from TSA, was linked to poor prognosis [35, 36].

### Integrated genomic characterisation of colorectal cancers (TCGA classification)

The TCGA network project collected colorectal tumour samples and corresponding germline DNA samples from 276 patients for exome sequencing of 224 cancers with paired normal samples, along with DNA SCNA analysis, promoter

methylation, messenger RNA (mRNA) and micro RNA (miRNA) studies. Ninety-seven samples underwent whole genome sequencing. The clinical and pathological characteristics reflected the typical cross-section of patients with CRC, so this data provides a valuable source of information to gain further insights into the molecular pathology of CRC [8].

The analysis revealed that the bowel cancers could be split into two major groups by mutation rate—non-hypermutated and hypermutated cancers—which by characteristics and frequency match well with the previously discussed CIN and MSI pathways (Fig. 1, Table 1). The hypermutated category was further subdivided in two subgroups. While the majority of tumours in this group (~13 % of the analysed tumours) were hypermutated cancers due to defective mismatch repair (dMMR) with a high mutation rate of 12–40 mutations/Mb, a small subgroup (~3 % of the analysed tumours) had an extremely high mutation rate of >40 mutations/Mb and were thus called ultramutated cancers. The dMMR of the hypermutated cancers resulted from acquired hypermethylation of the *MLH1* promoter in almost all cases, leading to the silencing of expression of *MLH1* and non-functioning mismatch repair, which is again in accordance with the previously discussed findings. Almost all of these tumours showed CIMP characteristics, with several other specifically tested genes also demonstrating promoter methylation. A small number of cancers showed either inherited (LS/HNPCC) or somatic MMR gene mutations. The ultramutated colorectal carcinomas had an extremely high mutation rate with a characteristic nucleotide base change spectrum with increased C-to-A transversions, resulting from the presence of a mutation that inactivates the proofreading function within the exonuclease domain of the polymerase E (POLE) DNA replicating enzyme, or rarely of *POLD1*. This resulted in failure to correct the misincorporation of nucleotides during DNA replication or repair by mutant POLE (or D1).

Other studies [44, 45] have shown that less than 0.1 % of CRC have inherited mutations at characteristic sites within the exonuclease domain of either *POLE* (p.Leu424Val) or *POLD1* (p.Ser478Asn), which are the basis of the polymerase-proofreading-associated polyposis (PPAP) syndrome that is characterised by increased colorectal adenomas and adenocarcinomas as well as increased risk of endometrial cancer in the case of inherited *POLD1* mutations [44]. The group of non-hypermutated cancers with a low mutation rate (<8 mutations/Mb) mostly demonstrated a high SCNA frequency, making up the majority (~84 %) of colorectal adenocarcinomas that were MSS due to an intact MMR pathway.

Ultramutated and hypermutated cancers were combined into a single group and compared with the low mutation rate MSS tumours. Overall, 32 genes were recurrently mutated and after removal of non-expressed genes, there were 15 and 17 recurrently mutated genes in the hypermutated and non-hypermutated bowel cancer groups, respectively. The significantly mutated genes in the hypermutated cancers included *ACVR2A* (63 %), *APC* (51 %), *TGFBR2* (51 %), *BRAF* (46 %), *MSH3* (40 %), *MSH6* (40 %), *MYOBI* (31 %), *TCF7L2* (31 %), *CASP8* (29 %), *CDC27* (29 %), *FZD3* (29 %), *MIER3* (29 %), *TCERG1* (29 %), *MAP7* (26 %), *PTPN12* (26 %) and *TP53* (20 %). The genes that were recurrently mutated in the non-hypermutated MSS colorectal cancers included mutations in *APC* (81 %), *TP53* (60 %), *KRAS* (43 %), *TTN* (31 %), *PIK3CA* (18 %), *FBXW7* (11 %), *SMAD4* (10 %), *NRAS* (9 %), *TCF7L2* (9 %), *FAM123B*, also known as *WTX*, (7 %), *SMAD2* (6 %), *CTNNB1* (5 %), *KIAA1804* (4 %), *SOX9* (4 %), *ACVR1B* (4 %), *GPC6* (40 %) and *EDNRB* (3 %). The tumour suppressor genes *ATM* and *ARID1A* showed a disproportionately high percentage of nonsense or frameshift mutations. The *KRAS* and *NRAS* mutations were activating oncogenic mutations at codons 12,

**Table 1** Characteristics of colorectal cancers in TCGA integrated molecular classification

Group	(1a) Ultramutated <i>POLE</i> mutant	(1b) Hypermutated dMMR/MSI	(2) CIN/SCNA-high, MSS
Mutation rate	++++	+++	+
Somatic copy number alterations	+/-	+	+++
Key molecular/genetic abnormality	<i>POLE</i> EDM proofreading mutation	Defective MMR/ <i>MLH1</i> promoter hypermethylation	Variety of mutated cancer genes; WNT pathway activation (mostly by <i>APC</i> mutation/inactivation)
Predominant histological type	Moderately differentiated adenocarcinoma	Mucinous, or signet ring, or poorly differentiated adenocarcinoma	Moderately differentiated adenocarcinoma
Proportion of all colorectal carcinomas	~3 %	~13 %	~84 %
Prognosis	Good (more data required)	Good/poor after relapse	Good-poor (depending on other characteristics)

CIN chromosomal instability, *POLE* DNA polymerase epsilon, *EDM* exonuclease domain mutant, *SCNA* somatic copy number alteration, *MMR* mismatch repair, *MSI* microsatellite instability

13 and 61, and the *BRAF* mutation was the classical V600E activating mutation, whereas the other genes almost entirely had inactivating mutations.

Colonic and rectal cancers were combined for the analysis of the non-hypermuted MSS group, as they showed no distinguishable molecular differences. SCNA patterns in non-hypermuted MSS tumours confirmed the previously well-documented [5] chromosomal arm-level changes of significant gains of 1q, 7p, 7q, 8p, 8q, 12q, 13q, 19q and 20p, and significant deletions of 1p, 4q, 5q, 8p, 14q, 15q, 17p (includes *TP53*) and 17q, 18q (includes *SMAD4*), 20p and 22q. Hypermuted MSI cancers had far fewer SCNAs, but a similar pattern of chromosomal arm gains and losses. There were 28 recurrent deletion peaks that included the genes *FHIT*, *RBFOX1*, *WWOX*, *SMAD4*, *APC*, *PTEN*, *SMAD3* and *TCF7L2*. Other studies have identified *PARK2* as another recurrently deleted gene on chromosome 6 in around a third of CRCs [46]. A chromosomal translocation generating a gene fusion of *TCF7L2* and *VTI1A* was seen in 3 % of CRC and also *NAV2-TCF7L1* fusion in three cancers. Focal amplifications were seen affecting *MYC*, *ERBB2*, *IGF2*, *USP12*, *CDK8*, *KLF5*, *HNF4A*, *WHSC1L1/FGFR1* and gains of *IRS2* [47].

The most frequently altered pathways by gene mutations, deletions, amplifications and translocations were activation of the WNT, MAPK and PI3K signalling pathways, and deactivation of the TGF- $\beta$  and P53 inhibitory pathways, which may be relevant for targeted therapies. The WNT signalling pathway was activated in 93 % of non-hypermuted and 97 % of hypermutated cancers, involving biallelic inactivation of *APC* or activation of *CTNNB1* in over 80 % of tumours, together with changes to many other genes involved in regulation of the WNT pathway (*TCF7L2*, *DKK*, *AXIN2*, *FBXW7*, *ARID1A*, *FAM123B*, *FZD10* and *SOX9*). Alterations affecting either the MAPK (*ERBB2*, *RAS* genes, *BRAF*) or PI3K (*PIK3CA*, *PIK3R1*, *PTEN*, *IGF2*, *IRS2*) signalling pathways were relatively common, often showing patterns of mutual exclusivity of gene mutations (for *RAS* and *BRAF* or for *PIK3CA*, *PIK3R1* and *PTEN*). The TGF- $\beta$  pathway was deregulated by alterations to *TGFBR1*, *TGFBR2*, *ACVR2A*, *ACVR1B*, *SMAD2*, *SMAD3* and *SMAD4* in 27 % of non-hypermuted MSS tumours and 87 % of hypermutated cancers. The P53 pathway was affected by mutations to *TP53* (60 %) and *ATM* (7 %) in a near mutually exclusive pattern in non-hypermuted MSS bowel cancers. An integrated data analysis showed that nearly all tumours displayed dysregulation of *MYC* transcriptional targets as a result of *MYC* activation by activated WNT signalling and/or dysregulation of TGF- $\beta$  signalling, indicating an important role for *MYC* in colorectal cancer. Using CRC resection data on stage, nodal status, distant metastasis and vascular invasion, some molecular changes were associated with aggressive features including those affecting *SCN5A*, *APC*, *TP53*, *PIK3CA*, *BRAF* and *FBXW7* as well as altered expression of

some miRNAs. Potential therapeutic approaches suggested by the TCGA classification are targeting of IGF2, IGF1R, ERBB2, ERBB3, MEK, AKT and mTOR proteins as well as possible WNT pathway inhibitors.

### Colorectal cancer gene expression profiling (CMS Classification)

Early attempts at gene expression profiling in order to stratify CRC were made by several groups, but showed little agreement with each other, suggesting different categories, and did not lead to a useful single consistent classification system [43, 48–53]. Subsequently, an international expert consortium [9] recently reached an agreement that describes four consensus molecular subtypes (CMS) after analysis of 18 different CRC gene expression datasets, including data from TCGA in conjunction with molecular data on mutations and SCNAs for a subset of the samples (Fig. 1).

CMS1 (MSI-immune, 14 %) CRC were hypermutated due to defective DNA mismatch repair with MSI and MLH1 silencing and accordingly CIMP-high with frequent *BRAF* mutations, while having a low number of SCNAs. This equates with the previously well-characterised sporadic MSI CRC subgroup. Gene expression profiling furthermore revealed evidence of strong immune activation (immune response, PD1 activation, NK cell, Th1 cell and cytotoxic T cell infiltration signatures) in CMS1, consistent with pathological descriptions of prominent tumour-infiltrating CD8+ cytotoxic T lymphocytes. Patients with the CMS1 subtype had a very poor survival rate after relapse.

The majority of CRC previously described as CIN was split into three subcategories based on transcriptomic profiling, which consequently were all characterised by high levels of SCNAs. CMS2 (canonical, 37 %) CRC predominantly displayed epithelial signatures with prominent WNT and *MYC* signalling activation, and more often displayed loss of tumour suppressor genes and copy number gains of oncogenes than the other subtypes. CMS2 patients had a better survival rate after relapse compared with the other subtypes. The CMS3 (metabolic, 13 %) subtype had fewer SCNAs and contained more hypermutated/MSI samples than CMS2 and CMS4, along with frequent *KRAS* mutations and a slightly higher prevalence of CIMP-low. Gene expression analysis of CMS3 found predominantly epithelial signatures and evidence of metabolic dysregulation in a variety of pathways. The CMS4 subtype (mesenchymal, 23 %) CRC showed increased expression of EMT genes and evidence of prominent transforming growth factor- $\beta$  activation, with expression of genes implicated in complement-associated inflammation, matrix remodelling, stromal invasion and

angiogenesis. Patients with the CMS4 subtype had a worse overall survival and worse relapse-free survival than patients of the other groups. Finally, there were some samples with mixed features (13 %) that possibly represent either a transition phenotype or intratumoural heterogeneity.

This CMS classification system has been suggested by the authors to be the most robust classification system currently available for CRC based on biological processes related to gene expression patterns and is suggested as a basis for future clinical stratification in trials and other studies with potential for subtype-based targeted interventions, although further studies are required to validate this assertion.

## Conclusion

In conclusion, integration of wide-ranging molecular data has generated two systems of classification of colorectal cancers (Fig. 1, Table 1). (A) TCGA classification—tumours with a very high mutation rate which can be further subdivided into either (1a) ultramutated colorectal cancers (~3 %) with DNA polymerase epsilon (POLE) proofreading domain mutations, or (1b) hypermutated colorectal cancers (~13 %) with microsatellite instability due to defective mismatch repair; and (2) colorectal cancers (~84 %) with a low mutation rate but a high frequency of DNA SCNAs. (B) The CMS classification describes four CMS groups—CMS1 (MSI-immune activation, 14 %), CMS2 (canonical, 37 %), CMS3 (metabolic, 13 %) and CMS4 (mesenchymal, 23 %), with a residual unclassified group (mixed features, 13 %). Further research is required to develop more easily applicable molecular tests, such as low-coverage high-throughput sequencing for DNA SCNA analysis and/or cancer gene panel mutation detection, and preferably easily applicable and useful immunohistochemical markers for these CMS subdivisions. Analysis of expression of the MMR proteins and/or MSI testing is currently efficient at identifying the group of defective mismatch repair MSI tumours (CMS1). Both classification systems agree on identification of this dMMR/MSI group, which has recently been shown to respond well to immune checkpoint blockade (antibodies to PD-1) that activates cytotoxic T cell attacks on tumour cells, which is suggested to be related to the large numbers of neo-antigens generated by dMMR [54, 55]. A straightforward and routinely applicable molecular test using PCR and sequencing for identification of POLE (and POLD1) proofreading mutations associated with ultramutated cancer may be performed in molecular pathology laboratories, although in the future a mutation-specific POLE antibody for immunohistochemistry may be developed to aid routine subclassification. Ultramutated cancers are likely to generate higher levels of neo-antigens and may also respond well to immune checkpoint blockade therapy. Selected transcript

expression profiling kits for CMS classification may be required for application of this system. Both classification systems have been proposed to allow better prognostication and are potentially important for future use in clinical trials and for multidisciplinary team discussions about post-surgical adjuvant treatment, including immune checkpoint blockade.

**Compliance with ethical standards** There was full compliance with ethical standards in the writing of this review article. No original research work with patient samples by the authors was involved.

**Funding** No funding was required for the writing of this review article.

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin D, Forman D, Bray F (2013) GLOBOCAN 2012 v1.0, cancer incidence and mortality worldwide: IARC CancerBase No. 11 [Internet]. International Agency for Research on Cancer. <http://globocan.iarc.fr>. Accessed 11/03/2016
2. Whiffin N, Hosking FJ, Farrington SM, Palles C, Dobbins SE, Zgaga L, Lloyd A, Kinnorsley B, Gorman M, Tenesa A (2014) Identification of susceptibility loci for colorectal cancer in a genome-wide meta-analysis. *Hum Mol Genet* 23(17):4729–4737
3. Dunlop MG, Dobbins SE, Farrington SM, Jones AM, Palles C, Whiffin N, Tenesa A, Spain S, Broderick P, Ooi L-Y (2012) Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. *Nat Genet* 44(7):770–776
4. Munkholm P (2003) The incidence and prevalence of colorectal cancer in inflammatory bowel disease. *Aliment Pharmacol Ther* 18:1–5. doi:10.1046/j.1365-2036.18.s2.2.x
5. Pouligiannis G, Ichimura K, Hamoudi RA, Luo F, Leung SY, Yuen ST, Harrison DJ, Wyllie AH, Arends MJ (2010) Prognostic relevance of DNA copy number changes in colorectal cancer. *J Pathol* 220(3):338–347
6. Arends MJ (2013) Pathways of colorectal carcinogenesis. *Appl Immunohistochem Mol Morphol* 21(2):97–102. doi:10.1097/PAL.0b013e3182849808
7. Ibrahim AE, Arends MJ, Silva A-L, Wyllie AH, Greger L, Ito Y, Vowler SL, Huang TH, Tavaré S, Murrell A (2010) Sequential DNA methylation changes are associated with DNMT3B overexpression in colorectal neoplastic progression. *Gut*:gut. 2010.223602
8. The Cancer Genome Atlas Network (326 collaborators). Comprehensive molecular characterization of human colon and rectal cancer (2012). *Nature* 487 (7407):330–337. doi: <http://www.nature.com/nature/journal/v487/n7407/abs/nature11252.html#supplementary-information>
9. Guinney J, Dienstmann R, Wang X, de Reynies A, Schlicker A, Soneson C, Marisa L, Roepman P, Nyamundanda G, Angelino P, Bot BM, Morris JS, Simon IM, Gerster S, Fessler E, De Sousa E Melo F, Missiaglia E, Ramay H, Barras D, Homicsko K, Maru D, Manyam GC, Broom B, Boige V, Perez-Villamil B, Laderas T,



- Salazar R, Gray JW, Hanahan D, Tabernero J, Bernards R, Friend SH, Laurent-Puig P, Medema JP, Sadanandam A, Wessels L, Delorenzi M, Kopetz S, Vermeulen L, Tejpar S (2015) The consensus molecular subtypes of colorectal cancer. *Nat Med* 21(11):1350–1356
10. Fearon ER, Vogelstein B (1990) A genetic model for colorectal tumorigenesis. *Cell* 61(5):759–767
  11. Frayling I, Arends M (2013) Adenomatous Polyposis Coli. In: Maloy S, Hughes K (eds) *Brenner's Encyclopedia of Genetics*, vol 1, 2nd edition edn. Academic, San Diego, pp 27–29
  12. Morin PJ, Sparks AB, Korinek V, Barker N, Clevers H, Vogelstein B, Kinzler KW (1997) Activation of  $\beta$ -catenin-Tcf signaling in colon cancer by mutations in  $\beta$ -catenin or APC. *Science* 275(5307):1787–1790. doi:10.1126/science.275.5307.1787
  13. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, Smits AM, Bos JL (1988) Genetic alterations during colorectal-tumor development. *N Engl J Med* 319(9):525–532
  14. Silva A-L, Dawson SN, Arends MJ, Guttula K, Hall N, Cameron EA, Huang TH, Brenton JD, Tavaré S, Bienz M (2014) Boosting Wnt activity during colorectal cancer progression through selective hypermethylation of Wnt signaling antagonists. *BMC Cancer* 14(1):891
  15. Bienz M, Clevers H (2000) Linking colorectal cancer to Wnt signaling. *Cell* 103(2):311–320
  16. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Rüschoff J, Fishel R, Lindor NM, Burgart LJ, Hamelin R (2004) Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst* 96(4):261–268
  17. Meyer LA, Broaddus RR, Lu KH (2009) Endometrial cancer and Lynch syndrome: clinical and pathologic considerations. *Cancer Control: J Moffitt Cancer Center* 16(1):14–22
  18. Poulogiannis G, Frayling IM, Arends MJ (2010) DNA mismatch repair deficiency in sporadic colorectal cancer and Lynch syndrome. *Histopathology* 56(2):167–179
  19. Gay LJ, Arends MJ, Mitrou PN, Bowman R, Ibrahim AE, Happerfield L, Luben R, McTaggart A, Ball RY, Rodwell SA (2011) MLH1 promoter methylation, diet, and lifestyle factors in mismatch repair deficient colorectal cancer patients from EPIC-Norfolk. *Nutr Cancer: Int J* 63(7):1000–1010. doi:10.1080/01635581.2011.596987
  20. Ibrahim AE, Arends MJ (2012) Molecular typing of colorectal cancer: applications in diagnosis and treatment. *Diagn Histopathol* 18(2):70–80
  21. Tomlinson IP, Novelli M, Bodmer W (1996) The mutation rate and cancer. *Proc Natl Acad Sci* 93(25):14800–14803
  22. Toft NJ, Winton DJ, Kelly J, Howard LA, Dekker M, te Riele H, Arends MJ, Wyllie AH, Margison GP, Clarke AR (1999) Msh2 status modulates both apoptosis and mutation frequency in the murine small intestine. *Proc Natl Acad Sci* 96(7):3911–3915. doi:10.1073/pnas.96.7.3911
  23. Toft NJ, Curtis LJ, Sansom OJ, Leitch AL, Wyllie AH, te Riele H, Arends MJ, Clarke AR (2002) Heterozygosity for p53 promotes microsatellite instability and tumorigenesis on a Msh2 deficient background. *Oncogene* 21(41):6299–6306
  24. Arends M, Ibrahim M, Happerfield L, Frayling I, Miller K (2008) Interpretation of immunohistochemical analysis of mismatch repair (MMR) protein expression in tissue sections for investigation of suspected Lynch/Hereditary Non-Polyposis Colorectal Cancer (HNPCC) syndrome. *UK NEQAS ICC & ISH Recommendations 1*
  25. Arends M, Frayling I, Happerfield L, Ibrahim M (2010) HNPCC/Lynch syndrome module: report of the immunohistochemical analysis of mismatch repair (MMR) protein expression. *UK NEQAS ICC & ISH Recommendations 8*
  26. Giardiello FM, Allen JI, Axilbund JE, Boland CR, Burke CA, Burt RW, Church JM, Dominitz JA, Johnson DA, Kaltenbach T, Levin TR, Lieberman DA, Robertson DJ, Syngal S, Rex DK (2014) Guidelines on genetic evaluation and management of Lynch syndrome: a consensus statement by the US multi-society task force on colorectal cancer. *Gastroenterology* 147(2):502–526. doi:10.1053/j.gastro.2014.04.001
  27. Domingo E, Niessen RC, Oliveira C, Alhopuro P, Moutinho C, Espín E, Armengol M, Sijmons RH, Kleibeuker JH, Seruca R (2005) BRAF-V600E is not involved in the colorectal tumorigenesis of HNPCC in patients with functional MLH1 and MSH2 genes. *Oncogene* 24(24):3995–3998
  28. Kawaguchi M, Yanokura M, Banno K, Kobayashi Y, Kuwabara Y, Kobayashi M, Nomura H, Hirasawa A, Susumu N, Aoki D (2009) Analysis of a correlation between the BRAF V600E mutation and abnormal DNA mismatch repair in patients with sporadic endometrial cancer. *Int J Oncol* 34(6):1541–1547
  29. Naguib A, Mitrou PN, Gay LJ, Cooke JC, Luben RN, Ball RY, McTaggart A, Arends MJ, Rodwell SA (2010) Dietary, lifestyle and clinicopathological factors associated with BRAF and K-ras mutations arising in distinct subsets of colorectal cancers in the EPIC Norfolk study. *BMC Cancer* 10:99. doi:10.1186/1471-2407-10-99
  30. Naguib A, Cooke JC, Happerfield L, Kerr L, Gay LJ, Luben RN, Ball RY, Mitrou PN, McTaggart A, Arends MJ (2011) Alterations in PTEN and PIK3CA in colorectal cancers in the EPIC Norfolk study: associations with clinicopathological and dietary factors. *BMC Cancer* 11:123. doi:10.1186/1471-2407-11-123
  31. Metcalf AM, Spurdle AB (2013) Endometrial tumour BRAF mutations and MLH1 promoter methylation as predictors of germline mismatch repair gene mutation status: a literature review. *Familial Cancer* 13(1):1–12. doi:10.1007/s10689-013-9671-6
  32. Berg M, Hagland HR, Søreide K (2014) Comparison of CpG island methylator phenotype (CIMP) frequency in colon cancer using different probe-and gene-specific scoring alternatives on recommended multi-gene panels. *PLoS One* 9(1), e86657
  33. Noffsinger AE (2009) Serrated polyps and colorectal cancer: new pathway to malignancy. *Annu Rev Pathol Mech Dis* 4:343–364
  34. Quirke P, Risio M, Lambert R, von Karsa L, Vieth M (2011) Quality assurance in pathology in colorectal cancer screening and diagnosis—European recommendations. *Virchows Arch* 458(1):1–19
  35. Bettington M, Walker N, Clouston A, Brown I, Leggett B, Whitehall V (2013) The serrated pathway to colorectal carcinoma: current concepts and challenges. *Histopathology* 62(3):367–386
  36. Bettington ML, Walker NI, Rosty C, Brown IS, Clouston AD, McKeone DM, Pearson S-A, Klein K, Leggett BA, Whitehall VL (2015) A clinicopathological and molecular analysis of 200 traditional serrated adenomas. *Mod Pathol* 28(3):414–427
  37. Rad R, Cadiñanos J, Rad L, Varela I, Strong A, Kriegl L, Constantino-Casas F, Eser S, Hieber M, Seidler B, Price S, Fraga Mario F, Calvanese V, Hoffman G, Ponstingl H, Schneider G, Yusa K, Grove C, Schmid Roland M, Wang W, Vassiliou G, Kirchner T, McDermott U, Liu P, Saur D, Bradley A (2013) A genetic progression model of BrafV600E-induced intestinal tumorigenesis reveals targets for therapeutic intervention. *Cancer Cell* 24(1):15–29. doi:10.1016/j.ccr.2013.05.014
  38. Feng Y, Bommer GT, Zhao J, Green M, Sands E, Zhai Y, Brown K, Burberry A, Cho KR, Fearon ER (2011) Mutant Kras promotes hyperplasia and alters differentiation in the colon epithelium but does not expand the presumptive stem cell pool. *Gastroenterology* 141(3):1003–1013. doi:10.1053/j.gastro.2011.05.007, e1010
  39. Tateyama H, Li W, Takahashi E, Miura Y, Sugiura H, Eimoto T (2002) Apoptosis index and apoptosis-related antigen expression in serrated adenoma of the colorectum: the saw-toothed structure may be related to inhibition of apoptosis. *Am J Surg Pathol* 26(2):249–256
  40. Longacre TA, Fenoglio-Preiser CM (1990) Mixed hyperplastic adenomatous polyps/serrated adenomas: a distinct form of colorectal neoplasia. *Am J Surg Pathol* 14(6):524–537
  41. Chetty R, Bateman AC, Torlakovic E, Wang LM, Gill P, Al-Badri A, Arends M, Biddlestone L, Burroughs S, Carey F (2014) A pathologist's survey on the reporting of sessile serrated adenomas/polyps. *J Clin Pathol* 67(5):426–430

42. Torlakovic E, Snover DC (1996) Serrated adenomatous polyposis in humans. *Gastroenterology* 110(3):748–755. doi:[10.1053/gast.1996.v110.pm8608884](https://doi.org/10.1053/gast.1996.v110.pm8608884)
43. Felipe De Sousa EM, Wang X, Jansen M, Fessler E, Trinh A, de Rooij LP, de Jong JH, de Boer OJ, van Leersum R, Bijlsma MF (2013) Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. *Nat Med* 19(5):614–618
44. Church DN, Briggs SE, Palles C, Domingo E, Kearsy SJ, Grimes JM, Gorman M, Martin L, Howarth KM, Hodgson SV (2013) DNA polymerase  $\epsilon$  and  $\delta$  exonuclease domain mutations in endometrial cancer. *Hum Mol Genet* 22(14):2820–2828
45. Briggs S, Tomlinson I (2013) Germline and somatic polymerase and  $\delta$  mutations define a new class of hypermutated colorectal and endometrial cancers. *J Pathol* 230(2):148–153
46. Poulogiannis G, McIntyre RE, Dimitriadi M, Apps JR, Wilson CH, Ichimura K, Luo F, Cantley LC, Wyllie AH, Adams DJ, Arends MJ (2010) PARK2 deletions occur frequently in sporadic colorectal cancer and accelerate adenoma development in Apc mutant mice. *Proc Natl Acad Sci U S A* 107(34):15145–15150. doi:[10.1073/pnas.1009941107](https://doi.org/10.1073/pnas.1009941107)
47. Day E, Poulogiannis G, McCaughan F, Mulholland S, Arends MJ, Ibrahim AE, Dear PH (2013) IRS2 is a candidate driver oncogene on 13q34 in colorectal cancer. *Int J Exp Pathol* 94(3):203–211
48. Budinska E, Popovici V, Tejpar S, D'Ario G, Lapique N, Sikora KO, Di Narzo AF, Yan P, Hodgson JG, Weinrich S (2013) Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer. *J Pathol* 231(1):63–76
49. Roepman P, Schlicker A, Tabernero J, Majewski I, Tian S, Moreno V, Snel MH, Chresta CM, Rosenberg R, Nitsche U, Macarulla T, Capella G, Salazar R, Orphanides G, Wessels LFA, Bernardis R, Simon IM (2014) Colorectal cancer intrinsic subtypes predict chemotherapy benefit, deficient mismatch repair and epithelial-to-mesenchymal transition. *Int J Cancer* 134(3):552–562. doi:[10.1002/ijc.28387](https://doi.org/10.1002/ijc.28387)
50. Sadanandam A, Lyssiotis CA, Homicsko K, Collisson EA, Gibb WJ, Wullschlegel S, Ostos LCG, Lannon WA, Grotzinger C, Del Rio M (2013) A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat Med* 19(5):619–625
51. Marisa L, de Reyniès A, Duval A, Selves J, Gaub MP, Vescovo L, Etienne-Grimaldi M-C, Schiappa R, Guenot D, Ayadi M, Kirzin S, Chazal M, Fléjou J-F, Benchimol D, Berger A, Lagarde A, Pencreach E, Piard F, Elias D, Parc Y, Olschwang S, Milano G, Laurent-Puig P, Boige V (2013) Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med* 10(5), e1001453. doi:[10.1371/journal.pmed.1001453](https://doi.org/10.1371/journal.pmed.1001453)
52. Schlicker A, Beran G, Chresta CM, McWalter G, Pritchard A, Weston S, Runswick S, Davenport S, Heathcote K, Castro DA, Orphanides G, French T, Wessels LF (2012) Subtypes of primary colorectal tumors correlate with response to targeted treatment in colorectal cell lines. *BMC Med Genet* 5(1):1–15. doi:[10.1186/1755-8794-5-66](https://doi.org/10.1186/1755-8794-5-66)
53. Perez Villamil B, Romera Lopez A, Hernandez Prieto S, Lopez Campos G, Calles A, Lopez Asenjo JA, Sanz Ortega J, Fernandez Perez C, Sastre J, Alfonso R, Caldes T, Martin Sanchez F, Diaz Rubio E (2012) Colon cancer molecular subtypes identified by expression profiling and associated to stroma, mucinous type and different clinical behavior. *BMC Cancer* 12(1):1–13. doi:[10.1186/1471-2407-12-260](https://doi.org/10.1186/1471-2407-12-260)
54. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, Skora AD, Luber BS, Azad NS, Laheru D (2015) PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med* 372(26):2509–2520
55. Xiao Y, Freeman GJ (2015) The microsatellite instable subset of colorectal cancer is a particularly good candidate for checkpoint blockade immunotherapy. *Cancer Discov* 5(1):16–18