UNIVERSITY OF CATANIA

PhD

in

Translational Biomedicine

XXX cycle

GENOME INSTABILITY AND GENE EXPRESSION PROFILE IN COLORECTAL CANCER

Giacomo Cinnirella

Coordinator of PhD:

Prof. Lorenzo Malatino

Tutor:

Prof. Daniele Filippo Condorelli

INDEX

1. ABSTRACT

Four main forms of genomic instability have been described in colorectal cancer (CRC): microsatellite instability (MSI), chromosomal instability (CIN), epigenome abnormalities (CIMP), and hypermutation-ultramutation.

The present thesis was focused on the two better characterized forms of genome instability: MSI and CIN. The aim of the present work was to set up a new classification based on MSI and CIN and to analyze gene expression profiles of the newly defined groups.

Microsatellite testing classifies tumor samples in two fundamental types: microsatellite-instable (MSI) and microsatellite-stable (MSS) tumors. This classification is well-established according to routine methodology and widely accepted guidelines (Boland et al., 1998). In the present thesis a detailed mutational profile analysis was performed for DNA mismatch repair (MMR) genes, the catalytic subunit of proofreading polymerases (*POLE* and *POLD1*) and a selected group of 50 among oncogenes and tumor suppressors, for a more accurate molecular description of MSI tumors.

Classification based on chromosomal instability is much less standardized and affected by some technical difficulties. In the present thesis, the recent proposal about the use of somatic broad copy number abnormalities (BCNAs) (Barresi et al., 2017) was adopted in order to identify and sub-classify CRC tumors. According to the proposed methodology and to new criteria established in the present work, MSS tumors were subdivided into high-BCNA (HB) and low-BCNA (LB) tumors. A mutational profile analysis - with the same methodology used for MSI tumors – was also applied to the LB group.

A further step of the present work was to correlate the classification based on microsatellite status and on the number of BCNAs with gene expression profiles from cancer samples.

HB tumors showed upregulation of intestinal epithelial genes, such as *NOX1*, *AREG*, *EREG*.

LB tumors and MSI tumors shared a pattern of upregulation of *REG4*, *AGR2*, *SPP1*, *CD55, MUC5B*, although expression of such genes was higher for MSI samples. Upregulation of these genes had previously been described for mucinous tumors. Indeed, LB and MSI groups were enriched for mucin-producing tumors.

In conclusion, taking into account the number of BCNAs, along with MSI status and with the mutational profile, two groups of MSS samples can be distinguished: HB tumors, characterized by the expression of a subset of epithelial genes, some of them involved in EGF signaling; and LB tumors, enriched for mucin-

producing tumors, which resemble MSI tumors for what concerns upregulation of a subset of genes involved in the colon mucus barrier and other cell-precursor markers.

2. INTRODUCTION

By the name of “colorectal cancer” (CRC) we refer to a heterogeneous disease, whose complex genetic bases still need further investigation. CRC tumors can show different forms of genome and epigenome instability, which can coexist or predominate over one another. Over the years, many attempts to classify CRC into molecular subtypes have been made, taking into account some or all of the main features of CRC tumors: chromosomal instability (CIN), microsatellite status, CpG island methylator phenotype (CIMP) and single nucleotide mutation rates (hypermutation-ultramutation).

2.1 GENOME INSTABILITY IN CRC AND CLASSIFICATION SYSTEMS

The different forms of genome and epigenome instability in CRC and the main classifications systems will be described in this section.

2.1.1 MICROSATELLITE INSTABILITY

Colorectal tumors, but also other tumor types, can be classified according to their microsatellite status. Microsatellites are repetitive sequences made up of one up to six nucleotides, scattered along the genome, with a number of tandem repetitions ranging from a few units up to several thousands. (Heinimann, 2013). Microsatellites represent approximately 3% of the human genome. Repetitive sequences are challenging templates for DNA polymerases (Kunkel, 2004), and phenomena of strand misalignments can give rise to insertion/deletion loops (IDLs), which can lead to insertions or deletions (Jiricny, 2013). Normally, the so-called mismatch repair (MMR) system ensures surveillance and correction of IDLs as well as base-base mismatches which can escape DNA polymerase proofreading activity. The Bethesda Panel (Boland et al., 1998), made up of 5 microsatellite markers (two mononucleotide repeats - BAT25 and BAT26 - and three dinucleotide repeats - D5S346, D2S123 and D17S250), or larger panels, allow for a classification of CRC based on the number (or percentage) of unstable markers (Vilar and Gruber, 2010): microsatellite instability-high (MSI-H), if instability is present at ≥2 loci (or >30% of tested loci for larger panels); microsatellite instability-low (MSI-L), if instability is restricted to one locus (or to 10-30% of tested loci for larger panels); microsatellite-stable (MSS), if there are no unstable loci among those tested (or <10% unstable loci for larger panels).

2.1.2 MISMATCH REPAIR (MMR) SYSTEM

MMR system, which is devoted to correction of base-base mismatches and IDLs (Figure 1), is made up of the following components: MutSα heterodimer (MSH2-MSH6), MutSβ heterodimer (MSH2–MSH3), MutLα heterodimer (MLH1-PMS2), Exonuclease 1 (EXO1), the single-strand DNA binding protein Replication Protein A (RPA), the sliding clamp Proliferating Cell Nuclear Antigen (PCNA), the clamp loader Replication Factor C (RFC) and DNA Polymerase δ (Polδ). MutSα recognizes base-base mismatches and IDLs of 1-3 extrahelical nucleotides, whereas MutSβ recognizes IDLs of 2-10 extrahelical nucleotides (Modrich, 2016; Sharma et al., 2014). EXO1 is a 5'-3' exonuclease which removes a fragment of single stranded DNA containing the mismatch. Because of its directionality, EXO1 needs accessible 5' termini, which are available on the lagging strand (5' ends of Okazaki fragments), but (usually) not on the leading strand. Therefore, once the mismatch is detected, MutLα is recruited to create multiple nicks on the neosynthesized DNA strand containing the mismatch (both 5' and 3' with respect to the mismatch), thanks to the endonuclease activity of the PMS2 subunit. The strand specificity of MutLα, ensuring nicks to be placed on the neosynthesized strand, depends on its directional interaction with PCNA. Indeed, PCNA is loaded onto the DNA by RFC at available 3' termini with a unique orientation (Modrich, 2016). DNA 3' termini can be found not only on the lagging strand (3' ends of Okazaki fragments) but also on the leading strand (3' terminus of the growing neosynthesized strand) (Jiricny, 2013; Peña-Diaz and Jiricny, 2012). RFC can also load PCNA on DNA without any preexisting nick or gap, but the efficiency of this process is low (Peña-Diaz and Jiricny, 2012). EXO1 then removes a DNA fragment of ∼200 nucleotides containing the mismatch from the nicked strand (Modrich, 2016). RPA binds to the single-stranded DNA generated during the excision process and limits the extent of the excision (Genschel and Modrich, 2009). Finally, the gap is filled in by DNA Polymerase δ and ligated by DNA ligase I (Zhang et al., 2005).

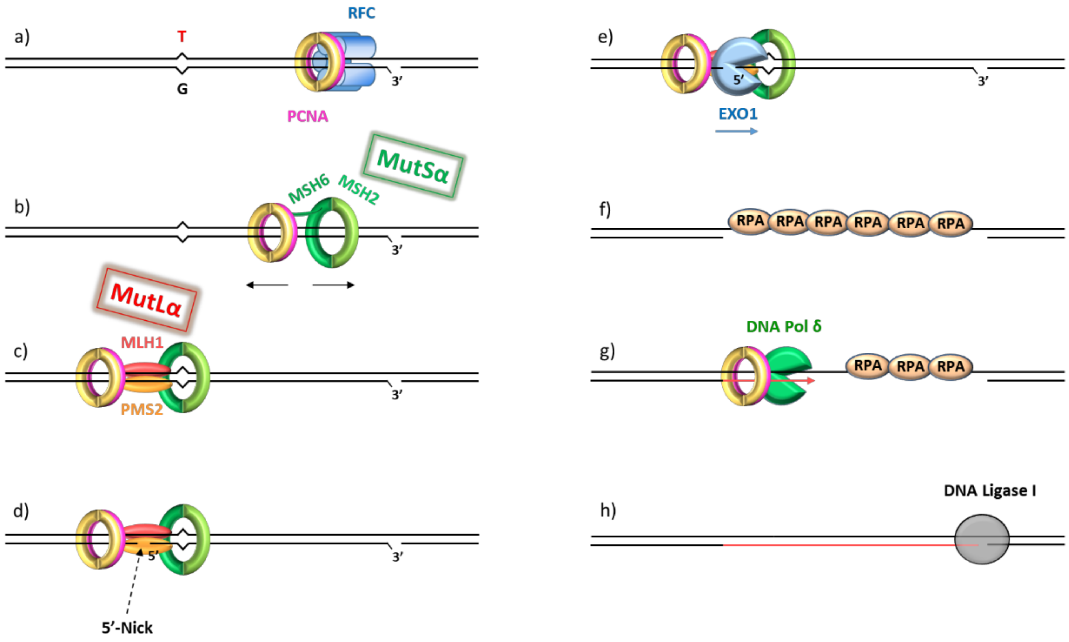


Figure 1: The MMR system in action. PCNA is loaded by RFC at 3' termini (a). MutSα scans the DNA for mismatches (b). Upon mismatch detection, MutLα is recruited and oriented through a directional interaction with PCNA (c), and creates multiple nicks on the DNA strand containing the mismatch (d). EXO1 removes a DNA fragment containing the mismatch (e) and RPA binds the single-stranded DNA (f). DNA Polymerase δ fills the gap (g). Finally, ligation of the nick is catalyzed by DNA ligase I (h). Redrawn and modified from: Modrich (2016).

2.1.3 LYNCH SYNDROME (LS)

Lynch syndrome is a cancer predisposition syndrome due to germline defects in one of the MMR genes. LS accounts for 1-3% of all CRCs (Aaltonen et al., 1998; Hampel et al., 2005; Peltomäki, 2016). As reviewed by Lynch et al. (2015), LS confers an increased risk of developing one or more (synchronous/metachronous) CRCs at an early age (<50 years of age), especially tumors of the proximal colon, with mucinous histology and lymphocytic infiltration. Such tumors do not develop in the context of a polyposis, so that LS was known in the past as "hereditary non-polyposis colorectal cancer" (HNPCC). LS also predisposes to tumors other than CRC, namely tumors of the endometrium, stomach, pancreas, ovaries, biliary tract, ureter or renal pelvis, brain (mainly glioblastomas), sebaceous gland adenomas/carcinomas and multiple keratoacanthomas (Muir–Torre syndrome), as well as carcinomas of the small intestine (Lynch et al., 2015; Peltomäki, 2016; Vasen et al., 1999). LS CRCs are MSI (Aaltonen et al., 1993) and show a near-diploid karyotype (Lengauer et al., 1998). MMR genes whose germline mutation can lead to LS are *MLH1* (accounting for 42% of LS cases), *MSH2* (33%), *MSH6* (18%) and *PMS2* (7.5%) (Plazzer et al., 2013). In LS, MMR genes seem to behave according to the "two hits" hypothesis: a germline mutation (first hit) in one of the MMR genes (point mutation or large rearrangement or –

rarely – constitutional epimutation) might cause haploinsufficiency and lead to formation of one or a few polyps, whereas a second hit affecting the MMR system (loss of the wild type allele, *MLH1* promoter hypermethylation) would be necessary for developing MSI (Lynch et al., 2015; Peltomäki, 2014).

2.1.4 CHROMOSOMAL INSTABILITY (CIN)

CIN refers to the rate at which cells gain or lose whole chromosomes or fractions of chromosomes along cell divisions (Geigl et al., 2008). Thereby, cancer cells show CIN if they are unable to keep their karyotype stable during cell proliferation, acquiring chromosomal abnormalities at a rate higher than normal. Measuring CIN is possible for cell cultures, or for blood tumors, but hardly achievable for solid tumors from surgery or biopsies (Geigl et al., 2008). Despite the term CIN being extensively used in the literature, what is often measured is not the rate of chromosomal changes but rather the amount of segmental aneuploidies. Thus, in recent papers, the measurement of copy number aberrations (CNAs) is being used for CRC molecular classification purposes instead of CIN. Since CIN leads to karyotype abnormalities, CNAs might be used as a surrogate – although imprecise – marker of CIN. A convenient approach when studying tumor CNAs could be focusing on broad copy number abnormalities (BCNAs), affecting a large percentage of a chromosomal arm or whole chromosomes (Barresi et al., 2017; Beroukhim et al., 2007). Identification of BCNAs is a more reliable approach compared to the study of smaller size CNAs – which can be termed focal copy number abnormalities (FCNAs) – the latter suffering from segmentation artefacts due to tumor heterogeneity and admixture of tumor and normal cells (Barresi et al., 2017).

2.1.5 CLASSIC CLASSIFICATION OF CRC

A simple classification of CRC takes into account the classic forms of genomic instability mentioned above (Figure 2):

- Tumors with chromosome instability (~85% of all CRCs): most of these tumors are of sporadic origin, and arise from the conventional adenoma-carcinoma sequence (Vogelstein et al., 1988). A small subset of CIN tumors (1% of all CRCs) arises in the context of the familiar adenomatous polyposis (FAP) syndrome, a cancer predisposition syndrome due to germline mutation in the *APC* gene (Groden et al., 1991; Kinzler and Vogelstein, 1996).
- Tumors with microsatellite instability (~15% of all CRCs): this class includes both Lynch syndrome tumors (2-4% of all CRCs), arising in the context of a cancer predisposition syndrome

due to germline defects in the mismatch repair genes, and sporadic MSI cases (10-13% of all CRCs). The latter are thought to arise through the serrated neoplasia pathway (Jass, 2007).

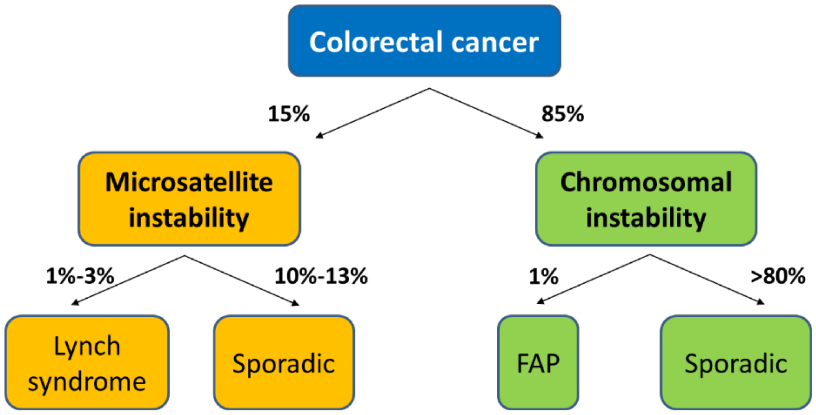


Figure 2: Classic classification of CRC.

2.1.6 CpG ISLAND METHYLATOR PHENOTYPE (CIMP)

CpG island methylator phenotype (CIMP) refers to an altered methylation pattern in the CpG islands within promoters of a large number of genes in tumor cells, not connected to the age-related increase of methylation levels (Toyota et al., 1999). Promoter methylation in such genes leads to transcriptional repression (Deaton and Bird, 2011). Methylation of *MLH1* in the context of CIMP is the dominant mechanism for the development of sporadic MSI CRCs (Weisenberger et al., 2006).

By a genome-scale analysis of aberrant DNA methylation, Hinoue et al. (2012) classified CRCs according to their CIMP status into three subgroups: CIMP-high (CIMP-H; ~22% of all CRCs), with a high frequency of cancer-specific promoter DNA hypermethylation; CIMP-low (CIMP-L; ~23% of the all CRCs), where hypermethylation only occurs at a subset of CIMP-H-associated markers; non-CIMP CRCs (~55% of all CRCs).

The same authors confirmed previous observations (Weisenberger et al., 2006) that CIMP-H strongly associates with *MLH1* promoter hypermethylation and *BRAF* V600E mutation. CIMP-H/MSI-H/*BRAF* mutant CRCs account for 9-12% of CRCs and are thought to arise from the serrated neoplasia pathway (Bettington et al., 2013; Jass, 2007).

CIMP and CIN appear to be inversely correlated in sporadic CRCs (Cheng et al., 2008; Goel et al., 2007; Hinoue et al., 2012). CIMP-positive samples are generally MSI and low-CIN, although a small a subset of CIMP-positive tumors can show a high degree of CIN (Cheng et al., 2008).

2.1.7 DNA POLYMERASES AND ULTRAMUTATION

Polymerase ε (Polε) and Polymerase δ (Polδ) are the two DNA polymerases responsible for replicating the nuclear genome (replicative polymerases). In the classic model (Figure 3), derived from yeast, human Polε replicates the leading strand (Shinbrot et al., 2014), whereas Polδ replicates the lagging strand, although Polδ might exert a role in replicating both the leading and the lagging strand (Johnson et al., 2015).

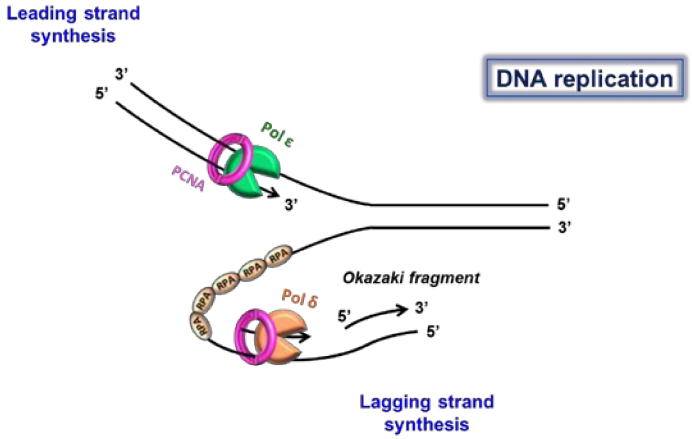


Figure 3: Current model of DNA polymerases at the replication fork.

Replicative polymerases have an error rate > 0, such errors being fundamental for evolution (Tomasetti et al., 2017). A percentage of replication errors might be corrected by the proofreading 3'–5' exonuclease activity of replicative polymerases. Polε and Polδ are hetero-tetramers, with the proofreading activity residing in the same protein subunit holding the catalytic activity, encoded by the *POLE* gene and the *POLD1* gene, respectively (Doublié and Zahn, 2014). *POLE* and *POLD1* exonuclease domains encompass amino acid residues 268–471 and 304–517, respectively (Briggs and Tomlinson, 2013). Without proofreading activity, the replication error rate would be of 1 misincorporated base in 10^5 nucleotides copied. With proofreading, the error rate is lowered to 1 in 10^7. Most errors escaping proofreading are repaired by the MMR system before the next round of replication, which takes the error rate as low as 1 in 10^9 nucleotides copied (Alberts et al., 2002; Preston et al., 2010).

Mutations in the proofreading domain of Polε or Polδ , either germline (Aoude et al., 2015; Bellido et al., 2016; Cancer Genome Atlas Network et al., 2012, 2013; Hansen et al., 2015; Palles et al., 2013; Rohlin et al., 2014) or somatic (Domingo et al., 2016; Erson-Omay et al., 2015; Stenzinger et al., 2014; Yoshida et

al., 2011) have been found in a wide spectrum of tumors, encompassing CRC, brain tumors, cutaneous melanoma, breast cancer, pancreatic cancer, ovarian cancer, tumors of the small intestine. These mutations are peculiar in that they cause very high rates of single nucleotide mutations (> 100 per 10^6) (Shinbrot et al 2014), such phenomenon being termed "ultramutation" in order to be distinguished from "hypermutation" (single nucleotide mutation rate between 10 and 100 per 10^6), the latter being typical of MSI samples. Germline exonuclease domain variants can be responsible for the so-called "polymerase proofreading-associated polyposis " (PPAP), defined as a dominant inherited susceptibility to multiple or large (>2 cm in diameter) colorectal adenomas and multiple or early-onset (usually <50 years of age) colorectal carcinomas (Briggs and Tomlinson, 2013; Palles et al., 2013). CRCs carrying *POLE* exonuclease domain mutations are usually MSS, although sometimes they might show microsatellite instability (Cancer Genome Atlas Network et al., 2012; Elsayed et al., 2015; Kim et al., 2013).

2.1.8 TCGA CLASSIFICATION OF CRC (2012)

In 2012, the Cancer Genome Atlas Network (TCGA) classified CRCs as follows:

- non-hypermutated CRCs (~84%), which were MSS and showed many somatic copy-number alterations (SCNAs), as well as mutations in well-known tumor suppressor genes (*APC, TP53, SMAD4*) and oncogenes (*KRAS, PIK3CA*) of the classic adenoma-carcinoma sequence (Vogelstein et al., 1988), along with mutations in *FBXW7*, *NRAS* and other genes.
- hypermutated cancers (~16%), here defined as having single nucleotide mutation rates >12 per 10^6. The majority of hypermutated cancers (~13% of the total cohort of this study) were MSI (most of them with *MLH1* methylation and CIMP) due to defective mismatch repair. The remaining hypermutated samples (~3% of the total cohort) were MSS, CIMP-negative and lacked *MLH1* methylation but generally had somatic mutations in one or more of the MMR genes and harbored mutations in *POLE* exonuclease domain. *POLE*-mutated tumors showed the highest point mutation rates of the whole study cohort.

As previously shown by others (Flohr et al., 1999; Jansen et al., 2016; Yoshida et al., 2011) and confirmed by TCGA data, mutations in MMR genes and *POLE/POLD1* genes tend to coexist in CRCs with high mutational rates.

2.1.9 CONSENSUS MOLECULAR SUBTYPES (CMS) OF CRC

In 2015, the Consensus Molecular Subtype (CMS) Consortium (Dienstmann et al., 2017; Guinney et al., 2015) described four CMS groups of CRC based on transcriptomic data from six studies (Table 1).

Table 1: Consensus molecular subtypes. Redrawn and modified from Guinney et al. (2015), figure 5.

CMS type	**CMS1**	**CMS2**	**CMS3**	**CMS4**
Description	**MSI Immune**	**Canonical**	**Metabolic**	**Mesenchymal**
Percentage of the cohort	14%	37%	13%	23%
Hypermutation	Hypermutated			
MSI status	Usually MSI		Mixed MSI status	
CIMP	CIMP-H		CIMP low	
SCNA	SCNA low	SCNA high	SCNA low	SCNA high
Point mutations in oncogenes	BRAF mutations		KRAS mutations	
Transcriptional profile	Immune infiltration and activation	WNT and MYC activation	Metabolic deregulation	Stromal infiltration TGFβ activation Angiogenesis

Of such groups, the CMS1 MSI-Immune subtype (14% of the total cohort) encompassed hypermutated, CIMP-H, usually MSI (76% of the CMS1 tumors) samples, with frequent $BRAF^{V600E}$ mutations (known to associate with MSI tumors and low prevalence of somatic copy number alterations (SCNAs). In line with deficient mismatch repair system, DNA repair proteins were overexpressed. Also, transcriptional immune activation and diffuse tumor immune infiltration were observed. On the other hand, CMS2, CMS3 and CMS4 subtypes showed higher numbers of SCNAs (the authors used SCNAs as a measure of CIN) with the CMS2 Canonical subtype (accounting 37% of the total cohort) harboring the highest frequency of SCNAs in oncogenes and tumor suppressor genes. CMS2 CRCs showed epithelial differentiation and marked activation of WNT and MYC downstream targets, along with upregulation of *EGFR*, *HER2*, *IGF2*, *IRS2* and amplification of the transcription factor *HNF4A*. CMS3 Metabolic subtype (13% of the total cohort) showed consistently lower number of SCNAs in comparison to the other CIN tumors. Moreover, 28% of CMS3 samples were hypermutated, 16% were MSI (with MSI samples falling among hypermutated tumors) and there was a high prevalence of the CIMP-L cluster (intermediate levels of gene hypermethylation). CMS3 tumors were also enriched for *KRAS* mutations and for several metabolic signatures (such as glutaminolysis and lipidogenesis), consistent with *KRAS* activation

remodeling of cell metabolism (Brunelli et al., 2014; Ying et al., 2012). CMS4 Mesenchymal subtype (23% of the total cohort) showed activation of genes involved in epithelial mesenchymal transition (EMT), transforming growth factor β (TGF β) signaling, extracellular matrix remodeling, complement inflammatory signaling, angiogenesis, and cancer stem cells signature. CMS4 Mesenchymal tumors transcriptional profile is thought to be strongly influenced by the remarkable stromal (mainly fibroblast) infiltration. Indeed, CMS4 samples showed reduced tumor purity compared to the other subtypes, that is higher admixture with non-cancer cells. Finally, samples with mixed features (13% of the total cohort) were observed: they might be due to intra-tumoral heterogeneity (potential mixtures of different CMS subtypes) or represent transition phenotypes among CMS subtypes.

2.2 INTESTINAL EPITHELIUM: DIFFERENTIATION PATHWAYS AND MUCUS BARRIER

In order to understand abnormalities in differentiation of CRC cells, it is necessary to provide up-to-date information on normal cell differentiation pathways in human colon and some insights into colon mucus barrier composition and functions.

2.2.1 INTESTINAL CRYPTS AND STEM CELLS

The inner layer of the small intestine presents with projections into the intestinal lumen known as villi, which extend the mucosal surface for optimizing the uptake of nutrients. Around each villus, there are invaginations known as crypts of Lieberkühn. New intestinal cells, which are needed to replace exfoliated ones, are formed at the crypt base and then climb up along the side of the crypt-villus axis, to substitute exfoliated cells. As reviewed by Clevers (2013), both villi and crypts of Lieberkühn contain absorptive enterocytes along with goblet cells secreting mucus, enteroendocrine cells secreting hormones, and tuft cells, which may act as sensors of luminal content. At the crypt base, two cell types are found: crypt base columnar (CBC) stem cells - which express the Leucine Rich Repeat Containing G Protein-Coupled Receptor 5 (LGR5) - and Paneth cells. $LGR5^+$ CBC stem cells can give rise to rapidly proliferating transit-amplifying (TA) cells, which in turn differentiate into enterocytes, goblet cells, enteroendocrine cells, tuft cells (Barker, 2014). $CD24^+$ Paneth cells not only secrete lysozyme and defensins, which help defending against bacteria, but are also essential to the intestinal stem cell niche, providing the $LGR5^+$ CBS stem cells with Wnt Family Member 3 (WNT3), epidermal growth factor (EGF), Transforming Growth Factor, Alpha (TGFα) (Sasaki et al., 2016; Sato et al., 2011). $LGR5^+$ CBC stem cells

also express Delta Like Canonical Notch Ligand DLL1 and DLL4 (Sasaki et al., 2016; Sato et al., 2011). The so called “position 4” or “+4” cells, located at the fourth position from the crypt base, are quiescent or slow cycling and might act as a reservoir capable of regenerating the LGR5^{+} CBC stem cells after tissue injury, acting as “facultative” stem cells (Barker, 2014; Beumer and Clevers, 2016; Buczacki et al., 2013; Clevers, 2013) (Figure 4).

Paneth cells are absent from colon crypts, where instead the so-called deep crypt secretory (DCS) cells - which are mucous-type cells - can be found intercalated with LGR5^{+} CBC stem cell at the crypt base (Altmann, 1983). It has been demonstrated in mice that DCS cells express Regenerating Family Member 4 (Reg4) and Proto-Oncogene C-Kit (cKit). Reg4^{+} cKit^{+} DCS cells express EGF and Notch ligands Dll1/Dll4 but do not produce Wnt ligand (Sasaki et al., 2016).

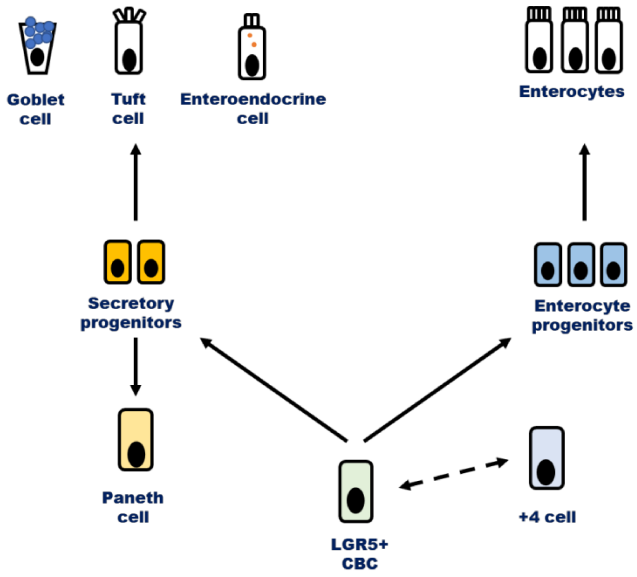


Figure 4: Stem cells and differentiation potential. Redrawn and modified from: Beumer and Clevers (2016), figure 2.

Among other intestinal stem cell markers, *OLFM4* is expressed by CBC stem cells in human small intestine and in human colon, and is a marker for LGR5^{+} stem cells (van der Flier et al., 2009). EPH Receptor B2 (EphB2), a tyrosine kinase transmembrane receptor which is a target of WNT signaling (van de Wetering et al., 2002) and Epithelial Cell Adhesion Molecule (EpCAM) are also intestinal stem cell markers (Dalerba et al., 2011).

2.2.2 INTESTINAL MUCUS BARRIER AND MUCINS

As reviewed by Birchenough et al. (2015), gastrointestinal epithelium is covered with mucus, which avoids direct contact between the luminal content, including bacteria, and the epithelial lining. A loosely attached single mucus layer is found in the small intestine, with large pores being penetrable to bacteria, but protection is provided by antimicrobial peptides (lysozyme, α-defensins) secreted by Paneth cells. In contrast, mucus in the colon is organized in two layers, an inner one tightly attached to the epithelium and impenetrable to bacteria, and an outer one hosting commensal bacteria, which have glycan-degrading enzymes and can use mucin glycans as a source of energy (Johansson et al., 2011).

The main component of intestinal mucus are mucins, which are glycoproteins, with *O*-glycans accounting for more than 50% of their total mass (Johansson and Hansson, 2016). Mucins can be distinguished into transmembrane mucins and gel-forming mucins (secreted mucins).

Some transmembrane mucins are constitutively expressed (MUC3, MUC4, MUC12, MUC13, MUC17, MUC20) whereas others (MUC1 and MUC16) are inducible in infection or cancer (Johansson and Hansson, 2016). Transmembrane mucins on the apical membrane of enterocytes contribute - with their oligosaccharide side chains - to the ∼10 nm-thick hydrophylic glycocalyx, also encompassing glycolipids of the cell membrane (Kufe, 2009). Members of the carcinoembryonic antigen cell adhesion molecules (CEACAM) family, such as the carcinoembryonic antigen CEA, CEACAM1, CEACAM6, CEACAM7, also contribute to the glycocalyx (Frängsmyr et al., 1999; Ou et al., 2009). The glycocalyx, together with the tight junctions among neighboring epithelial cells, behave as a diffusion barrier (Pelaseyed et al., 2014). The major glycocalyx transmembrane mucins are MUC3, MUC12, and MUC17 (Pelaseyed et al., 2014). MUC17 is preferentially expressed in the small intestine, although it is also present in the transverse colon (Johansson and Hansson, 2016).

The major gel-forming mucin in the intestine is MUC2, whose monomers multimerize to form a net-like structure (Nilsson et al., 2014). Other gel-forming mucins are MUC5AC and MUC6 (preferentially secreted by gastric surface epithelium and gastric glands, respectively) (Nordman et al., 2002), as well as MUC5B, which is expressed by a subset of colonic goblet cells at the crypt base (colocalizing with MUC2-containing goblet cells), but its expression decreases (until disappearing) when moving towards the upper part of the crypt (van Klinken et al., 1998; Larsson et al., 2011). Genes encoding the secreted mucins MUC2, MUC5AC, MUC5B and MUC6 are found in cluster on chromosome 11 (Kufe, 2009).

It has been demonstrated by Schutte et al. (2014) in the small intestine of mice that detachment of mucus from the epithelium requires cleavage of MUC2 N-terminus by meprin β, a membrane-bound zinc-dependent metalloendopeptidase which is released from the enterocyte apical membrane after microbial challenge and detaches MUC2 from its goblet cell anchor. The same authors showed that MUC2 release also requires its unfolding, in order to uncover meprin β cleavage sites. Once MUC2 is secreted by goblet cells, it undergoes expansion and organizes in stratified sheets (Ambort et al., 2012); both MUC2 unfolding and expansion require a pH increase and a Ca^{2+} decrease, which - in the small intestine - is accomplished through bicarbonate secretion by the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR), expressed on the apical membrane of adjacent enterocytes, allowing for the exit of Cl^- and HCO_3^- (Ambort et al., 2012; Frizzell and Hanrahan, 2012; Gustafsson et al., 2012; Schutte et al., 2014). CFTR expression in colon is lower than that in small intestine, urging further studies to better elucidate colonic mechanisms of mucus secretion (Crawford et al., 1991; Gustafsson et al., 2012; Strong et al., 1994). SLC26A3 is an electroneutral Cl^-/HCO_3^- exchanger mainly expressed in colon, especially in differentiated surface colonic epithelial cells (Höglund et al., 1996; Melvin et al., 1999), and is considered to be a marker of mature enterocytes (Dalerba et al., 2011). Coupled activity of SLC26A3 and the Na^+/H^+ exchanger SLC9A3 (NHE3) on apical membranes of epithelial cells results in electroneutral NaCl absorption (Alper et al., 2011) and apical HCO_3^- secretion (Jacob et al., 2002). SCL26A3 shows an activating interaction with CFTR, increasing CFTR channel overall open probability (Ko et al., 2004).

Goblet cells are the main cell type responsible for mucus production, and secrete not only MUC2 but also other proteins such as AGR2, FCGBP, TFF3, CLCA1, ZG16 (Faderl et al., 2015; Pelaseyed et al., 2014). Anterior gradient protein 2 homologue (AGR2) is a disulfide isomerase located in the endoplasmic reticulum (ER), where it helps protein folding by formation of disulfide bonds (Park et al., 2009). AGR2 is essential for MUC2 production by goblet cells, as well as for the biosynthesis of MUC5AC and MUC5B (Schroeder et al., 2012). AGR2 can also be secreted into the gastrointestinal mucus (Bergstrom et al., 2014). FCGBP (Fc globulin-binding protein), after an autocatalyzed cleavage reaction forming a reactive C-terminal end, can covalently attach to MUC2 and might be involved in mucus cross-linking (Johansson et al., 2009, 2011). Moreover, FCGBP covalently binds to Trefoil Factor 3 (TFF3) (Albert et al., 2010).

Intestinal mucus also contains proteins secreted by Paneth cells: lysozyme, α-defensins, Deleted In Malignant Brain Tumors 1 (DMBT1), as well as MUC2 (Pelaseyed et al., 2014). DMBT1 is a secreted glycoprotein belonging to the superfamily of scavenger receptor cysteine-rich (SRCR) proteins

(Mollenhauer et al., 1997). It is involved in mucosal immunity, since it can bind to secretory IgA (Ligtenberg et al., 2004) and act as a putative receptor for pathogens opsonized by the collectins SP-D and SP-A (Mollenhauer et al., 2000). It might also bind directly to bacteria regardless of opsonization (Mollenhauer et al., 2007). DMBT1 is expressed by epithelial cells of the crypt base and of the midcrypts throughout normal small intestine and colon (Renner et al., 2007), and has been implicated in epithelial differentiation (Kang and Reid, 2003). Interaction between DMBT1 and dimeric TFF3 has been described (Madsen et al., 2013), and both proteins might associate with gel-forming mucins to form a net which concentrates protective factors (Madsen et al., 2013). DMBT1 expression is up-regulated during inflammatory bowel diseases (Madsen et al., 2013; Renner et al., 2007).

Trefoil factors are peptides secreted by mucus-producing cells throughout the gastrointestinal tract: TFF1 and TFF2 are expressed in the stomach and in duodenal glands, whereas TFF3 is expressed by goblet cells of the small intestine and colon (Aihara et al., 2017; Playford, 1997). Trefoil factors are able to dimerize and exert some of their functions in the dimeric form (Muskett et al., 2003). As reviewed by Aihara et al. (2017), TFF peptides show protective effects on gastrointestinal mucosa: they contribute to mucus stabilization and stimulate cell migration to cover areas of damaged mucosa.

2.3 CRC PATHOLOGY

The great majority of CRCs (>90%) are adenocarcinomas, originating from epithelial cells of the intestinal mucosa. Histological tumor grading is given according to the entity of gland formation, which decreases from well-differentiated to poorly differentiated tumors (Fleming et al., 2012). About 10% of CRCs are instead mucinous adenocarcinomas, while ∼1% are signet ring cell adenocarcinomas (Nitsche et al., 2013). Mucinous adenocarcinomas are characterized by large glandular structures and extracellular mucin accounting for >50% of tumor volume (for lower percentages of extracellular mucin – as long as >10% - the term “adenocarcinoma with mucinous differentiation” or “with mucinous features” is used), whereas signet-ring cells adenocarcinomas show intracytoplasmic mucin pushing the nucleus to the cell periphery (Fleming et al., 2012). Sometimes, intracellular mucin can be present in mucinous adenocarcinomas as well, resulting in a signet-ring cell morphology; on the other hand, signet-ring adenocarcinomas can also produce some extracellular mucin (Sung et al., 2008). Prognosis is better for mucinous adenocarcinomas than for signet-ring cell adenocarcinomas (Sung et al., 2008). Among mucinous CRC tumors, those showing MSI have better prognosis than MSS ones (Nitsche et al., 2013).

3. AIM OF THE WORK

The present thesis was focused on the two better characterized forms of genome instability: MSI and CIN. The aim of the present work was to set up a new classification based on MSI and CIN and to analyze gene expression profile of the newly defined groups.

A first step of this thesis was to define criteria for classifying a cohort of CRC patients according to genome instability.

MSI classifies tumor samples in two fundamental types: microsatellite-instable (MSI) and microsatellite-stable (MSS) tumors. This classification is well-established according to routine methodology and widely accepted guidelines (Boland et al., 1998). Moreover, in the present thesis a detailed mutational profile analysis was performed for DNA mismatch repair (MMR) genes, the catalytic subunit of proofreading polymerases (POLE and POLD1) and a selected group of 50 among oncogenes and tumor suppressors, for a more accurate molecular description of MSI tumors.

On the contrary, classification based on chromosomal instability is much less standardized and affected by some technical difficulties. In the present thesis, the recent proposal about the use of somatic broad copy number abnormalities (BCNAs) (Barresi et al., 2017) was adopted in order to identify and sub-classify CRC tumors.

A second step of the present work was to correlate the classification based on microsatellite status and on the number of BCNAs with gene expression profiles from cancer samples.

4. MATERIALS AND METHODS

4.1 SAMPLES

The 48 samples analyzed in this thesis belonged to a cohort of 35 patients who underwent surgery for resection of primary invasive CRC at “Centro Clinico Diagnostico S.r.l. G.B. Morgagni” in Catania (Italy). All patients gave informed consent for this study, which was approved by the Ethics Committee of ASL3 of Catania (Italy). All specimens were frozen and stored at -80°C until DNA extraction. Samples had been previously tested for microsatellite instability with five markers belonging to the Bethesda panel (D2S123, D5S346, D17S250, BAT25 and BAT26) and one additional marker (BAT40) (Barresi et al., 2017). The sample cohort consisted of 7 MSI and 41 MSS tumor samples. If two biopsies were taken from the same tumor mass (at a distance of at least 1 cm from each other), the two samples were termed T1 and T2 (double-sampling pair). If a synchronous tumor was present and biopsied, this sample was termed T3. A biopsy of adjacent phenotypically normal colonic tissue (at a distance of 3-6 cm from the tumor) was taken for 31 patients (tumor/normal pairs). Copy number analysis by Affymetrix SNP 6.0 arrays was performed for all the tumor samples, and for their normal pair if available. Whole-Transcript Expression analysis was performed for 45 of the 48 CRC samples, and 25 normal tissue samples were also included. Targeted NGS sequencing was performed for 15 samples from 14 patients, including all the 7 MSI samples and 8 MSS samples. Details on the sample cohort can be found in Table 2.

Table 2: CRC sample cohort and normal controls.

Tumor samples					Normal colonic tissue		
Sample name	**Microsatellite status**	**HTA**	**MMR/Polymerases Panel Sequencing**	**CHS Panel Sequencing**	**Sample name**	**MMR/Polymerases and CHS Panel Sequencing**	**HTA**
PAA1_T1	MSS	Yes	-	-	PAA1_N	-	Yes
PAA3_T2	MSI	-	Yes	Yes	PAA3_N	Yes	-
P3_T1	MSI	-	Yes	Yes	P3_N	Yes	-
P11_T1	MSS	Yes	-	-	-	-	-
P13_T1	MSI	Yes	Yes	Yes	P13_N	Yes	Yes
P15_T1	MSS	Yes	-	-	-	-	-
P17_T1	MSS	Yes	-	-	-	-	-
P19_T1	MSS	Yes	Yes	Yes	P19_N	Yes	Yes
P23_T1	MSS	Yes	-	-	-	-	-
P29_T1	MSI	Yes	Yes	Yes	P29_N	Yes	-
P31_T1	MSS	Yes	-	-	P31_N	-	Yes
P37_T1	MSS	Yes	-	-	P37_N	-	Yes
P37_T2	MSS	Yes	-	-			
P37_T3	MSS	Yes	-	-			
P41_T1	MSI	Yes	Yes	Yes	P41_N	Yes	-
P43_T1	MSS	Yes	-	-	P43_N	-	Yes
P47_T1	MSS	Yes	Yes	Yes	P47_N	Yes	Yes
P49_T1	MSS	Yes	-	-	P49_N	-	Yes
P59_T2	MSI	Yes	Yes	Yes	P59_N	Yes	Yes
P63_T1	MSS	Yes	-	-	P63_N	-	Yes
P65_T1	MSS	Yes	Yes	Yes	P65_N	Yes	Yes
P65_T2	MSS	Yes	-	-			
P67_T1	MSS	Yes	-	-	P67_N	-	Yes
P67_T2	MSS	Yes	-	-			
P69_T1	MSS	Yes	-	-	P69_N	-	Yes
P69_T2	MSS	Yes	-	-			
P71_T1	MSS	Yes	-	-	P71_N	-	Yes
P71_T2	MSS	Yes	-	-			
P73_T1	MSS	Yes	Yes	Yes	P73_N	Yes	Yes
P75_T1	MSI	Yes	Yes	Yes	P75_N	Yes	Yes
P77_T1	MSS	Yes	-	-	P77_N	-	Yes
P77_T2	MSS	Yes	-	-			
P79_T1	MSS	Yes	-	-	P79_N	-	Yes
P79_T2	MSS	Yes	-	-			

P83_T1	MSS	Yes	-	-	P83_N	-	Yes
P83_T2	MSS	Yes	-	-			
P85_T1	MSS	Yes	Yes	Yes	P85_N	Yes	-
P85_T2	MSS	Yes	Yes	Yes			
P87_T1	MSS	Yes	-	-	P87_N	-	Yes
P89_T2	MSS	Yes	-	-	P89_N	-	Yes
P91_T1	MSS	Yes	-	-	P91_N	-	Yes
P93_T1	MSS	Yes	-	-	P93_N	-	Yes
P93_T2	MSS	Yes	-	-			
P95_T1	MSS	Yes	-	-	P95_N	-	Yes
P95_T2	MSS	Yes	-	-			
P97_T1	MSS	Yes	-	-	P97_N	Yes	Yes
P97_T2	MSS	Yes	Yes	Yes			
PCC3_T1	MSS	-	Yes	Yes	PCC3_N	Yes	-

4.2 GENOMIC DNA EXTRACTION AND RNA EXTRACTION

Genomic DNA was extracted from specimens by using the QIAamp DNA Mini Kit (QIAGEN, Venlo, Netherlands). Total RNA was extracted from specimens by using the RNeasy Mini (QIAGEN, Venlo, Netherlands). Extracted DNA and RNA were quantified on a NanoDrop ND-1000 spectrophotometer (Thermo Scientific, Waltham, MA, USA).

4.3 NEXT-GENERATION SEQUENCING (NGS) ON THE ION TORRENT PGM™ PLATFORM

4.3.1 TARGETED SEQUENCING

We used two panels of primer pairs for NGS targeted sequencing. The first panel, called MMR/Polymerases Panel, was custom-designed by using the Ion AmpliSeq™ Designer tool, and targeted the coding regions and the exon-intron boundaries (up to 50 bases from each exon en) of MMR genes and *POLE*/*POLD1* (Table 3, Table 4). Further details on this custom panels designed by our laboratory are reported in Supplementary Table 1.

Table 3: Features of the MMR/Polymerases panel.

Primer pairs	Number of pools	Average amplicon length	Panel Size	Input DNA
525 pairs	2 pools	157 bp	42870 bp	20 ng (10 ng DNA x 2 pools)

Table 4: Genes targeted by the MMR/Polymerases Panel.

#	Gene	RefSeq NM	Description (GeneCards)	Ensemble Transcript ID	Chromosome (Ensembl)
1	*MLH1*	NM_000249.3	MutL Homolog 1	ENST00000231790.6	3p22.2
2	*MLH3*	NM_001040108.1	MutL Homolog 3	ENST00000355774.6	14q24.3
3	*MSH2*	NM_000251.2	MutS Homolog 2	ENST00000233146.6	2p21
4	*MSH3*	NM_002439.4	MutS Homolog 3	ENST00000265081.6	5q14.1
5	*MSH6*	NM_000179.2	MutS Homolog 6	ENST00000234420.9	2p16.3
6	*PMS2*	NM_000535.5	PMS1 Homolog 2, Mismatch Repair System Component	ENST00000265849.11	7p22.1
7	*POLD1*	NM_001256849.1	DNA Polymerase Delta 1, Catalytic Subunit	ENST00000440232.6	19q13.3
8	*POLE*	NM_006231.3	DNA Polymerase Epsilon, Catalytic Subunit	ENST00000320574.9	12q24.33

Hotspot sites covering approximately 2,800 COSMIC mutations in 50 well-known tumor oncogenes and tumor suppressor genes were sequenced by using the commercially available Ion AmpliSeq™ Cancer Hotspot Panel v2. From here onwards, this panel will be referred to as the CHS Panel. Details on the CHS Panel genes can be found in Table 5 and Supplementary Table 2.

Table 5: Features of the Ion AmpliSeq™ Cancer Hotspot Panel v2.

Primer pairs	Number of pools	Average amplicon length	Panel Size	Input DNA
207 pairs	1 pool	154 bp	31727 bp	10 ng

The whole coding region of *MLH1* was targeted by the MMR/Polymerases Panel, while only *MLH1* COSM26085 (c.1151T>A, p.Val384Asp) hotspot was targeted by the CHS panel.

Barcoded libraries were prepared according to the Ion AmpliSeq™ DNA and RNA Library Preparation User Guide (Pub. No. MAN0006735) instructions. The Ion AmpliSeq™ Library Kit 2.0 (Cat. No. 4475345) was used. Libraries were quantified by using Qubit™ Fluorometer (TermoFisher Scientific, Waltham, MA, USA) and diluted to obtain a concentration of ~100 pM per primer pool. At this point, the two pools of the MMR/Polymerases were mixed with one another and then processed in a single tube (this step was unnecessary for the CHS Panel, which was made up of a single pool). An emulsion PCR was performed on the Ion PGM™ OneTouch 2 instrument in order to obtain a clonal amplification of amplicons on non-

magnetic beads called Ion PGM™ Template OT2 200 Ion Sphere™ Particles (ISPs). Enrichment of the template-positive ISPs was performed on the Ion OneTouch™ ES instrument in order to eliminate empty beads. Non-optical DNA sequencing was performed on the Ion PGM™ system using Ion 314™ Chips (1,3 million wells chips) or Ion 316™ Chips (6.3 million wells).

4.3.2 ANALYSIS OF SEQUENCING DATA

Variant prioritization was performed applying two filters:

1. Sequencing quality filter: coverage ≥ 30 reads; mutated allele coverage ≥ 3 reads; percentage of mutant allele reads ≥ 8% (or compatible with homozygous or heterozygous state for germline variants); individual evaluation of variants in homopolymer tracts.
2. Functional filter: location in coding sequences or splice sites (within ±2 nucleotides from exon-intron boundaries); non-synonymous variants; highest population allele frequency <0.01.

All filtered variants in the sequenced targets (MMR genes, *POLE* and *POLD1* genes, CHS genes), whether missense, nonsense, frameshift or splice site variants, were searched for functional information in the following databases:

- COSMIC database at http://cancer.sanger.ac.uk/cosmic (Forbes et al., 2017). COSMIC variants are reported in the database with a prediction of their functional consequence by the Functional Analysis through Hidden Markov Models (FATHMM) algorithm. Such prediction has been considered when analyzing data of the present work.
- ClinVar database at https://www.ncbi.nlm.nih.gov/clinvar/ (Landrum et al., 2016)

Moreover, filtered variants in MMR genes were searched for their clinical significance in the database of the International Society for Gastrointestinal Hereditary Tumours (InSiGHT), Variant Interpretation Committee (VIC), at www.insight-database.org/classifications (Thompson et al., 2013).

4.4 GENOME-WIDE DNA COPY NUMBER AND SNP GENOTYPING ANALYSIS

High-resolution genome-wide DNA copy number and SNP genotyping analysis was performed on Affymetrix SNP 6.0 array using 500 ng of DNA according to manufacturer's instructions (Affymetrix, Inc.,

Santa Clara, CA, USA). Array scanning and data analysis were performed by using the Affymetrix® "GeneChip Command Console" (AGCC) and the "Genotyping Console™" (GTC) version 3.0.1 software (Barresi et al., 2017). Broad copy number abnormalities (BCNAs), defined as gains or losses involving more than 25% of a chromosomal arm or numerical aberrations of whole chromosomes, were identified by using the BroCyA bioinformatics tool (Barresi et al., 2017). Along with copy number (gains and losses), SNP arrays also detected broad copy-neutral loss-of-heterozygosity (CN-LOH) regions, that is genomic segments >3Mb with unvaried copy number but lacking heterozygosity for the assayed polymorphic markers.

4.5 WHOLE-TRANSCRIPT EXPRESSION ANALYSIS

Whole-Transcript Expression Analysis was performed from 100 ng of total RNA by amplification and target hybridization to the Gene-Chip Human Transcriptome Array 2.0, following the Manufacturer's instructions (Cat. No. 902310, Cat. No. 900720; Affymetrix, Inc., Santa Clara, CA, USA). Array scanning and data analysis had been performed by using Affymetrix® Expression Console™ software version 1.4 (Affymetrix, Inc., Santa Clara, CA, USA) and the Affymetrix® Transcriptome Analysis Console (TAC) Software (Affymetrix, Inc., Santa Clara, CA, USA).

Analysis of genes with no gene symbol assigned by Affymetrix, genes described by Affymetrix as "uncharacterized LOC", genes encoding small nucleolar RNAs (snoRNAs), small Cajal body-specific RNAs (scaRNAs), snRNAs (small nuclear RNAs), RNA 5S ribosomal genes and pseudogenes, and histone cluster genes was out of the purposes of the present thesis, as well as the analysis of transcripts on chr1_gl000191_random, chr4_ctg9_hap1, chr4_gl000193_random, chr6_apd_hap1, chr6_cox_hap2, chr6_dbb_hap3, chr6_mann_hap4, chr6_mcf_hap5, chr6_qbl_hap6, chr6_ssto_hap7, chr7_gl000195_random, chr17_ctg5_hap1, chr17_gl000204_random, chr19_gl000209_random, chrUn_gl000211, chrUn_gl000212, chrUn_gl000218, chrUn_gl000219, chrUn_gl000220, chrUn_gl000222, chrUn_gl000223, chrUn_gl000228. Thereby, such entries were ruled out from the analysis.

5. RESULTS

5.1 PATHOLOGY AND MOLECULAR FEATURES OF THE PATIENTS' COHORT

Pathology features, microsatellite status and information on HTA and targeted sequencing for each sample of the patient's cohort can be found in Table 6.

Table 6: Pathology, microsatellite instability status and information on HTA and Targeted Sequencing Availability. *AJCC*, American Joint Committee on Cancer.

Sample name	Anatomical site	Histology	Gender	Age at surgery	Tumor stage (AJCC)	HTA	Microsatellite Status	MMR/Polymerases Panel Sequencing	CHS Panel Sequencing
PAA1_T1	sigmoid colon	adenocarcinoma	male	71	stage 4	Yes	MSS	-	-
PAA3_T2	right colon	mucinous adenocarcinoma	male	66	stage 2	-	MSI	Yes	Yes
P3_T1	right colon	mucinous adenocarcinoma	male	41	stage 2	-	MSI	Yes	Yes
P11_T1	left colon	adenocarcinoma	female	75	stage 3	Yes	MSS	-	-
P13_T1	right colon	mucinous adenocarcinoma	male	46	stage 3	Yes	MSI	Yes	Yes
P15_T1	right colon	adenocarcinoma	male	62	stage 3	Yes	MSS	-	-
P17_T1	sigmoid colon	adenocarcinoma	male	66	stage 2	Yes	MSS	-	-
P19_T1	right colon	adenocarcinoma	male	38	stage 2	Yes	MSS	Yes	Yes
P23_T1	rectum	adenocarcinoma	female	74	stage 4	Yes	MSS	-	-
P29_T1	left colon	mucinous adenocarcinoma	male	42	stage 3	Yes	MSI	Yes	Yes
P31_T1	right colon	adenocarcinoma	male	75	stage 4	Yes	MSS	-	-
P37_T1	left colon	mucinous adenocarcinoma	female	48	stage 3	Yes	MSS	-	-
P37_T2	left colon	mucinous adenocarcinoma	female	48	stage 3	Yes	MSS	-	-
P37_T3	left colon	mucinous adenocarcinoma	female	48	stage 3	Yes	MSS	-	-
P41_T1	right colon	mucinous adenocarcinoma	male	81	stage 3	Yes	MSI	Yes	Yes
P43_T1	right colon	adenocarcinoma	male	32	stage 4	Yes	MSS	-	-
P47_T1	left colon	mucinous adenocarcinoma	male	77	stage 2	Yes	MSS	Yes	Yes
P49_T1	left colon	mucinous adenocarcinoma	male	84	stage 2	Yes	MSS	-	-
P59_T2	right colon	adenocarcinoma	male	90	stage 3	Yes	MSI	Yes	Yes
P63_T1	sigmoid colon	adenocarcinoma	female	88	stage 2	Yes	MSS	-	-
P65_T1	rectum	mucinous adenocarcinoma	male	71	stage 2	Yes	MSS	Yes	Yes
P65_T2	rectum	mucinous adenocarcinoma	male	71	stage 2	Yes	MSS	-	-
P67_T1	right colon	mucinous adenocarcinoma	female	86	stage 2	Yes	MSS	-	-

P67_T2	right colon	mucinous adenocarcinoma	female	86	stage 2	Yes	MSS	-	-
P69_T1	left colon	adenocarcinoma	female	50	stage 2	Yes	MSS	-	-
P69_T2	left colon	adenocarcinoma	female	50	stage 2	Yes	MSS	-	-
P71_T1	right colon	adenocarcinoma	male	71	stage 2	Yes	MSS	-	-
P71_T2	right colon	adenocarcinoma	male	71	stage 2	Yes	MSS	-	-
P73_T1	right colon	signet-ring cell adenocarcinoma	male	69	stage 3	Yes	MSS	Yes	Yes
P75_T1	right colon	adenocarcinoma	female	68	stage 2	Yes	MSI	Yes	Yes
P77_T1	right colon	mucinous adenocarcinoma	male	79	stage 3	Yes	MSS	-	-
P77_T2	right colon	mucinous adenocarcinoma	male	79	stage 3	Yes	MSS	-	-
P79_T1	right colon	adenocarcinoma	female	84	stage 3	Yes	MSS	-	-
P79_T2	right colon	adenocarcinoma	female	84	stage 3	Yes	MSS	-	-
P83_T1	right colon	mucinous adenocarcinoma	female	76	stage 3	Yes	MSS	-	-
P83_T2	right colon	mucinous adenocarcinoma	female	76	stage 3	Yes	MSS	-	-
P85_T1	right colon	mucinous adenocarcinoma	female	81	stage 3	Yes	MSS	Yes	Yes
P85_T2	right colon	mucinous adenocarcinoma	female	81	stage 3	Yes	MSS	Yes	Yes
P87_T1	rectum	mucinous adenocarcinoma	female	79	stage 2	Yes	MSS	-	-
P89_T2	right colon	adenocarcinoma	female	69	stage 2	Yes	MSS	-	-
P91_T1	right colon	signet-ring cell adenocarcinoma	male	73	stage 3	Yes	MSS	-	-
P93_T1	rectum	adenocarcinoma	female	50	stage 3	Yes	MSS	-	-
P93_T2	rectum	adenocarcinoma	female	50	stage 3	Yes	MSS	-	-
P95_T1	sigmoid colon	adenocarcinoma	male	73	stage 3	Yes	MSS	-	-
P95_T2	sigmoid colon	adenocarcinoma	male	73	stage 3	Yes	MSS	-	-
P97_T1	sigmoid colon	adenocarcinoma	male	73	stage 2	Yes	MSS	-	-
P97_T2	sigmoid colon	adenocarcinoma	male	73	stage 2	Yes	MSS	Yes	Yes
PCC3_T1	sigmoid colon	adenocarcinoma	male	76	stage 4	-	MSS	Yes	Yes

5.2 RESULTS FROM SEQUENCING

5.2.1 GERMLINE VARIANTS FROM THE MMR/POLYMERASES PANEL

The germline variants found with the MMR/Polymerases Panel are reported in Table 7 and Supplementary Table 3.

Table 7: Germline mutations of MMR/Polymerase Panel genes in CRC samples.

Gene	cDNA Change	Protein Change		Sample	Microsatellite Status
MLH1	c.546-2A>G	p.Arg182Serfs*6	heterozygous	P3_T1	MSI
MSH2	c.2536C>T	p.Gln846*	heterozygous	P29_T1	MSI
MSH6	c.4001+1_4001+2insTAAC	-	heterozygous	P75_T1	MSI
			heterozygous	P47_T1	MSS
PMS2	c.1789A>T	p.Thr597Ser	heterozygous	P65_T1	MSS
POLE	c.2083T>A	p.Phe695Ile	heterozygous	P29_T1	MSI
	c.1007A>G	p.Asn336Ser	heterozygous	P19_T1	MSS
POLD1	c.2017G>A	p.Glu673Lys	heterozygous	P3_T1	MSI
	c.2275G>A	p.Val759Ile	heterozygous	P29_T1	MSI

No germline variants with a population allele frequency < 1% were identified in the CHS panel genes.

5.2.2 SOMATIC VARIANTS FROM THE MMR/POLYMERASES PANEL

The following somatic variants were found in MSI samples (Table 8).

Table 8: Somatic mutations of MMR/Polymerase Panel genes in CRC samples.

Gene	cDNA Change	Protein Change	Sample
MLH1	c.1276C>T	p.Gln426*	PAA3_T2
	c.1459C>T	p.Arg487*	P13_T1
MLH3	c.4011G>T	p.Glu1337Asp	P41_T1
MSH2	c.2327_2328insT	p.Cys778Leufs*9	P59_T2
MSH3	c.554A>T	p.Asp185Val	P13_T1
MSH6	c.1082G>A	p.Arg361His	P29_T1
			P13_T1
	c.3163G>A	p.Ala1055Thr	P3_T1
POLE	c.1630G>A	p.Val544Met	P3_T1
	c.2132C>T	p.Ser711Phe	P41_T1
	c.2780A>G	p.Asn927Ser	P3_T1
	c.3455A>G	p.Gln1152Arg	P41_T1
	c.4570C>T	p.Pro1524Ser	P3_T1
	c.4603G>A	p.Gly1535Ser	P59_T2
POLD1	c.1732G>A	p.Gly578Ser	P13_T1

Among MSS sample, only one patient (PCC3_T1) showed two variants: *MLH3* c.55A>C (p.Ile19Leu) and *POLE* c.4652A>C (p.His1551Pro). In the remaining MSS samples (P65_T1, P73_T1, P85_T1, P85_T2, P19_T1, P47_T1, P97_T2) no variants in the MMR/polymerases genes were detected.

The number of somatic variants per patient is reported in Figure 5, while further details on variants can be found in Supplementary Table 4. As shown in Figure 5, MSI CRC samples were significantly enriched (two-tailed t-test p-value = 0.0088, p-value <0.01) for somatic variants of the MMR/Polymerases panel genes (average number of variants = 2.14, standard deviation (SD) = 1.57) compared to MSS CRCs (average number of variants = 0.25, SD = 0.71). However, all the mutations in *POLE* and *POLD1* (Table 8, Supplementary Table 2) were located outside the exonuclease domain (which encompasses amino acids 268-471 for *POLE* and 304-517 for *POLD1*). Most of the variants in MMR genes were missense changes, followed by nonsense and frameshift variants.

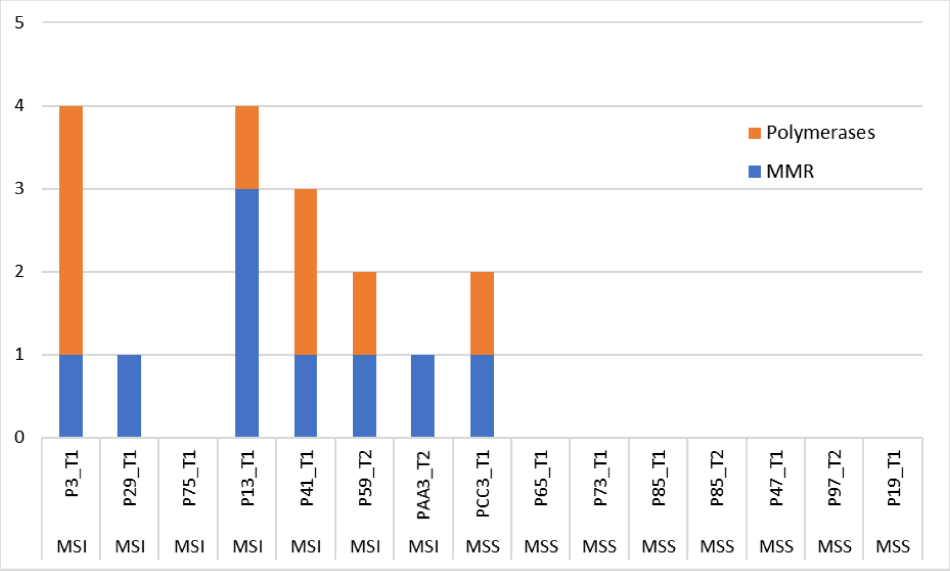


Figure 5: Distribution of MMR/Polymerases mutations across CRC samples. The number of variants is reported on the y axis.

5.2.3 SOMATIC MUTATIONS FROM THE CANCER HOTSPOT PANEL

Several mutations in classic oncogenes and tumor suppressors were identified in both MSI and MSS samples (Table 9). Details can be found in Supplementary Table 5.

Table 9: Somatic mutations of CHS Panel genes in CRC samples.

Gene	cDNA Change	Protein Change	Sample	Microsatellite Status
KRAS	c.35G>A	p.Gly12Asp	P29_T1	MSI
			PAA3_T2	MSI
	c.35G>T	p.Gly12Val	P75_T1	MSI
			P19_T1	MSS
	c.40G>A	p.Val14Ile	P41_T1	MSI
	c.175G>A	p.Ala59Thr	P29_T1	MSI
TP53	c.473G>A	p.Arg158His	P97_T2	MSS
	c.524G>A	p.Arg175His	P13_T1	MSI
	c.631_632delAC	p.Thr211Phefs*4	P75_T1	MSI
	c.746G>T	p.Arg249Met	P75_T1	MSI
APC	c.2626C>T	p.Arg876*	P97_T2	MSS
	c.4104_4105insG	p.Pro1369Alafs*6	P47_T1	MSS
BRAF	c.1790T>A	p.Leu597Gln	P47_T1	MSS
	c.1798_1799insAGA	p.Val600delinsGluMet	P41_T1	MSI
PIK3CA	c.1633G>A	p.Glu545Lys	P29_T1	MSI
	c.3073A>G	p.Thr1025Ala	P59_T2	MSI
ALK	c.3599C>T	p.Ala1200Val	P47_T1	MSS
CDH1	c.1115C>A	p.Pro372His	P3_T1	MSI
FBXW7	c.1177C>T	p.Arg393*	P59_T2	MSI
SMAD4	c.1009G>A	p.Glu337Lys	P47_T2	MSS
HNF1A	c.833G>A	p.Arg278Gln	P85_T1	MSS
			P85_T2	MSS
PTEN	c.343G>T	p.Asp115Tyr	P85_T1	MSS
			P85_T2	MSS
STK11	c.1069G>A	p.Glu357Lys	P85_T1	MSS
			P85_T2	MSS
RET	c.2767C>T	p.Leu923Phe	P85_T1	MSS

Distribution of CHS mutations across samples can be seen in Figure 6.

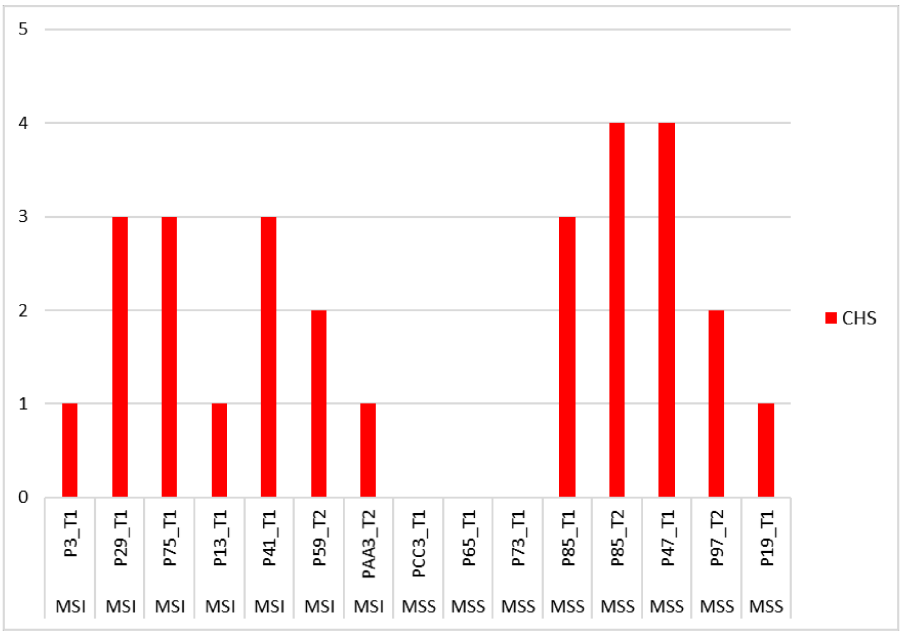


Figure 6: Distribution of CHS mutations across CRC samples. Number of variants is reported on the y axis.

In contrast to what observed in the case of MMR /Polymerases genes, the number of somatic mutations of the CHS Panel genes in MSI tumors (average number of variants = 2.00, SD = 1.00) and in MSS tumors (average number of variants 1.75, SD = 1.75) did not differ significantly from one another (two-tailed t-test p-value = 0.75) (Figure 6).

5.2.4 DESCRIPTION OF GERMLINE AND SOMATIC SEQUENCE VARIANTS IN MMR/POLYMERASES AND CHS PANEL IN INDIVIDUAL PATIENTS

In this section a description of all variants found in each patient is reported, along with an analysis of available evidence from public database about their functional significance. Moreover, whenever possible, the chromosomal aberrations found by SNP array (broad gains, broad losses and CN-LOH, see Table 10) were taken into account.

5.2.4.1 TUMORS FROM LYNCH SYNDROME PATIENTS

Two germline heterozygous pathogenic mutations of MMR genes (InSiGHT Class 5) have been found in two different MSI tumors (Table 7), allowing respective patients to be diagnosed with Lynch syndrome. The first germline heterozygous change, *MLH1* c.546-2A>G, was found in **P3_T1 sample**. It leads to a splice site mutation *MLH1* p.Arg182Serfs*6 and is predicted to cause skipping of exon 7 (Dieumegard et

al., 2000; Planck et al., 1999). The other germline heterozygous mutation, *MSH2* c.2536C>T, causing p.Gln846* premature stop codon, was found in **P29_T1** sample. Indeed, the two diagnosed Lynch syndrome patients (P3_T1 and P29_T1) underwent surgery at a young age (41 years and 42 years, respectively), and their CRCs presented with mucinous aspect, consistent with Lynch syndrome typical features.

P3_T1 Lynch syndrome CRC, in addition to the above-mentioned germline mutation affecting *MLH1*, showed a somatic variant of uncertain significance (InSiGHT Class 3) in *MSH6* (c.3163G>A, p.Ala1055Thr, detected in 9 of the reads). Such variant is not recorded in COSMIC database, whereas a different change at the same codon, namely *MSH6* c.3164C>T (p.Ala1055Val), is reported as COSM3186029 (FATHMM prediction: none, score 0.00). Moreover, P3_T1 sample presented with variants in DNA replicative polymerases. At the germline level, heterozygous *POLD1* p.Glu673Lys was found. This variant lies in the *Pfam DNA_pol_B* conserved domain (PF00136) but has uncertain functional significance according to ClinVar. At the somatic level *POLE* p.Val544Met, *POLE* p.Asn927Ser and *POLE* p.Pro1524Ser were detected. *POLE* somatic variants all map outside the exonuclease domain and are not recorded in any of the consulted databases. Both somatic *MSH6* and *POLE* variants were detected in approximately 9-11% of the sequencing reads. Considering the CHS panel, only a missense change (p.Pro327His) in Cadherin 1 (*CDH1*) gene was detected, this variant being absent from the consulted databases.

The other Lynch syndrome sample, **P29_T1**, apart from its pathogenic germline *MSH2* mutation, also harbored two additional germline heterozygous variants. One, *POLD1* p.Val759Ile, has been classified as likely benign by most (but not all) of ClinVar submitters and has been previously described as a somatic variant in one colon adenocarcinoma sample (COSM3692994). The other one, *POLE* p.Phe695Ile, despite an overall population allele frequency of T = 0.01097 in Exac and of T = 0.008 in the 1000 Genomes Project Phase 3 (data not shown in Supplementary Table 3), was above the polymorphism threshold in several populations, including the ethnic group closest to our Sicilian population (Toscani in Italy, T = 0.023). Somatic *MSH6* p.Arg361His, classified as InSiGHT Class 3 variant of uncertain significance, was found in 45% of the reads. In the same patient, a terminal copy-neutral loss of heterozygosity (CN-LOH) at 2p has been observed (see Table 10 for a list of chromosomal aberrations); however, neither *MSH2* nor *MSH6* were included in this region of LOH. From the CHS panel, two mutations in *KRAS* were identified, p.Gly12Asp (51% of the reads) and p.Ala59Thr (43% of the reads). *KRAS* codon 12 is a classic mutational hotspot. Mutations in *KRAS* p.Ala59Thr have also been previously reported in gastrointestinal tumors (see COSM546; Lee et al., 2003; Yuen et al., 2002) and this codon is included among the hotspots for *KRAS* mutation tests (Weyn et al., 2017). *PIK3CA* p.Glu545Lys, detected

in 46% of the reads, is one of the most common *PIK3CA* hotspot activating mutations (Zhao and Vogt, 2008), with changes at codon 545 encompassing ~10% of *PIK3CA* mutations in CRC (Bader et al., 2005).

5.2.4.2 NON-LYNCH SYNDROME MSI TUMORS

P75_T1 carried germline heterozygous *MSH6* c.4001+1_4001+2insTAAC, which introduces an additional TAAC tetranucleotide within a pre-existing sequence of $(TAAC)_3$ located at the junction between exon 9 and intron 9-10. This variant, in the equivalent form of *MSH6* c.4001+12_4001+15dupACTA, is reported in ClinVar as Benign/Likely benign (Variation ID: 182672). No somatic mutations in the MMR/Polymerases Panel were found for sample P75_T1. The CHS Panel detected *KRAS* p.Gly12Val (46% of the reads). Since there is CN-LOH at 12p including *KRAS* gene, both *KRAS* alleles should be mutated. Moreover, two *TP53* mutations were found: p.Thr211Phefs*4 (27%), due to a deletion of a GT dinucleotide within a $(GT)_2$ tandem repetition at the genomic level; and p.Arg249Met (29%).

P13_T1 sample showed somatic *MLH1* p.Arg487* premature stop codon mutation (45% of the reads). Actually, a CN-LOH involving a region of the short arm of chromosome 3 including *MLH1* gene was identified in this sample, accounting for a double somatic hit in *MLH1*. The specific nucleotide change c.1459C>T detected in *MLH1*, causing p.Arg487* premature stop codon, is not reported in COSMIC but a different nucleotide change with the same effect, that is *MLH1* c.1458_1459delCCinsTT (p.Arg487*) is recorded with COSM4515964 (FATHMM: none, score 0.00). Moreover, P13_T1 sample harbored an InSiGHT Class 3 *MSH6* p.Arg361His variant (19% of reads) not found in any of the consulted databases; a *MSH3* p.Asp185Val variant (18% of the reads) for which no information is available from the consulted databases, whereas Exac and the 1000 Genomes Project Phase 3 both report two other changes at the same codon, *MSH3* c.553G>C (p.Asp185His, C = 0.000008509) and *MSH3* c.554A>G (p.Asp185Gly, G = 0.00004255, rs144012714). The last variant detected in P13_T1 sample was *POLD1* p.Gly578Ser (25% of the reads), located outside the exonuclease domain, with a very low population allele frequency in Exac database (A = 0.00002826) but for which no functional information is available. The percentage of sequencing reads for *MSH6* p.Arg361His, *MSH3* p.Asp185Val and *POLD1* p.Gly578Ser (18-25%) might be compatible with a single dominant clone, with the *MLH1* p.Arg487* mutation being present at the double of the frequency (45%) because of the CN-LOH event. Admixture with other tumor clones or with non-tumor tissue might explain why, despite the CN-LOH event, *MLH1* p.Arg487* mutation is limited to 45% of the reads. The CHS panel only detected the frequent hotspot mutation *TP53* p.Arg175His

(35.5%). Since there is a broad loss of a region of 17p including *TP53* gene, such mutation became hemizygous in the tumor (deletion-LOH).

P41_T1 sample showed somatic variants *MLH3* p.Glu1337Asp (10%), *POLE* p.Ser711Phe (9%) and *POLE* p.Gln1152Arg (9%). All these variants were identified in a percentage of approximately 9-10% of the sequencing reads. However, none of these variants has been found in the consulted databases. All the above-mentioned variants of *POLE* lie outside the exonuclease domain. Sample P41_T1 showed loss of 19w, where *POLD1* is located, but loss of *POLD1* seems to be restricted to a small percentage of tumor tissue cells (as estimated by $\log_2$-ratio intensity signal in SNP6 array analysis). As regards the CHS panel, three mutations were found. The first one was a missense change in *ATM*, p.Arg337His (17% of the reads; COSM21301). The second one was *BRAF* c.1798_1799insAGA, which consists in an in-frame insertion of three nucleotides at the genomic level, changing valine at codon 600 into glutamate and inserting a methionine residue between codon 600 and codon 601, leading to p.Val600delinsGluMet insertion/deletion (18% of the reads), not found in any of the consulted databases. COSMIC database reports instead c.1798delGinsTACA (p.Val600delinsTyrMet), COSM1159850 (FATHMM: none, score 0.00). The last mutation identified with the CHS panel was *KRAS* p.Val14Ile (14% of the reads), which is known to reduce KRAS GTPase activity, both intrinsic and GTPase activating protein (GAP)-stimulated, compared to the wild type protein, although not as much impaired as in the case of KRAS p.Gly12Asp (Schubbert et al., 2006). *KRAS* p.Val14Ile has been previously found in tumors of the large intestine, as well as in other tumor types (COSM12722, Y.-J. Lee et al., 2016).

P59_T2 showed somatic *MSH2* frameshift variant p.Cys778Leufs*9 (40% of the reads), due to insertion of a single base (c.2327_2328insT) in a $(T)_5$ homopolymer tract, which would therefore convert into a $(T)_6$ homopolymer with a shift in the reading frame; this variant has not been found in any of the consulted database and has not been assigned any InSiGHT class. Classification is given instead for *MSH2* c.2334C>A (p.Cys778*), which has been attributed InSiGHT Class 5, being also considered to be pathogenic by ClinVar (Variation ID: 90956). Moreover, *POLE* p.Gly1535Ser variant, located outside the exonuclease domain, was detected in 35% of the reads. The CHS panel allowed detection of *FBXW7* p.Arg393* (33% of the reads), and *PIK3CA* p.Thr2015Ala (40% of the reads).

PAA3_T2 sample showed InSiGHT Class 5 *MLH1* p.Gln426* premature stop codon (15% of the total reads)., and *KRAS* p.Gly12Asp (15% of the reads).

5.2.4.3 MSS TUMORS

PCC3_T1 showed a somatic *MLH3* p.Ile19Leu in 62% of the reads. This variant is not present in any of the consulted databases. Moreover, *POLE* p.His1551Pro (8%) was identified, not recorded in any of the databases as well. Exac database reports a different POLE variant at the same codon, namely *POLE* c.4651C>T (p.His1551Tyr) with an allele frequency of A = 0.000008274, also present in COSMIC database, COSM1605832 (FATHMM prediction: Pathogenic, score 0.93). No mutations were found with the CHS panel.

P65_T1 carried a germline heterozygous *PMS2* p.Thr597Ser, which is classified as an InSiGHT Class 1 (benign) variant. No somatic variants were detected.

P73_T1 carried no germline variants and no somatic variants in any of the sequenced genes.

MSS tumor samples **P85_T1** and **P85_T2** represent two biopsies taken at different sites of the same patient's surgical specimen. Interestingly, P85_T1 showed 0 BCNAs and 0 CN-LOH events, while P85_T2 showed 2 BCNAs and 0 CN-LOH events. No mutations were detected at the germline level in any of the two samples, as well as no mutations were found with the MMR/Polymerases panel. Considering results from the CHS panel, P85_T1 and P85_T2 shared the majority of sequence variants, although at different percentages of the sequenced reads: *HNF1A* p.Arg278Gln (22% vs 32%), *PTEN* p.Asp115Tyr (18% vs 25%), *STK11* p.Glu357Lys (23% vs 28%). As can be seen, percentage of mutated reads was higher for P85_T2. Morevover, P85_T2 also showed *RET* p.Leu923Phe (9%), not detected in P85_T1. Mutations of *HNF1A* have been described in a subset of hepatocellular adenomas with remarkable steatosis but neither significant inflammation nor cytologic abnormalities (Zucman-Rossi et al., 2006). Of note, *HNF1A* p.Arg278Gln has been found mutated at the germline level in a monoallelic fashion in at least one case of hepatocellular adenoma (Jeannot et al, 2010) and at the somatic level in one carcinoma of the small intestine (COSM1359416). No functional information is available for *PTEN* p.Asp115Tyr. Interestingly, COSMIC database reports a different change at *PTEN* codon 115, namely *PTEN* c.343G>A (p.Asp115Asn), COSM1166807 (FATHMM prediction: Pathogenic, score 0.94), while ClinVar classifies one more variant at the same codon, *PTEN* c.344A>G (p.Asp115Gly), as likely pathogenic (Variation ID: 224542). *STK11* p.Glu357Lys is classified by ClinVar as a variant of uncertain significance. *RET* p.Leu923Phe has been previously identified in colorectal cancer (Dallol et al., 2016) but no functional information is available, whereas three different variants at the same codon are reported in COSMIC: *RET* c.2767C>A (p.Leu923Ile; COSM6006463, FATHMM prediction: none, score 0.00), *RET* c.2767C>G (p.Leu923Val;

COSM48745, FATHMM prediction: Pathogenic, score 0.96) and *RET* c.2768T>A (p.Leu923His; COSM6006464, FATHMM prediction: none, score 0.00).

P47_T1 carried germline heterozygous *MSH6* c.4001+1_4001+2insTAAC (reported in ClinVar as Benign/Likely benign, Variation ID: 182672). No mutations were found with the MMR/Polymerases panel. As regards the CHS panel, four mutations were found. The first one, *ALK* p.Ala1200Val (19% of the reads), COSM317003 (FATHMM prediction: Pathogenic, score 0.98), not reported before for CRC. The second one, *APC* c.4104_4105insG (p.Pro1369Alafs*6), due to an insertion of a guanosine within a non-homopolymer sequence context, was found in 37% of the reads and is not recorded in any of the consulted databases. Third, *BRAF* p.Leu597Gln (21%), COSM1125 (FATHMM prediction: Pathogenic, score 0.99) previously found in metastatic melanoma (Amanuel et al., 2012) as well as in other tissues and, finally, *SMAD4* p.Glu337Lys (23%, COSM417827, FATHMM prediction: Pathogenic, score 0.99).

P97_T2 did not carry any germline variant, and no somatic variants were detected with the MMR/Polymerases panel. The CHS panel allowed detection of *APC* p.Arg876* (65%, COSM18852, frequently found in tumors of the large intestine) and *TP53* p.Arg158His (70%, COSM10690, present in a wide spectrum of malignancies). P97_T2 showed a broad loss on chromosome 5q including *APC*, (Figure 7) configuring a deletion LOH for *APC*, with the only copy left being mutated.

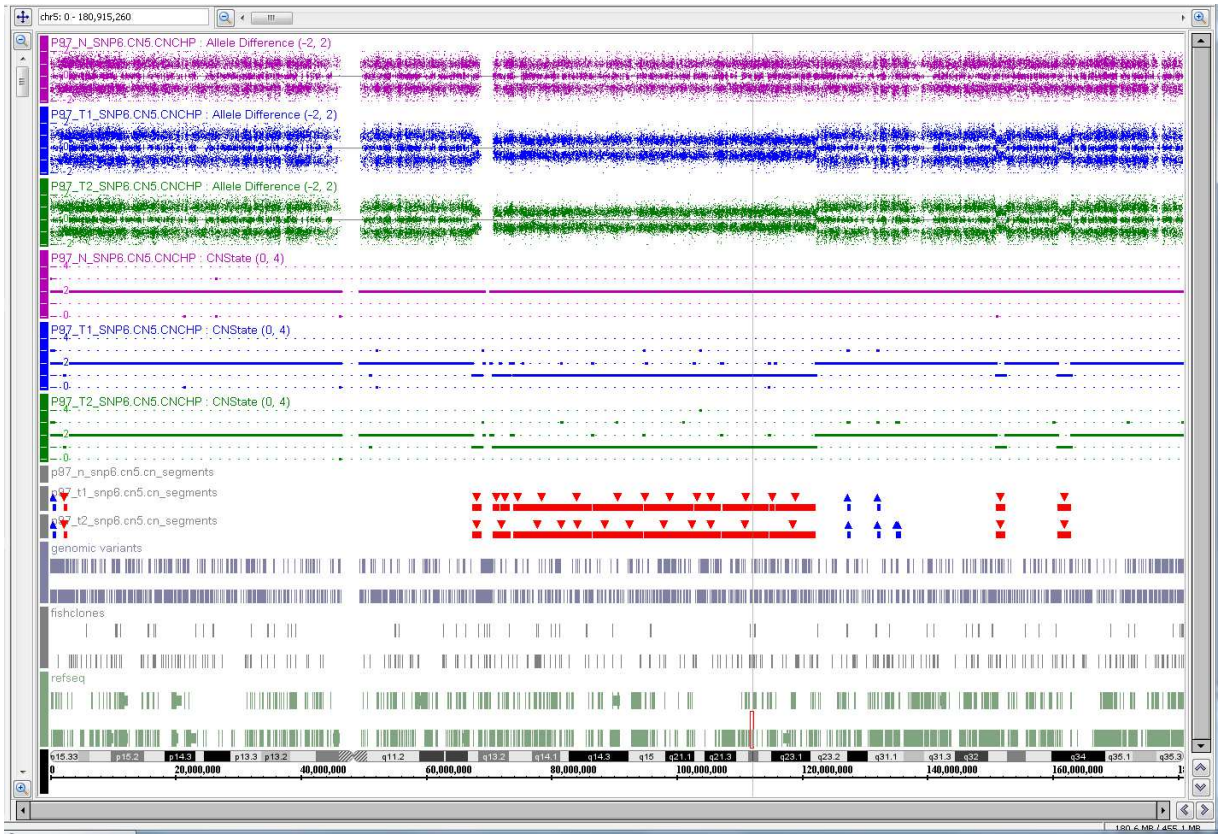


Figure 7: Broad loss on chromosome 5q, involving APC gene, in sample P97_T2.

P19_T1 showed the highest number of chromosomal abnormalities in our cohort. P19_T1 carried a germline heterozygous *POLE* p.Asn336Ser variant which, although located within the exonuclease domain, has been considered to be benign/likely benign in ClinVar database. No somatic variants have been detected with the MMR/Polymerases panel, whereas only *KRAS* p.Gly12Val (27.5%) has been detected with the CHS panel.

Table 10: List of BCNAs and CN-LOH for sequenced samples.

Sample	Total BCNAs	Gain	Loss	Total CN-LOH	CN-LOH
P3_T1	0	-	-	0	-
P29_T1	1	13q	-	2	2p, 9q
P75_T1	1	8w	-	3	11q, 12p, 22q
P13_T1	3	11q, 23p	17p	1	3p
P41_T1	3	8w, 13q	19w	1	2p,18q
P59_T2	5	7w, 8w, 9w, 13q, 20w		0	-
PAA3_T2	6	7w, 8w, 9w, 20w	22q, 24w	0	-
P65_T1	0	-	-	0	-

P73_T1	0	-	-	0	-
PCC3_T1	0	-	-	0	-
P85_T1	0	-	-	0	-
P85_T2	2	7w, 13q	-	0	-
P47_T1	5	8q, 13q, 23w	4q, 18w	2	5q, 10q
P97_T2	10	7w, 13q, 20q, 23w	4w, 5q, 8p, 12p, 18w, 20p	2	9
P19_T1	16	3q, 7w, 8q, 9w, 10p, 17q, 20q, 23p	1w, 4q, 6w, 8p, 17w, 18w, 19p, 20p	6	5w, , 10q, 11q, 14q,

X chromosome is numbered as chromosome 23; Y chromosome is numbered as chromosome 24.

5.3 GENOME INSTABILITY CLASSIFICATION OF CRC

MSS CRC samples were divided in two groups, according to their level of genome instability: high-BCNA (HB) and low-BCNA (LB) groups. To divide MSS samples into HB and LB, we used as a threshold the average value of BCNAs (2.71) found among MSI samples – which are usually regarded as near-diploid (Curtis et al., 2000; Hugen et al., 2015; Lengauer et al., 1998; Nakao et al., 2004) – plus one standard deviation (SD = 2.21), the sum being rounded up to 5 BCNAs. Samples in the MSI group and LB group had ≤5 BCNAs (Table 11 and Table 12), whereas sample in the HB group had >5 BCNAs (Table 13). The average BCNA number for LB samples was 1.73 (SD = 1.90); for the HB group, the average BCNA number was of 12.76 (SD = 4.36). The MSI group consisted of 5 tumors from 5 patients. The LB group included 11 tumors from 8 patients; the HB group included 29 tumors from 20 patients.

In one case, two samples from the same tumor presented with different BCNAs, and were assigned to different groups: P83_T1 (9 BCNAs) to the HB group, P83_T2 (0 BCNAs) to the LB group.

A comparison of the distribution of BCNAs in different chromosomes between HB and LB tumors is reported in Figure 8. Apart from 16p, whose gains have been found only in the LB group, all gains found in LB CRCs are also found in HB ones. This is also valid for losses: chromosomes affected in the LB group were the same chromosomes commonly affected in the HB group.

The same comparison regarding distribution of BCNAs per chromosome was made between HB and MSI tumors, as reported in Figure 9. Gains of chromosome 8w were more common among MSI samples.

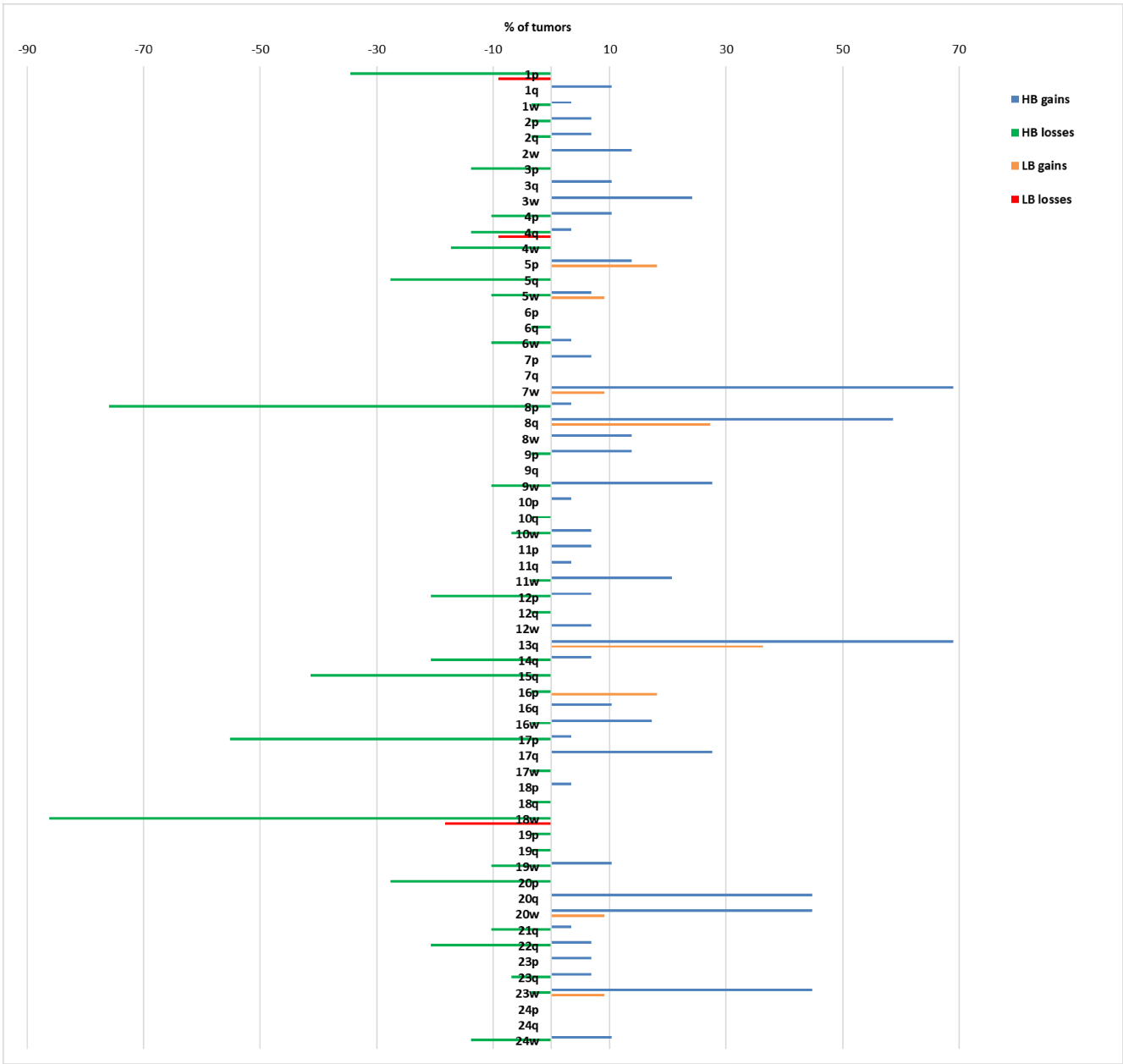


Figure 8: Distribution of BCNAs across HB and LB tumors. On the x axis, percentage of tumors harboring a specific BCNA.

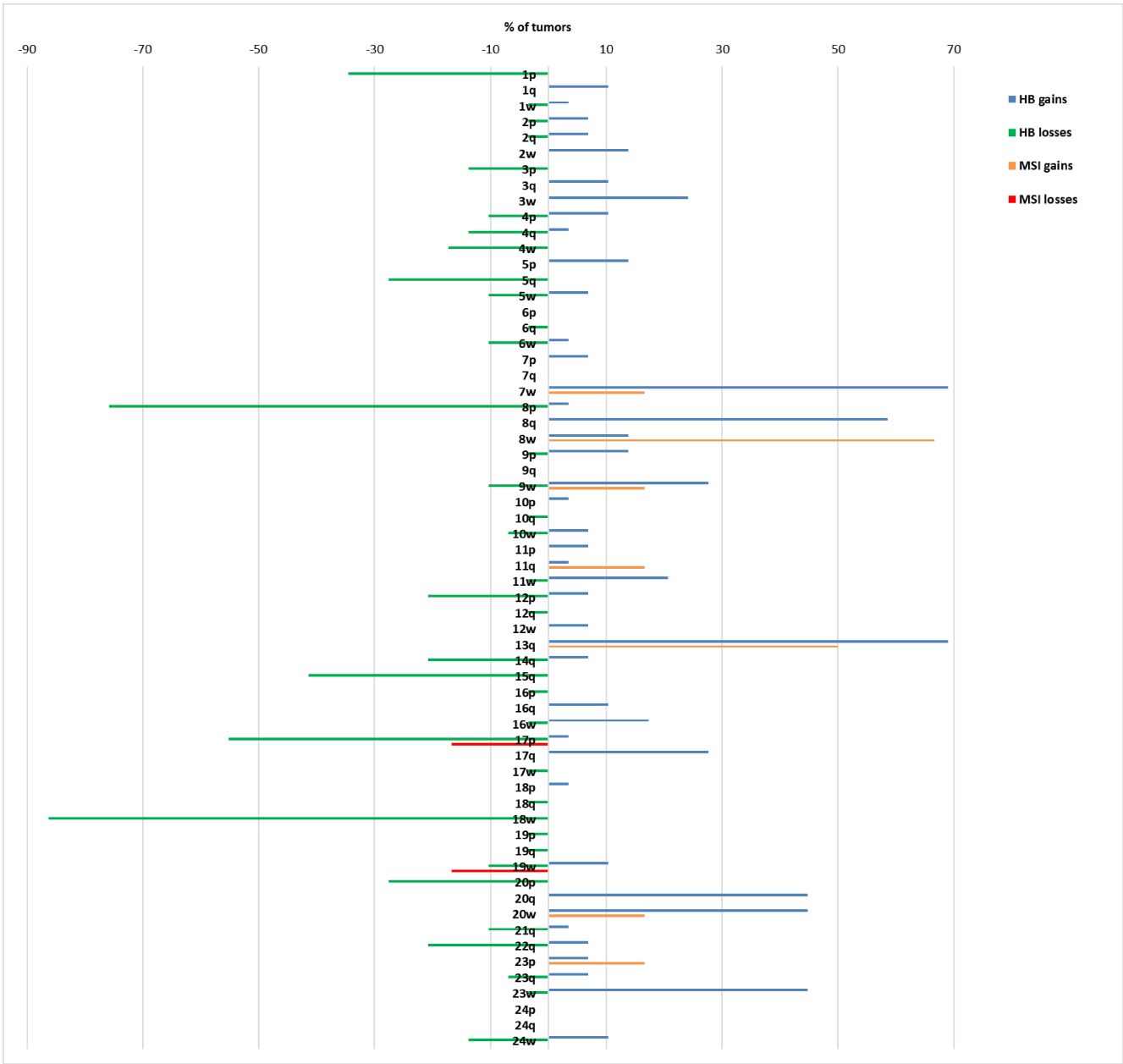


Figure 9: Distribution of BCNAs across HB and MSI tumors. On the x axis, percentage of tumors harboring a specific BCNA.

In conclusion, three sample groups were obtained as reported in Table 11 and Table 12 and

Table **13**.

Table 11: MSI CRC sample group for HTA analysis. Clinicopathological features, microsatellite instability status and molecular cytogenetics. *AJCC*, American Joint Committee on Cancer.

#	Sample name	Anatomical site	Histology	Gender	Age at surgery	Tumor stage (AJCC)	Total Broad Gains	Total Broad Losses	Total BCNAs	HTA	Microsatellite status	MMR/Polymerases Panel Sequencing	CHS Panel Sequencing
1	P13_T1	right colon	mucinous adenocarcinoma	male	46	stage 3	2	1	3	Yes	MSI	Yes	Yes
2	P29_T1	left colon	mucinous adenocarcinoma	male	42	stage 3	1	0	1	Yes	MSI	Yes	Yes
3	P41_T1	right colon	mucinous adenocarcinoma	male	81	stage 3	2	1	3	Yes	MSI	Yes	Yes
4	P59_T2	right colon	adenocarcinoma	male	90	stage 3	5	0	5	Yes	MSI	Yes	Yes
5	P75_T1	right colon	adenocarcinoma	female	68	stage 2	1	0	1	Yes	MSI	Yes	Yes

Table 12: LB CRC sample group for HTA analysis. Clinicopathological features, microsatellite instability status and molecular cytogenetics. *AJCC*, American Joint Committee on Cancer.

#	Sample name	Anatomical site	Histology	Gender	Age at surgery	Tumor stage (AJCC)	Total Broad Gains	Total Broad Losses	Total BCNAs	HTA	Microsatellite status	MMR/Polymerases Panel Sequencing	CHS Panel Sequencing
1	P23_T1	rectum	adenocarcinoma	female	74	stage 4	0	2	2	Yes	MSS	-	-
2	P47_T1	left colon	mucinous adenocarcinoma	male	77	stage 2	3	2	5	Yes	MSS	Yes	Yes
3	P49_T1	left colon	mucinous adenocarcinoma	male	84	stage 2	2	0	2	Yes	MSS	-	-
4	P65_T1	rectum	mucinous adenocarcinoma	male	71	stage 2	0	0	0	Yes	MSS	Yes	Yes
5	P65_T2	rectum	mucinous adenocarcinoma	male	71	stage 2	0	0	0	Yes	MSS	-	-
6	P67_T1	right colon	mucinous adenocarcinoma	female	86	stage 2	4	0	4	Yes	MSS	-	-
7	P67_T2	right colon	mucinous adenocarcinoma	female	86	stage 2	4	0	4	Yes	MSS	-	-
8	P73_T1	right colon	signet-ring cell adenocarcinoma	male	69	stage 3	0	0	0	Yes	MSS	Yes	Yes
9	P83_T2	right colon	mucinous adenocarcinoma	female	76	stage 3	0	0	0	Yes	MSS	-	-
10	P85_T1	right colon	mucinous adenocarcinoma	female	81	stage 3	0	0	0	Yes	MSS	Yes	Yes
11	P85_T2	right colon	mucinous adenocarcinoma	female	81	stage 3	2	0	2	Yes	MSS	Yes	Yes

Table 13: HB CRC sample group for HTA analysis. Clinicopathological features, microsatellite instability status and molecular cytogenetics. *AJCC*, American Joint Committee on Cancer.

#	Sample name	Anatomical site	Histology	Gender	Age at surgery	Tumor stage (AJCC)	Total Broad Gains	Total Broad Losses	Total BCNAs	HTA	Microsatellite status	MMR/Polymerases Panel Sequencing	CHS Panel Sequencing
1	PAA1_T1	sigmoid colon	adenocarcinoma	male	71	stage 4	9	8	17	Yes	MSS	-	-
2	P11_T1	left colon	adenocarcinoma	female	75	stage 3	10	8	18	Yes	MSS	-	-
3	P15_T1	right colon	adenocarcinoma	male	62	stage 3	9	8	17	Yes	MSS	-	-
4	P17_T1	sigmoid colon	adenocarcinoma	male	66	stage 2	8	11	19	Yes	MSS	-	-
5	P19_T1	right colon	adenocarcinoma	male	38	stage 2	8	8	16	Yes	MSS	Yes	Yes
6	P31_T1	right colon	adenocarcinoma	male	75	stage 4	7	4	11	Yes	MSS	-	-
7	P37_T1	left colon	mucinous adenocarcinoma	female	48	stage 3	4	6	10	Yes	MSS	-	-
8	P37_T2	left colon	mucinous adenocarcinoma	female	48	stage 3	4	6	10	Yes	MSS	-	-
9	P37_T3	left colon	mucinous adenocarcinoma	female	48	stage 3	6	7	13	Yes	MSS	-	-
10	P43_T1	right colon	adenocarcinoma	male	32	stage 4	12	5	17	Yes	MSS	-	-
11	P63_T1	sigmoid colon	adenocarcinoma	female	88	stage 2	13	8	21	Yes	MSS	-	-
12	P69_T1	left colon	adenocarcinoma	female	50	stage 2	5	3	8	Yes	MSS	-	-
13	P69_T2	left colon	adenocarcinoma	female	50	stage 2	5	3	8	Yes	MSS	-	-
14	P71_T1	right colon	adenocarcinoma	male	71	stage 2	6	6	12	Yes	MSS	-	-
15	P71_T2	right colon	adenocarcinoma	male	71	stage 2	6	6	12	Yes	MSS	-	-
16	P77_T1	right colon	mucinous adenocarcinoma	male	79	stage 3	2	5	7	Yes	MSS	-	-
17	P77_T2	right colon	mucinous adenocarcinoma	male	79	stage 3	2	9	11	Yes	MSS	-	-
18	P79_T1	right colon	adenocarcinoma	female	84	stage 3	6	2	8	Yes	MSS	-	-
19	P79_T2	right colon	adenocarcinoma	female	84	stage 3	7	8	15	Yes	MSS	-	-
20	P83_T1	right colon	mucinous adenocarcinoma	female	76	stage 3	5	4	9	Yes	MSS	-	-
21	P87_T1	rectum	mucinous adenocarcinoma	female	79	stage 2	6	0	6	Yes	MSS	-	-
22	P89_T2	right colon	adenocarcinoma	female	69	stage 2	9	3	12	Yes	MSS	-	-
23	P91_T1	right colon	signet-ring cell adenocarcinoma	male	73	stage 3	5	2	7	Yes	MSS	-	-
24	P93_T1	rectum	adenocarcinoma	female	50	stage 3	11	8	19	Yes	MSS	-	-
25	P93_T2	rectum	adenocarcinoma	female	50	stage 3	9	6	15	Yes	MSS	-	-
26	P95_T1	sigmoid colon	adenocarcinoma	male	73	stage 3	3	8	11	Yes	MSS	-	-
27	P95_T2	sigmoid colon	adenocarcinoma	male	73	stage 3	12	8	20	Yes	MSS	-	-
28	P97_T1	sigmoid colon	adenocarcinoma	male	73	stage 2	4	7	11	Yes	MSS	-	-
29	P97_T2	sigmoid colon	adenocarcinoma	male	73	stage 2	4	6	10	Yes	MSS	Yes	Yes

5.4 GENE EXPRESSION PROFILE

An analysis was run on the Affymetrix® Transcriptome Analysis Console (TAC) Software (Affymetrix UK Ltd.) comparing MSI, LB and HB samples, with all the 25 normal tissue samples (N) (Table 2) being used a control group.

5.4.1 EXPRESSION OF MLH1 IN MSI SAMPLES

Expression of *MLH1* in the MSI group was compared to that in the HB group, in the LB group and in normal tissue samples. Figure 10 reports bi-weight average signal (Robust Multi-array Average, RMA) values for *MLH1* gene in the four groups. *MLH1* gene expression decrease could represent a surrogate marker of *MLH1* promoter hypermethylation.

As expected, *MLH1* expression appeared to be significantly reduced in the MSI group compared to normal tissues (2.12-fold decrease) and to the HB and LB tumors.

Somatic hypermethylation of both alleles of *MLH1* promoter is the primary cause of microsatellite instability in sporadic CRCs (Lynch et al., 2015), as could be the case for P75_T1, P41_T1 and P59_T2.

P29_T1 tumor from a Lynch syndrome patient also showed reduced *MLH1* expression. In the “two hit” model for Lynch syndrome, somatic *MLH1* promoter hypermethylation and consequent reduced gene expression may represent the second hit (Valo et al., 2015).

P13_T1 showed the highest RMA value of the MSI group. This tumor harbored a *MLH1* p.Arg487* premature stop codon mutation (45% of the sequencing reads) and a CN-LOH on chromosome 3 including *MLH1*.

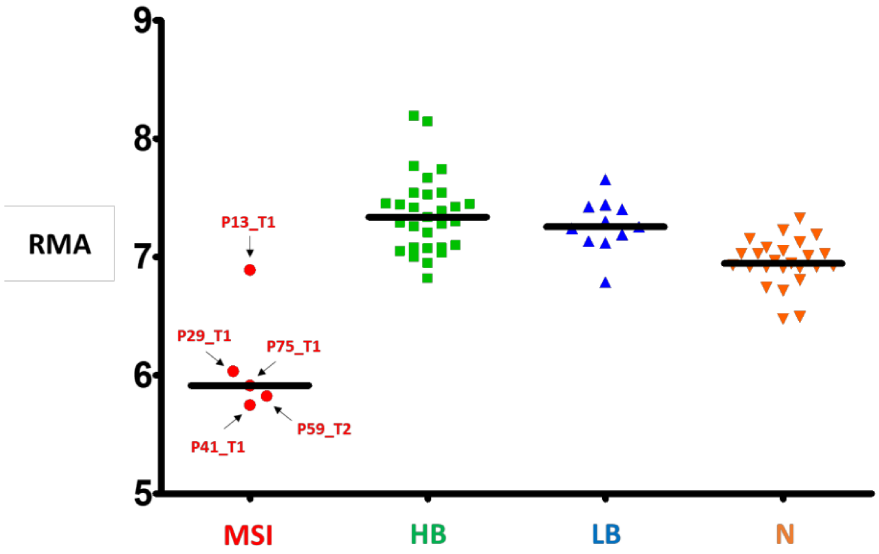


Figure 10: MLH1 gene expression is decreased in the MSI group compared to HB, LB and normal tissues. On the y axis, Bi-weight Average Signal (Robust Multi-array Average, RMA). Horizontal bars, median value.

5.4.2 DIFFERENTIALLY EXPRESSED GENES IN CRC GROUPS COMPARED TO NORMAL TISSUES

The first step was looking for those genes strongly upregulated in all CRC groups compared to normal colonic tissue, regardless of the extent of variation in gene upregulation among CRC tumor groups. Therefore, two filters were imposed to be met simultaneously: FDR p-value <0.05 (filter 1); fold change (FC) increase in each of the three CRC groups compared to normal colonic tissue (HB vs N, LB vs N, MSI vs N) >3.5. The top 20 upregulated genes in CRCs, listed in descending order according to fold change increase (not shown) of gene expression in the largest group (HB) compared to normal tissue (N), can be found in Figure 11.

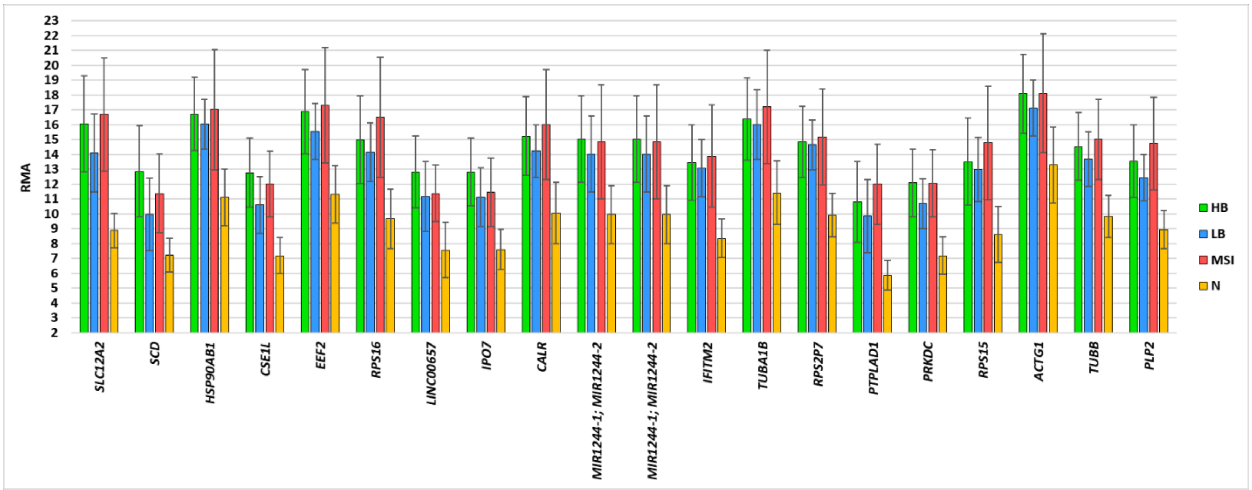


Figure 11: Top 20 genes upregulated in all CRC groups (HB, LB, MSI) compared to normal colonic tissue (N). Two filters have been imposed simultaneously: FDR p-value (all conditions) <0.05 (filter 1); FC increase in each of the three CRC groups compared to normal colonic tissues (HB vs N, LB vs N, MSI vs N) >3.5 (filter 2). Genes are listed in descending order according to fold change increase (not shown) of gene expression in the largest group (HB) compared to normal tissues (N). RMA values and corresponding standard deviations are reported on the y axis. The minimum value on the y axis is 2.

The *SLC12A2* gene, encoding the Solute Carrier Family 12 Member 2, also known as Na-K-Cl cotransporter isoform 1 (NKCC1) was on top of the list. It encodes a membrane channel allowing for chloride (Cl^-) entry (coupled to Na^+ and K^+ entry) across the basolateral membranes of gastrointestinal chloride-secreting epithelial cells (Frizzell and Hanrahan, 2012). Chloride secretion across the apical membrane is then achieved through apical channels such as CFTR. SLC12A2 is expressed in intestinal crypt cells (van de Wetering et al., 2002) and indeed Cl^- secretion appears to be a feature of crypt cells, which is lost during differentiation to enterocytes (Matthews et al., 1998; Welsh et al., 1982). A role for NKCC1 in cancer has been demonstrated for stomach, esophagus and brain tumors (Garzon-Muvdi et al., 2012; Shiozaki et al., 2011, 2014).

The second position was held by the *SCD* gene, encoding the Stearoyl-CoA desaturase-1, a delta-9-desaturase catalyzing the introduction of double bonds into position C9 of fatty acid chains, also required for the biosynthesis of monounsaturated fatty acids, which are important for cancer cells to satisfy their high demand for membrane lipids (Igal, 2010). Moreover, SCD1 is a regulator of de novo fatty acid synthesis in cancer cells since it influences the activity of acetylCoA carboxylase (ACC), the enzyme catalyzing the synthesis of malonylCoA from carboxylation of cytosolic acetylCoA, the first step of de novo fatty acids synthesis (Scaglia et al., 2009).

Another upregulated gene, *CSE1L*, encodes the Chromosome segregation 1-like protein, which can be found both in the nucleus, where it binds to a subset of p53 target promoters and regulates their transcription (Tanaka et al., 2007), and in the cytoplasm, where it interacts with microtubules and

promotes cell migration (Tai et al., 2010). Most CRCs (up to 99%) are positive for CSE1L protein expression, - whereas normal colorectal glands are weakly stained - and tumor CSE1L immunochemical staining correlates with tumor invasiveness (Tai et al., 2013).

To select those genes which were upregulated at a similar extent in all the three CRC groups compared to normal colonic tissues, we imposed an additional filter to the above-mentioned ones: when comparing each tumor group with the other two, the difference in gene expression shall not be significant (fold changes in gene expression of HB vs LB, HB vs MSI and LB vs MSI comparisons must all be in the range between -2 and 2). Genes were listed in descending order according to fold change increase (not shown) of gene expression in the largest group (HB) compared to normal tissue (N) (Figure 12).

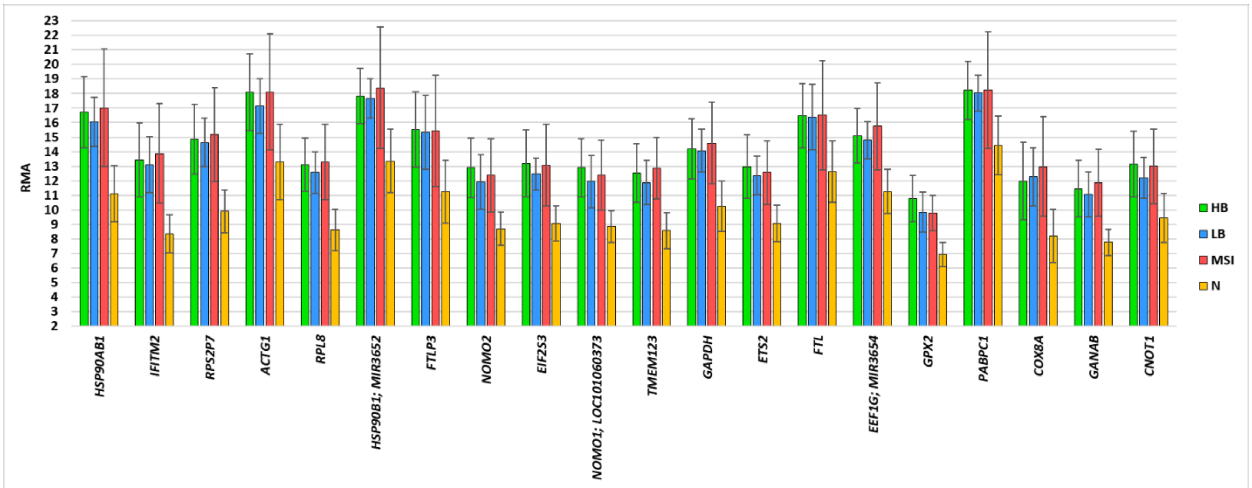


Figure 12: Top 20 genes upregulated in all tumors (HB, LB, MSI) at similar extent compared to normal colonic tissue (N). Three types of filter have been imposed simultaneously: FDR p-value (all conditions) <0.05 (filter 1); FC increase in each of the three CRC groups compared to normal colonic tissues (HB vs N, LB vs N, MSI vs N) >3.5 (filter 2); for each of the listed genes, the three tumor groups do not significantly differ in gene expression (fold changes of HB vs LB, HB vs MSI and LB vs MSI comparisons are all between -2 and 2) (filter 3). Genes are listed in descending order according to fold change increase (not shown) of gene expression in the largest group (HB) compared to normal tissues (N). RMA values and corresponding standard deviations are reported on the y axis. The minimum value on the y axis is 2.

Some genes, which could be seen in Figure 11 as generally upregulated in tumors, such as heat shock protein 90kDa alpha (cytosolic), class B member 1 (*HSP90AB1*), interferon induced transmembrane protein 2 (*IFITM2*), actin, gamma 1 (*ACTG1*), also turned to be upregulated at a similar extent (Figure 12). Others, such as NODAL modulator 2 (*NOMO2*), glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), glutathione peroxidase 2 (gastrointestinal) (*GPX2*), appeared in the top 20 only when a similar level of upregulation was considered (Figure 12).

CD44, a gene upregulated at a similar extent in all tumor groups, although outside the top 20 ranking of Figure 12 (it occupies the 21st position of the ranking, not shown), encodes a transmembrane glycoprotein of the cell adhesion molecules (CAMs) family participating in cell-cell and cell-extracellular matrix (ECM) interactions (Goodison et al., 1999). CD44 can interact with ECM components such as hyaluronic acid, collagen, fibronectin, fibrinogen, laminin, chondroitin sulphate, osteopontin, matrix metalloproteinase 9 (MMP9) as well serglycin (Naor et al., 1997; Orian-Rousseau, 2010). *CD44* is expressed in the proliferative compartment at the colon crypt bottom (van de Wetering et al., 2002). Intestinal stem cells express *CD44* alternative splicing variant isoforms (CD44v), but lack CD44 standard (CD44s) isoforms (Zeilstra et al., 2014). Colorectal cancer stem cells express CD44v6, containing the exon v6, and such CD44v6^{+} cells show activation of β-catenin and a higher epithelial–mesenchymal transition (EMT) and metastatic phenotype (Todaro et al., 2014).

The next step was searching for those genes strongly downregulated in all CRC groups compared to normal colonic tissue, regardless of the extent of variation in gene downregulation among CRC tumor groups. Therefore, two conditions were imposed to be met simultaneously: FDR p-value <0.05 (filter 1); FC decrease in each of the three CRC groups compared to normal colonic tissues (HB vs N, LB vs N, MSI vs N) <-3.5. The top 20 downregulated genes among CRCs, listed in ascending order according to fold change decrease (not shown) of gene expression in the largest group (HB) compared to normal tissue (N), can be found in Figure 13.

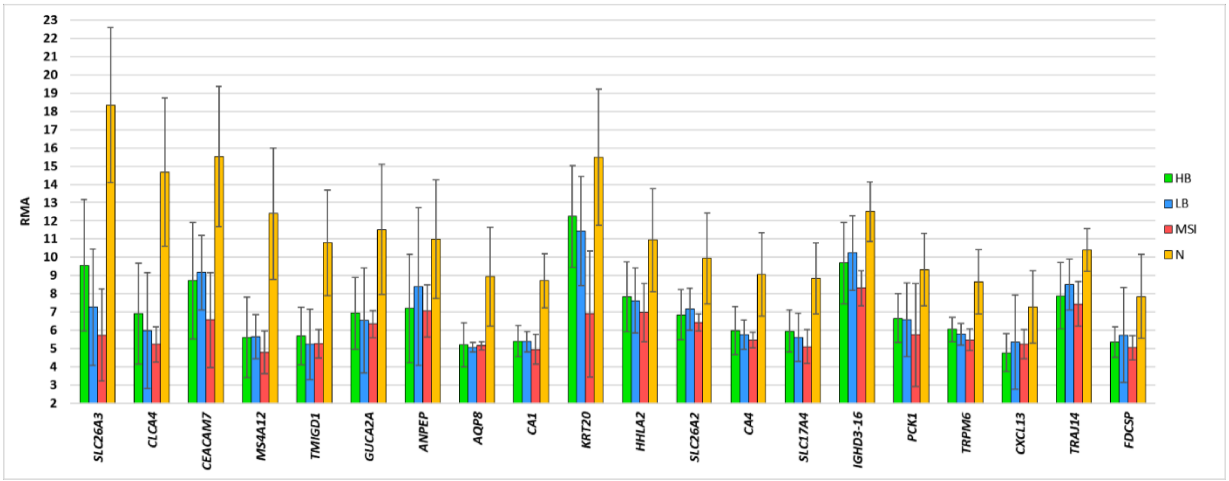


Figure 13: Top 20 genes downregulated in all CRC groups (HB, LB, MSI) compared to normal colonic tissue (N). Two filters have been imposed simultaneously: FDR p-value <0.05 (filter 1); FC decrease in each of the three CRC groups compared to normal colonic tissues (HB vs N, LB vs N, MSI vs N) <-3.5 (filter 2). Genes are listed in ascending order according to fold change decrease (not shown) of gene expression in the largest group (HB) compared to normal tissues (N). RMA values and corresponding standard deviations are reported on the y axis. The minimum value on the y axis is 2.

At the first position, there was *SLC26A3* gene, encoding a Cl^-/HCO_3^- exchanger with preferential expression in the colon, especially in differentiated cells (Höglund et al., 1996; Melvin et al., 1999), followed by *CLCA4*, encoding the Chloride Channel Accessory 4, which is also preferentially expressed in the colon (Agnel et al., 1999). Among downregulated genes several markers of mature enterocytes were found (*SLC26A3*, *MS4A12*, *CA1)* (Dalerba et al., 2011), as well as CEACAM7, which is expressed on the apical membrane of mature enterocytes (Schölzel et al., 2000).

To select those genes which were downregulated at a similar extent in all the three CRC groups compared to normal colonic tissues, we imposed an additional condition to the above-mentioned ones: if we compare each tumor group with the other two, the difference in gene expression should not be significant (fold changes in gene expression of HB vs LB, HB vs MSI and LB vs MSI comparisons must all be in the range between -2 and 2). Genes were listed in ascending order according to fold change decrease (not shown) of gene expression in the largest group (HB) compared to normal tissue (N). The top 20 entries can be seen in Figure 14.

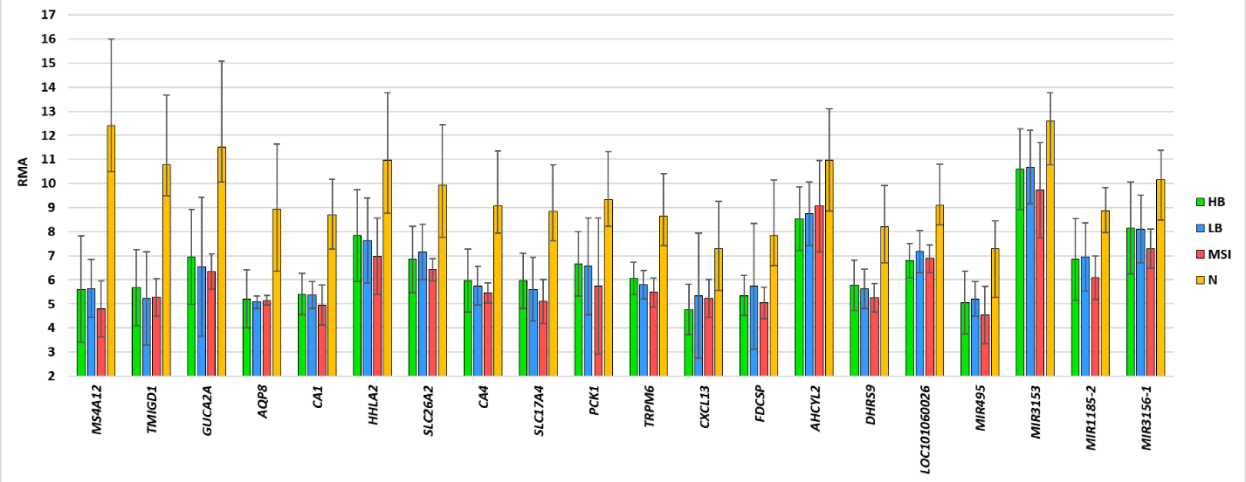


Figure 14: Top 20 genes downregulated in all tumors (HB, LB, MSI) at similar extent compared to normal colonic tissue (N). Three types of filter have been imposed simultaneously: FDR p-value (all conditions) <0.05 (filter 1); FC decrease in each of the three CRC groups compared to normal colonic tissues (HB vs N, LB vs N, MSI vs N) <-3.5 (filter 2); for each of the listed genes, the three tumor groups do not significantly differ in gene expression (fold changes of HB vs LB, HB vs MSI and LB vs MSI comparisons are all between -2 and 2) (filter 3). Genes are listed in ascending order according to fold change decrease (not shown) of gene expression in the largest group (HB) compared to normal tissues (N). RMA values and corresponding standard deviations are reported on the y axis. The minimum value on the y axis is 2.

Most of the genes generally downregulated in CRC compared to normal tissue (*MS4A12*, *TMIGD1*, *GUCA2A*, *CA1*, *CA4*, *SLC26A2*, *SLC17A4*) are also downregulated at a similar extent in the three CRC groups (HB, LB and MSI) (as can be seen by comparing Figure 13 and Figure 14). On the other hand, *SLC26A3* gene expression, despite being in all the three tumor groups, is strikingly lower in MSI tumors

(6330.5-fold decrease compared to normal) than in LB (2169.2-fold decrease against normal) and HB tumors (443.28-fold decrease against normal).

Afterwards, significant differential gene expression against normal tissues was investigated for each single tumor group, regardless of changes in gene expression being concordant among tumor groups. Genes were selected under the following conditions: FDR p-value <0.05 and fold change >3.5 or <-3.5 for upregulated and downregulated genes, respectively. The top 30 genes for each comparison can be found in Figure 15.

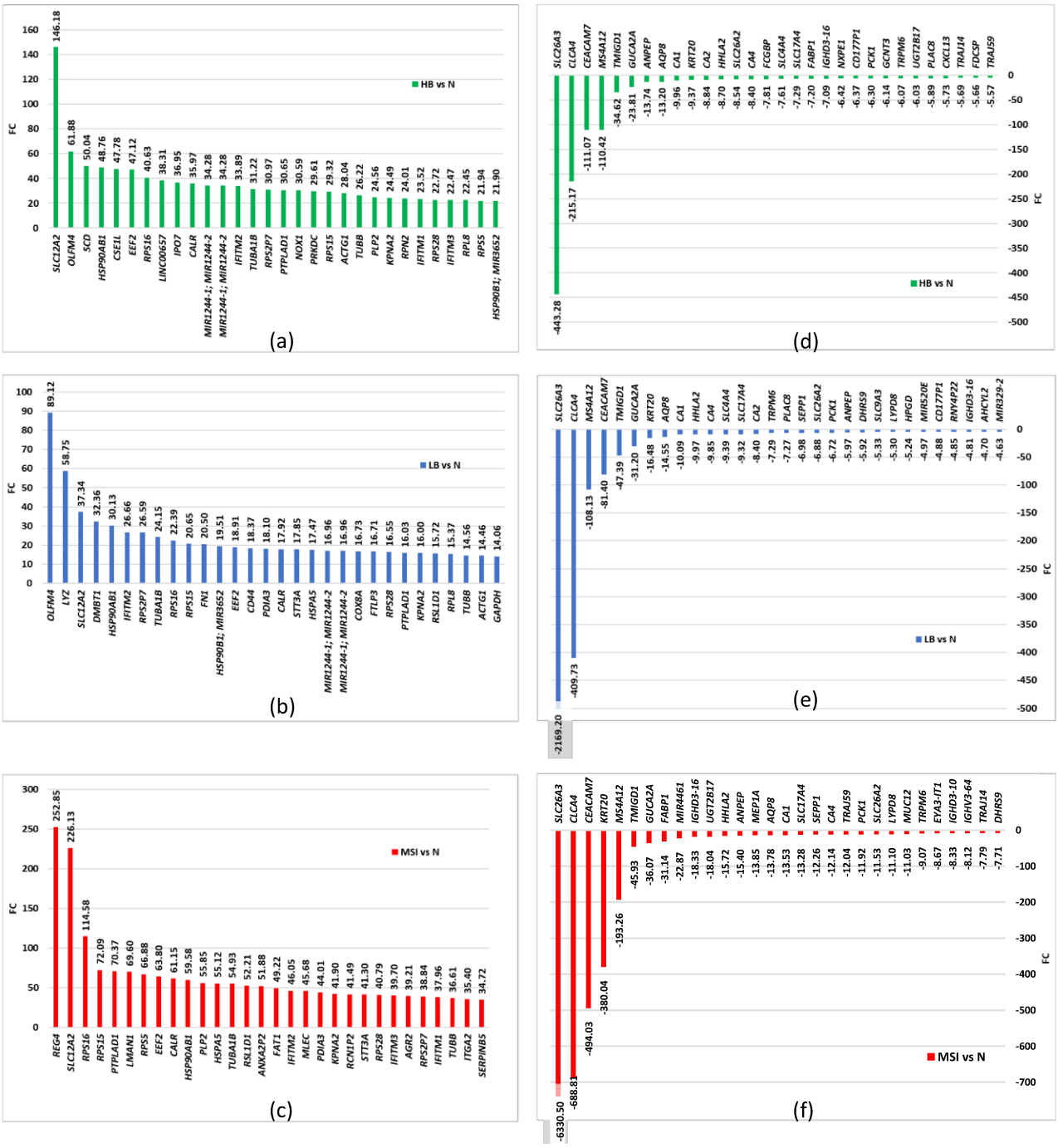


Figure 15: Top 30 differentially expressed genes between CRC groups and normal tissue. (a, b, c) Genes upregulated in HB CRCs (a), LB CRCs (b) and MSI CRCs (c) compared to normal colonic tissues. (d, e, f) Genes downregulated in HB CRCs (d), LB CRCs (e) and MSI CRCs (f) compared to normal colonic tissue. Fold change values for change in gene expression are reported at the tip of corresponding columns. (e) In LB CRCs, *SLC26A3* gene showed a 2169.20-fold decrease compared to normal (exceeding y axis minimum value; fold change value for this gene is not in scale). (f) In MSI CRCs, *SLC26A3* gene showed a 6330.50-fold decrease compared to normal (exceeding y axis minimum value; fold change value for this gene is not in scale).

5.4.3 DIFFERENTIALLY EXPRESSED GENES ACROSS CRC GROUPS

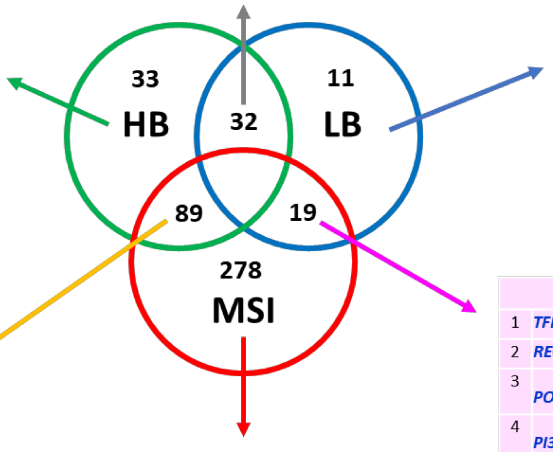
In order to characterize those genes who are not only differentially expressed compared to normal tissue, but also specifically upregulated in each of the three CRC groups compared to the other two, we imposed the following conditions to be met simultaneously: FDR p-value <0.05 (filter 1); FC in gene expression of the examined CRC group compared to normal colonic tissue > 2 or < -2 (filter 2); FC in gene expression of the examined CRC group compared to either of the other tumor groups >2 (filter 3). A detailed list of the selected genes can be found in Supplementary Table 6, Supplementary Table 7 and Supplementary Table 8.

To find specifically downregulated genes in one CRC group compared to the other two, we proceeded as before, but modifying filter 3 in the way that FC in gene expression of the examined CRC group compared to either of the other tumor groups shall be <-2. A detailed list of the selected genes can be found in Supplementary Table 12, Supplementary Table 13 and Supplementary Table 14.

To characterize those genes which were not only differentially expressed compared to normal tissues, but also significantly upregulated at similar extent in two CRC groups compared to the remaining one, we applied the following conditions: FDR p-value <0.05 (filter 1); FC in gene expression of each of the two examined CRC groups compared to normal colonic tissues > 2 or < -2 (filter 2); FC in gene expression for each of the two examined CRC groups compared to the remaining one >2 (filter 3). A detailed list of the selected genes can be found in Supplementary Table 9, Supplementary Table 10 and Supplementary Table 11.

To find specifically downregulated genes in one CRC group compared to the other two, we proceeded as before, but modifying filter 3 in the way that FC in gene expression of each of the examined CRC group compared to remaining one shall be <-2. A detailed list of the selected genes can be found in Supplementary Table 15, Supplementary Table 16 and Supplementary Table 17.

Two Venn diagrams were drawn, one for differentially upregulated genes across CRC groups (Figure 16) and the other one for differentially downregulated genes across CRC groups (Figure 17).



HB			
1	NOX1	18	FARP1; FARP1-IT1
2	SCD	19	POFUT1; MIR1825
3	AREG; AREGB	20	ACSL5
4	AREG; AREGB	21	FAM111B
5	TSPAN6	22	NKD1
6	SLC26A3	23	TM9SF3
7	CEACAM6	24	HSPH1
8	CFTR	25	MACC1
9	PRDX5	26	CCL24
10	VIL1	27	CDHR1
11	HUNK	28	RNF43
12	IPO7	29	RNF157
13	ACE2	30	TM4SF1
14	LINC00657	31	TFRC
15	AXIN2	32	PRPF6
16	EREG	33	PROX1
17	ATP9A		

HB and LB					
1	OLFM4	12	FABP1	23	LYPD8
2	KRT20	13	MIR4461	24	HOXB9
3	CFTR	14	TGFBI	25	IGHD2-21
4	SLC26A3	15	LYZ	26	TRAJ59
5	DMBT1	16	SPINK1	27	TGFBR2
6	DACH1	17	MUC12	28	MIR514B
7	KIAA1324	18	CCL20	29	CDHR5
8	DPEP1	19	USP12	30	TRAJ2
9	CCAT1	20	IGHD3-16	31	IGHD3-10
10	CYP2B6	21	PRSS8	32	IGHV3-64
11	CEACAM7	22	EYA3-IT1		

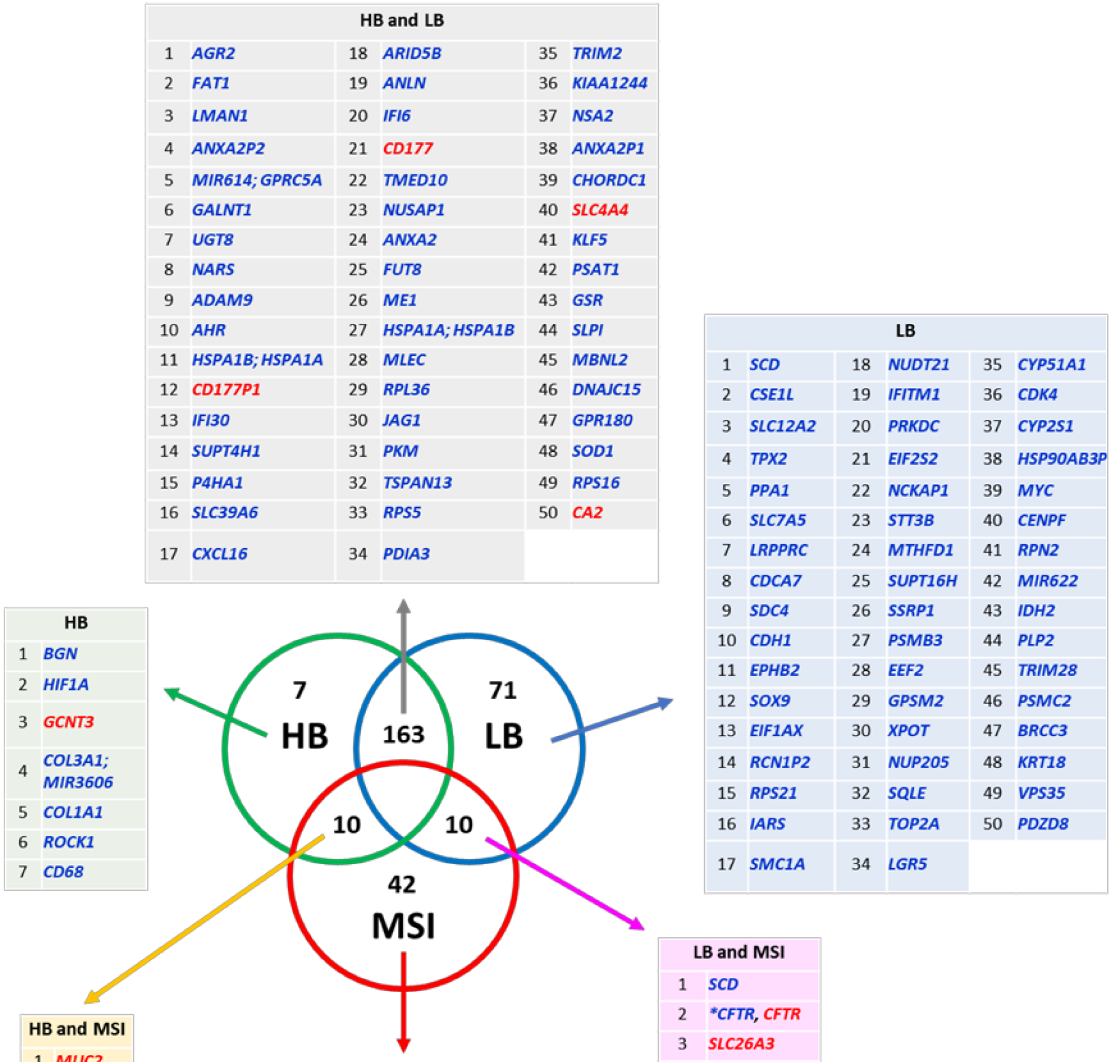
LB	
1	LYZ
2	MUC2
3	LAPTM5
4	DMBT1
5	CTSE
6	FABP1
7	SRGN
8	ARHGDIB
9	TCN1
10	NXPE1
11	ANPEP

LB and MSI			
1	TFF1	11	EMP3
2	REG4	12	SPP1
3	POSTN	13	FCER1G
4	PI3	14	HSPA4L
5	MUC5B	15	COL1A1
6	BGN	16	SERPINB5
7	HIF1A	17	CREB3L1
8	SAMD5	18	ROCK1
9	CD55	19	CD68
10	COL3A1; MIR3606		

HB and MSI			
1	SCD	26	PRKDC
2	CSE1L	27	EIF2S2
3	SLC12A2	28	NCKAP1
4	TPX2	29	STT3B
5	PPA1	30	MTHFD1
6	SLC7A5	31	SUPT16H
7	AGMAT	32	ECH1
8	LRPPRC	33	SSRP1
9	PM20D2	34	PSMB3
10	CDCA7	35	EEF2
11	SDC4	36	GPSM2
12	CLDN1	37	XPOT
13	UTP20	38	EXOSC5
14	CDH1	39	NUP205
15	GSS	40	SQLE
16	ARPC1A	41	TOP2A
17	EPHB2	42	LGR5
18	SOX9	43	CYP51A1
19	EIF1AX	44	CDK4
20	RCN1P2	45	CYP2S1
21	RPS21	46	HSP90AB3P
22	IARS	47	HIBADH
23	SMC1A	48	MYC
24	NUDT21	49	CENPF
25	IFITM1	50	RPN2

MSI					
1	REG4	18	HSPA4L	35	TMEM66
2	SERPINB5	19	CREB3L1	36	PLA2G2A
3	PLA2G4A	20	UGT8	37	SS18
4	AGR2	21	NARS	38	ASAH1
5	MUC5B	22	KITLG	39	PLA2G16
6	SPON1	23	ADAM9	40	CXCL16
7	CD55	24	AHR	41	ARID5B
8	SPP1	25	DUSP6	42	ANLN
9	DPP4	26	HSPA1B; HSPA1A	43	IFI6
10	FAT1	27	DAPK1	44	LY6E
11	LMAN1	28	IFI30	45	IRS2
12	SAMD5	29	PBK	46	TMED10
13	AFAP1-AS1	30	SUPT4H1	47	GATA6
14	IQGAP2	31	HTR1D	48	NT5E
15	ANXA2P2	32	P4HA1	49	NUSAP1
16	MIR614; GPRC5A	33	SLC6A14	50	CRYZ
17	GALNT1	34	SLC39A6		

Figure 16: Differentially upregulated genes across different CRC groups. Among genes differentially expressed in comparison to normal colonic tissue (FC <-2 or FC >2), we looked for those specifically upregulated (FC >2) in a single group compared to the other two, and for those upregulated (FC >2) in either two groups when compared to the remaining one. Within each analyzed group or group intersection, the number of overexpressed genes is reported. Top ranking overexpressed genes (up to 50) were detailed for each group. Blue, genes upregulated in comparison to normal colonic tissue; red, genes downregulated in comparison to normal colonic tissue. FDR p-value <0.05 was used.



HB and LB					
1	AGR2	18	ARID5B	35	TRIM2
2	FAT1	19	ANLN	36	KIAA1244
3	LMAN1	20	IFI6	37	NSA2
4	ANXA2P2	21	CD177	38	ANXA2P1
5	MIR614; GPRC5A	22	TMED10	39	CHORDC1
6	GALNT1	23	NUSAP1	40	SLC4A4
7	UGT8	24	ANXA2	41	KLF5
8	NARS	25	FUT8	42	PSAT1
9	ADAM9	26	ME1	43	GSR
10	AHR	27	HSPA1A; HSPA1B	44	SLPI
11	HSPA1B; HSPA1A	28	MLEC	45	MBNL2
12	CD177P1	29	RPL36	46	DNAJC15
13	IFI30	30	JAG1	47	GPR180
14	SUPT4H1	31	PKM	48	SOD1
15	P4HA1	32	TSPAN13	49	RPS16
16	SLC39A6	33	RPS5	50	CA2
17	CXCL16	34	PDIA3		

HB	
1	BGN
2	HIF1A
3	GCNT3
4	COL3A1; MIR3606
5	COL1A1
6	ROCK1
7	CD68

LB					
1	SCD	18	NUDT21	35	CYP51A1
2	CSE1L	19	IFITM1	36	CDK4
3	SLC12A2	20	PRKDC	37	CYP2S1
4	TPX2	21	EIF2S2	38	HSP90AB3P
5	PPA1	22	NCKAP1	39	MYC
6	SLC7A5	23	STT3B	40	CENPF
7	LRPPRC	24	MTHFD1	41	RPN2
8	CDCA7	25	SUPT16H	42	MIR622
9	SDC4	26	SSRP1	43	IDH2
10	CDH1	27	PSMB3	44	PLP2
11	EPHB2	28	EEF2	45	TRIM28
12	SOX9	29	GPSM2	46	PSMC2
13	EIF1AX	30	XPOT	47	BRCC3
14	RCN1P2	31	NUP205	48	KRT18
15	RPS21	32	SQLE	49	VPS35
16	IARS	33	TOP2A	50	PDZD8
17	SMC1A	34	LGR5		

LB and MSI	
1	SCD
2	*CFTR, CFTR
3	SLC26A3
4	IPO7
5	LINC00657
6	AXIN2
7	TM9SF3
8	HSPH1
9	RNF43
10	TFRC

HB and MSI	
1	MUC2
2	MIR548I4
3	UGT2B17
4	FCGBP
5	FABP1
6	NXPE1
7	MUC4
8	ANPEP
9	SI
10	UGT2B15

MSI					
1	OLFM4	15	TGFBI	29	MIR376A2
2	KRT20	16	IGHD2-15	30	PPP1R9A
3	CFTR	17	A1CF	31	MIR196A1
4	SLC26A3	18	SULT1B1	32	LYPD8
5	ATP10B	19	RNY5	33	IGHD2-21
6	VIL1	20	MUC12	34	TRAJ25
7	MEP1A	21	UGT2B17	35	UGT2B15
8	IL2RG	22	CCL20	36	TRAJ59
9	LRRC19	23	MIR4437	37	TGFBR2
10	CYP2B6	24	HMGN5	38	MIR514B
11	CEACAM7	25	MLH1	39	CDHR5
12	FABP1	26	IGHD3-16	40	TRAJ2
13	TFCP2L1	27	EYA3-IT1	41	IGHD3-10
14	MIR4461	28	MIR4782	42	IGHV3-64

Figure 17: Downregulated genes across different groups. Among genes differentially expressed in comparison to normal colonic tissue (FC <-2 or FC >2), we looked for those specifically downregulated (FC <-2) in a single group compared to the other two, and for those downregulated (FC <-2) in either two groups when compared to the remaining one. Within each analyzed group or group intersection, the number of overexpressed genes is reported. Top ranking overexpressed genes (up to 50) were detailed for each group. Blue, genes upregulated in comparison to normal colonic tissue; red, genes downregulated in comparison to normal colonic tissue. *For the "LB and MSI" intersection group, *CFTR* gene was upregulated in LB compared to normal tissues (blue) but downregulated in MSI compared to normal tissues (red). FDR p-value <0.05 was used.

5.4.3.1 GENES DIFFERENTIALLY EXPRESSED IN HB COMPARED TO LB AND MSI

A gene selectively upregulated in HB CRCs was *NOX1* (Figure 16), encoding the NADPH Oxidase 1, a membrane-bound flavin dehydrogenase preferentially expressed in normal colon epithelial cells (Lambeth et al., 1999), which catalyzes one-electron reduction of oxygen to superoxide anion. NOX1 may activate Wnt/β-catenin and Notch signaling and play a role in colon stem cell fate, since *NOX1*-deficient mice show unbalanced differentiation towards goblet cells rather than towards absorptive enterocytes (Coant et al., 2010). NOX1 is involved in reactive oxygen species (ROS)-mediated mitogenic responses to growth factors (Juhasz et al., 2017; Lambeth et al., 1999). Moreover, NOX1-mediated increase in ROS levels results in increased generation of 8-oxoguanine DNA oxidative lesions (Chiera et al., 2008). *NOX1* gene is upregulated by the RAS/MAPK pathway with subsequent raise in ROS levels, affecting malignant transformation and neoplastic growth (Mitsushita et al., 2004). *NOX1* is overexpressed in adenomas and differentiated CRCs (Fukuyama et al., 2005; Laurent et al., 2008).

AREG (encoding amphiregulin) and *EREG* (encoding epiregulin) gene upregulation was also typical of the HB group. Two copies of the *AREG* gene are found on chromosome 4q13.3. *EREG* is located upstream of *AREG* genes, on the same chromosome region. Both *AREG* and *EREG* are members of the epidermal growth factor (EGF) family, and their transcripts encode transmembrane precursors (Pro-AREG and Pro-EREG, respectively), which are subsequently cleaved by ADAM17 (Sahin et al., 2004) to produce soluble AREG and EREG, which can bind to EGF receptors and stimulate downstream signaling. *AREG* overexpression is common in CRC (Ciardiello et al., 1991), as well as *EREG* overexpression, both being inversely correlated with promoter methylation (Lee et al., 2016a). Qu et al. (2016) demonstrated that *EREG* upregulation during the adenoma–carcinoma transition is associated with promoter demethylation. Activation of EGF signaling and *AREG*/*EREG* overexpression have been found to be important features of CIN CRCs, especially in the distal colon (Lee et al., 2016a; Missiaglia et al., 2014).

HB tumors also overexpressed *AXIN2*, an important negative regulator of Wnt/β-catenin signaling pathway (Thorvaldsen et al., 2017). Wnt signaling activation is common in CRC, and is accompanied by upregulation of Wnt pathway feedback inhibitors, especially of *AXIN2* (Lustig et al., 2002), which would act as an oncogene rather than as a tumor suppressor, and would contribute to trigger the epithelial-mesenchymal transition (EMT) program and the metastatic phenotype (Wu et al., 2012).

5.4.3.2 GENES DIFFERENTIALLY EXPRESSED IN LB COMPARED TO HB AND MSI

LB tumors show differential overexpression of *LYZ* (lysozyme) and *MUC2* gene. Lysozyme is an antibacterial product secreted by Paneth cells and lymphocytes, whereas MUC2 is the most abundant constituent of gastrointestinal mucus. MUC2 expression is generally reduced in CRC, with the exception of mucinous CRC, which preserve MUC2 expression (Gratchev et al., 2001; Iwase et al., 2005).

DMBT1 gene product is a secreted glycoprotein expressed by epithelial cells of the crypt base and of the midcrypts throughout normal small intestine and colon (Renner et al., 2007). *DMBT1* is involved in epithelial differentiation (Kang and Reid, 2003), and its protein expression progressively increases during colorectal carcinogenesis from early to late stages (Peng et al., 2016).

LAPTM5, encoding Lysosomal Protein Transmembrane 5 - a protein involved in macrophage activation (Glowacka et al., 2012) – was also upregulated in the LB group.

5.4.3.3 GENES DIFFERENTIALLY EXPRESSED IN MSI COMPARED TO HB AND LB

MSI CRCs show a peculiar pattern of gene upregulation compared to normal tissue. The gene showing highest upregulation was *REG4*, whose protein product belongs to the calcium (C-type) dependent lectin superfamily (Hartupee et al., 2001). REG4 has been shown to activate the Akt/GSK-3β/β-catenin signaling and to promote cancer cell migration and invasion (Bishnupuri et al., 2014; Rafa et al., 2010). High expression of REG4 has been found in neuroendocrine cells of both normal small intestine and normal colon, but also in colorectal cancer (Oue et al., 2007), especially mucinous colorectal adenocarcinomas (Kaprio et al., 2014; Oue et al., 2005), as well as in gastric adenocarcinomas with intestinal mucin phenotype (Oue et al., 2005), intestinal-type intraductal papillary mucinous neoplasms of the pancreas (Nakata et al., 2009) and mucinous ovarian cancer (Huang et al., 2014; Lehtinen et al., 2016).

SERPINB5, which is also known as MASPIN (mammary serine proteinase inhibitor), belongs to the serpin superfamily of protease inhibitors, which also includes α1-antitrypsin, plasminogen activator inhibitor 1 and 2, angiotensinogen, antithrombin, and others (Law et al., 2006). Although MASPIN has been regarded as a tumor suppressor (Sheng et al., 1996; Yin et al., 2005), its protein expression levels evaluated by immunohistochemistry (IHC) associate with MSI CRCs (Bettstetter et al., 2005) and with serrated lesions of the colon (hyperplastic polyps, sessile serrated adenomas/polyps) whereas conventional colorectal adenomas are negative for MASPIN or only show focal expression (Rubio &

Kaufeldt, 2015.). Indeed, sporadic MSI tumors are thought to originate through the serrated neoplasia pathway (Jass, 2007).

PLA2G4A (Phospholipase A2 Group IVA), also known as *CPLA2* (cytosolic PLA2), is a member of the PLA2 family of phospholipases that hydrolyze glycerophospholipids, with PLA2G4A showing a preference for those containing arachidonic acid (AA) in position 2 (Hu et al., 2011). AA can be used for Prostaglandin E2 (PGE_2) synthesis by inducible Cyclooxygenase-2 (COX2), and PGE_2 induces expansion of colorectal cancer stem cells by stimulating NF-κB via Prostaglandin E Receptor 4 (EP4)-mediated activation of PI3K and MAPK pathways (Wang et al., 2015). PGE_2 can stimulate tumor growth by amplifying inflammatory responses in the tumor microenvironment via Prostaglandin E receptor 2 (EP2), which is expressed by infiltrating neutrophils and by tumor-associated fibroblasts (TAFs) of tumor stroma (Aoki and Narumiya, 2017). In human breast cancer, it has been demonstrated that cPLA2α has a role in TGF-β-induced epithelial-mesenchymal transition (EMT) via the PI3K/Akt/GSK-3β/β-catenin pathway, with an impact on tumor cells invasion and migration. cPLA2α-derived PGE_2 can promote G1 progression by modulating Forkhead box protein O1 (FOXO1) activity via PI3K/AKT signaling (Naini et al., 2016). Furthermore, cPLA2α plays a role in cell cycle re-entry of quiescent prostate cancer cells, and inhibition of cPLA2α impairs cell cycle re-entry (Yao et al., 2015). Deletion of cPLA2α in $Apc^{Min/+}$ mice lead to a consistent reduction in tumor number in the small intestine (Hong et al., 2001). Interestingly, KIT Ligand is one of the factors capable of upregulating expression of cPLA2 (Murakami et al., 1995).

AGR2, encoded by the *AGR2* (Anterior Gradient 2, Protein Disulphide Isomerase Family Member) gene, is a protein required for MUC2 production by goblet cells (Park et al., 2009). AGR2 can also be secreted into the gastrointestinal mucus (Bergstrom et al., 2014) and extracellular AGR2 has been shown to activate stromal fibroblasts and make them acquire cancer-associated fibroblast (CAF)-like invasive properties in gastric signet-ring cell carcinoma (Tsuji et al., 2015). AGR2 can induce expression of MUC1 (a cancer-associated mucin) in pancreatic ductal adenocarcinoma cells (Norris et al., 2013).

MUC5B encodes for a gel-forming mucin expressed by a subset of colonic goblet cells at the crypt base (colocalizing with MUC2-containing goblet cells). MUC5B expression normally decreases until it disappears moving towards the upper part of the crypt (van Klinken et al., 1998; Larsson et al., 2011).

F-Spondin, encoded by the *SPON1* gene, is an extracellular matrix glycoprotein and a member of the Thrombospondin Type 1 Repeat (TSR) Superfamily (Adams and Tucker, 2000). It is involved in axon guidance during development in mice (Burstyn-Cohen et al., 1998; Tzarfati-Majar et al., 2001). Although F-spondin protein expression has been detected in ovarian carcinomas (Pyle-Chenault et al., 2005),

relevant information about an association between F-spondin and CRC could not be found in the literature.

CD55, previously known as Decay Accelerating Factor For Complement (DAF), is a phosphatidylinositol-anchored protein belonging to the membrane-bound complement-regulatory proteins, which inhibit autologous complement cascade activation (Lublin and Atkinson, 1989). CD55 immunohystochemical positivity is rarely observed in normal colonic epithelium; on the other hand, it increases in colorectal adenomas and in carcinomas, especially low grade ones (Koretz et al., 1992; Shang et al., 2014). Increased CD55 expression correlates with poor prognosis in CRC patients (Durrant et al., 2003).

Osteopontin (OPN), encoded by the *SPP1* gene, is a secreted extracellular matrix glycophosphoprotein belonging to the Small Integrin-Binding LIgand N-linked Glycoproteins (SIBLINGs) family (Bellahcène et al., 2008). OPN is involved in cell adhesion by interacting with cell surface integrin receptors and with integral membrane adhesion molecule CD44 (Huang et al., 2016). OPN specifically binds to CD44v6 variant (Orian-Rousseau, 2010) and induces CD44v6 expression, which is related to cancer cell migration and metastatic features (Todaro et al., 2014). OPN is overexpressed in CRC patients, where it appears to be associated with metastasis, since it impairs homotypic adhesion of CRC cells with each other, while promoting heterotypic adhesion between CRC cells and endothelial cells (Huang et al., 2012). *SPP1* was remarkably overexpressed in the MSI group studied in the present thesis.

Dipeptidyl Peptidase 4 (encoded by the *DPP4* gene), also known as cluster differentiation antigen CD26, is a surface glycoprotein and a member of the prolyl peptidase family (Lam et al., 2014), cleaving peptides at their N-terminus to release N-terminal dipeptides with proline (or alanine) at the penultimate position (Mentlein, 1999). CD26 binds proteins of the ECM like fibronectin and collagen, and degrades ECM, promoting migration and invasion (Ghersi et al., 2006). CD26 is a marker of a subset of colorectal cancer stem cells (CSCs) with metastatic potential, with differential expression between metastatic CRC and primary tumors (Pang et al., 2010). CD26 can be found in many human tissues, including the small and large intestine; its soluble form (sCD26) can be found in plasma, serum, secretions, urine and other biological fluids (Dinjens et al., 1989; Mentlein, 1999). *DPP4* is upregulated in CRC compared to normal mucosa, whereas plasmatic DPP4 activity is lower in CRC patients compared to healthy individuals, and behaves as a prognostic factor associated with worse overall and disease-free survival (Larrinaga et al., 2015).

FAT Atypical Cadherin 1 (encoded by the *FAT1* gene) belongs to the FAT gene family, whose members encode membrane proteins called protocadherins. FAT1 protein can be cleaved in its extracellular

region by furin before being expressed at the cell surface, the final receptor resulting from the association of the extracellular fragment with the transmembrane and cytoplasmic domains (Sadeqzadeh et al., 2011; Sopko and McNeill, 2009); an alternative furin-independent cleavage has been detected in human melanoma cells, generating a p65 membrane-bound cytoplasmic fragment not associated with the extracellular fragment; an intact non-cleaved form can also be expressed on the cell surface (Sadeqzadeh et al., 2011). Moreover, in pancreatic cells a soluble isoform of FAT1 is released by ADAM10 cleavage (Wojtalewicz et al., 2014). Silencing of *ADAM10* in a colon cancer cell line (HCT15) leads to accumulation of non-cleaved FAT1 protein (Pileri et al., 2016). *ADAM10* expression is increased in MSI samples compared to HB and LB tumors. Cytoplasmic domain of FAT1 binds β-catenin and keeps it at the cell membrane, preventing its translocation to the nucleus; mutations or deletions of FAT1 (found in several types of cancer including 7.7% of CRCs) allow for increased translocation of β-catenin to the nucleus and activation of Wnt/β-catenin pathway (Morris et al., 2013a, 2013b). FAT1 would therefore behave as a tumor suppressor, although on the other hand high expression of FAT1 has been reported to predict shorter relapse-free and overall survival in B-cell acute lymphoblastic leukemia (de Bock et al., 2011). In CRC, high FAT1 expression and plasma membrane localization has been observed regardless of β-catenin pathway activation.

LMAN1, also known as Endoplasmic Reticulum-Golgi Intermediate Compartment Protein 53 (ERGIC-53), encodes a non-glycosylated mannose-specific membrane lectin which works as a carrier transporting glycoproteins from the ER to the Golgi (Hauri et al., 2000), and is frequently mutated in MSI-H CRC as an early event, with a high frequency of biallelic mutations and loss of LMAN1 protein expression (despite the gene being transcribed) (Roeckel et al., 2009).

In contrast to MSI tumors, *AGR2*, *FAT1* and *LMAN1* were downregulated in HB and LB samples.

5.4.3.4 GENES DIFFERENTIALLY EXPRESSED IN HB AND LB COMPARED TO MSI

OLFM4, which is a marker for Lgr5+ intestinal stem cells (van der Flier et al., 2009), was upregulated in both HB and LB tumors compared to normal tissues. *OLFM4* has been characterized for being both a target and a negative regulator of the Wnt/β-catenin and NF-κB pathways in intestinal stem cells (Chin et al., 2008; Liu et al., 2016). Liu et al. (2016) hypothesized that OLFM4 might inhibit Akt phosphorylation and increase GSK-3β-mediated β-catenin degradation, or forming nonfunctional complexes with the Frizzled receptors; the same authors suggested OLFM4 protein loss might contribute to colon-cancer progression after mutations in APC. *OLFM4* has been found upregulated in a subset of

adenomas and early-stage CRCs (stage I/II of the TNM classification) (Sobin et al., 2011), while its expression declines in advanced disease (TNM stages III/IV). *OLFM4* expression in *OLFM4*-positive tumors is significantly higher than that in wild-type crypt base columnar cells, suggesting *OLFM4* is a marker of a subset of colorectal cancer stem cells (van der Flier et al., 2009). Seko et al. (2009) showed that OLFM4 IHC expression is a prognostic marker for survival, with OLFM4-positive CRCs showing better survival than OLFM4-negative CRCs.

DACH1 encodes a transcription factor regulating the transcription of some genes involved in proliferation. Its expression is higher in the lower half of normal colorectal crypts, where proliferative precursors reside (Vonlanthen et al., 2014). Murine Dach1 expression has been found to be enriched in intestinal stem cells (Munoz et al., 2012).

Fatty acid binding protein 1 (*FABP1*) was downregulated in all the three tumor groups (HB, LB, MSI) compared to normal tissues, but MSI samples showed the deepest decrease, so that *FABP1* resulted to be upregulated in both HB and LB CRCs compared to MSI (although its expression was 3.52-fold higher in LB compared to HB samples). FABP1 is an intracellular protein involved in lipid metabolism and transport of long chain fatty acids. In the intestine, FABP1 is expressed by the absorptive enterocytes, but not by crypt cells (Gajda and Storch, 2015). FABP1 expression is decreased in the consensus molecular subtype CMS1 of colorectal carcinomas, which is enriched for MSI tumors (Wood et al., 2017).

5.4.3.5 GENES DIFFERENTIALLY EXPRESSED IN HB AND MSI COMPARED TO LB

AGMAT gene was upregulated in HB and MSI tumors compared to normal tissues and to LB tumors. *AGMAT* encodes agmatinase, an enzyme catalyzing hydrolysis of the polyamine agmatine to putrescine and urea (Mistry et al., 2002). Putrescine can be further converted into polyamines spermidine and spermine. Polyamines are known to contribute to colorectal carcinogenesis, by increasing cell proliferation and inflammation (Babbar and Gerner, 2011). Agmatine is a component of human cells, however humans do not have a functional L-arginine carboxylase enzyme (Coleman et al., 2004), and produce putrescine through decarboxylation of L-ornithine by L-ornithine decarboxylase. Agmatine intake comes from diet, since plants and bacteria (including intestinal microbial flora) can decarboxylate L-arginine to agmatine (Coleman et al., 2004), and colonic bacteria are a source of polyamines (Babbar and Gerner, 2011).

5.4.3.6 GENES DIFFERENTIALLY EXPRESSED IN LB AND MSI COMPARED TO HB

REG4, *SERPINB5*, *MUC5B, CD55* and *SPP1* were upregulated in both MSI and LB CRCs, although upregulation was higher in MSI CRCs.

A gene which was expressed at a similar extent by LB and MSI tumors was Trefoil Factor 1 (*TFF1*). Trefoil factors TFF1, TFF2 and TFF3 are secreted by mucus-producing cells throughout the gastrointestinal tract, with TFF1 and TFF2 showing preferential expression in normal stomach and Brunner glands and TFF3 being preferentially expressed by goblet cells of the small intestine and colon (Aihara et al., 2017; Playford, 1997). TFF peptides show protective effects on gastrointestinal mucosa, contributing to mucus stabilization and stimulating cell migration to cover areas of damaged mucosa, but they can also destabilize cell adhesion by disrupting the E-Cadherin/β-catenin complex, along with playing a role in cell survival signaling (Aihara et al., 2017).

Another gene expressed by both LB and MSI tumors compared to HB ones was *POSTN*, encoding periostin (POSTN), also known as osteoblast-specific factor 2 (OSF-2). POSTN is a matricellular protein belonging to the family of fasciclins (being homologous to Fasciclin I), originally characterized as a protein preferentially expressed in periosteum and in periodontal ligament, its expression being inducible by TGF-β (Horiuchi et al., 1999). POSTN is involved in bone turnover, bone formation and response to mechanical loading (Bonnet et al., 2009). POSTN is indeed expressed by other normal tissues, such as heart valves and colon, while high fluctuations in its expression among tissue samples have been observed for the small intestine, breast, skin, kidney and ovaries (Norris et al., 2007; Tilman et al., 2007). POSTN interacts with collagen type I, fibronectin, Notch1 precursor, Tenascin-C, (Kii et al., 2010; Kudo, 2011; Norris et al., 2007; Tanabe et al., 2010), as well as BMP-1, promoting BMP-1-mediated proteolytic activation of lysyl oxidase (LOX), which then catalyzes covalent collagen cross-linking (Maruhashi et al., 2010). Overexpression of POSTN has been detected across different tumors, including CRC (Bao et al., 2004), breast cancer (Shao et al., 2004), epithelial ovarian cancer (Ismail et al., 2000), non-small cell lung cancer (NSCLC) (Sasaki et al., 2001), pancreatic cancer (Baril et al., 2007) and others (Ishiba et al., 2015). In normal colonic mucosa, POSTN is produced and secreted by pericryptal fibroblasts of colonic crypts, whereas in colonic adenomas and adenocarcinomas POSTN expression by pericryptal fibroblasts decreases and cancer-associated fibroblasts (CAFs) become the main source of POSTN (Kikuchi et al., 2008). In CRC, POSTN can bind to $\alpha_V\beta_3$ integrins, activate the Akt/PKB survival pathway (such effect was not observed for $\alpha_V\beta_5$ integrins) and promote tumor angiogenesis and metastasis (Bao et al., 2004). High immunohistochemical stromal POSTN expression correlates with

aggressive clinicopathological features in CRCs and associates with the BRAF-mutant/CIMP-high/MSS subgroup of tumors arising from the serrated neoplasia pathway (Oh et al., 2017). POSTN is also both a marker and an inductor of EMT (Morra and Moch, 2011).

5.4.4 HISTOPATHOLOGY AND GENOME-INSTABILITY CLASSIFICATION OF CRC TUMORS

In the MSI group, 4/5 tumors (80%) were proximal (right colon) and only 1/5 (20%) was distal (left colon) (Table 11). In the LB group, excluding 3 tumors of the rectum, 6/8 (75%) tumors were proximal (right colon) and 2/8 (25%) were distal (left colon and sigmoid colon) (Table 12). Finally, in the HB group, excluding 3 rectal tumors, 13/26 (50%) CRCs were proximal (right colon) and 13/26 (50%) were distal (left colon and sigmoid colon) (Table 13). Several CRC tumors showed mucinous histology or mucinous differentiation in the pathology reports (Table 11, Table 12, Table 13). It is interesting to note that in the LB group 7/8 patients (87%) showed tumors with mucin production (6 mucinous adenocarcinomas and 1 signet ring cell carcinoma). This was in agreement with the above-mentioned prominent expression of *MUC2*, a colon mucus barrier marker, in the present transcriptomic analysis. In the MSI group, 3/5 patients (60%) showed tumors with mucin production (3 mucinous adenocarcinomas). Indeed, *MUC5B*, encoding a secreted mucin, and *AGR2*, encoding a disulfide isomerase which helps MUC2 and MUC5B folding in the endoplasmic reticulum, were highly expressed in the MSI group, in agreement with the enrichment in mucinous tumors found among MSI samples. On the contrary, in the HB group only 3/20 patients (15%) showed tumors with mucin production (2 mucinous adenocarcinoma and 1 signet ring adenocarcinoma).

6. DISCUSSION

6.1 TARGETED SEQUENCING

As reported in the results section, MSI CRC samples were significantly enriched for somatic variants of the MMR/Polymerases panel compared to MSS CRCs. Ultra-hypermutation is known to be more common among MSS CRCs with mutations in *POLE* exonuclease domain (Kim et al., 2013), and results from a recent study on a large number of CRCs showed *POLE* exonuclease domain mutations to be mutually exclusive with mismatch repair deficiency (Domingo et al., 2016). Nevertheless, *POLE*-exonuclease domain and non-exonuclease domain mutations have been also identified among MSI tumors (Cancer Genome Atlas Network et al., 2012; Kim et al., 2013; Palles et al., 2013) and it has been reported that mutations in MMR genes and in the exonuclease domain of *POLE/POLD1* tend to associate in CRCs with a high mutational burden (Flohr et al., 1999; Jansen et al., 2016; Yoshida et al., 2011). Except for germline *POLE* p.Asn336Ser variant, found in P19_T1 MSS sample, all *POLE*/*POLD1* variants detected in the present thesis were outside the exonuclease domain, and the potential clinical significance of such non-exonuclease domain variants in replicative polymerases is of unclear interpretation (Briggs and Tomlinson, 2013). Moreover, none of the somatic *POLE*/*POLD1* variants identified in the present work was recorded in any of the consulted databases. We could not evaluate the single nucleotide mutation rates, which might act as a proof-of-principle of *POLE*/*POLD1* pathogenicity, because of the small size of the targeted sequencing panels that have been used. For what concerns germline variants in replicative polymerases, apart from *POLE* p.Asn336Ser variant in P19_T1 MSS sample, all the other variants were detected in Lynch syndrome patients: P3_T1 patient showed *POLD1* p.Glu673Lys, located in a conserved domain but with uncertain functional significance according to ClinVar. P29_T1 showed *POLD1* p.Val759Ile and *POLE* p.Phe695Ile (Table 7 and Supplementary Table 3), of which the former has been classified as likely benign by most ClinVar submitters and the latter is above the polymorphism threshold in several populations, including the ethnic group closest to our Sicilian population (minor allele frequency = 0.023 in Tuscanian population, data from the 1000 Genomes Project Phase 3).

Of note, P13_T1 showed a double somatic hit in *MLH1* gene, with a *MLH1* p.Arg487* premature stop codon mutation located at a region of copy-neutral loss of heterozygosity on chromosome 3. The same sample presented additional MMR variants: an InSiGHT Class 3 *MSH6* p.Arg361His (19% of reads) and a

MSH3 p.Asp185Val variant (18% of the reads). Double somatic hits are reported to explain microsatellite instability in up to 27.8% of MSI tumors (Mensenkamp et al., 2014; Sourrouille et al., 2013).

Moreover, HTA data showed a significant downregulation of the *MLH1* gene in the MSI group compared to normal tissues and to HB and LB tumors, in accordance with the frequent *MLH1* promoter methylation observed among MSI tumors (Deng et al., 1999; Herman et al., 1998).

In contrast to what observed for the MMR/Polymerases genes, the number of somatic variants in the CHS Panel genes, which are common oncogenes and tumor suppressors, did not significantly differ between MSI tumors and MSS tumors (Figure 6).

Lynch syndrome tumors showed a lower number of BCNAs (0 for P3_T1 and 1 for P29_T1) compared to the majority of other MSI tumors. Of note, among MSS tumors, there were three samples (PCC3_T1, P73_T1, P65_T1) with no BCNAs and no variants in the CHS genes. P73_T1 did not either show any sequence variant in the MMR/Polymerases panel; P65_T1 had a germline benign *PMS2* p.Thr597Ser variant; PCC3_T1 had a *MLH3* p.Ile19Leu which should not be functionally relevant with regard to the MMR system, since the tumor is MSS, and a *POLE* p.His1551Pro variant outside the exonuclease domain. Thereby, since these MSS tumors have a low number of BCNAs and are MSS, they do not show any of the forms of genomic instability investigated in this work; additionally, no relevant sequence variants were identified in the examined genes. Exome sequencing might be useful for a further understanding of the genetic alterations in such tumors.

6.2 GENE EXPRESSION PROFILE

HB tumors showed strong expression of epithelial markers, such as *NOX1*, which contributes to oxidative DNA damage and ROS-driven tumor growth, as well as activation of Wnt/β-catenin and Notch signaling pathways (Coant et al., 2010; Juhasz et al., 2017; Lambeth et al., 1999). Epithelial markers upregulated in HB tumors also include the epidermal growth factor (EGF) family members amphiregulin (*AREG*) and epiregulin (*EREG*), which are often overexpressed in CRC, especially by CIN tumors of the distal colon (Ciardiello et al., 1991; Lee et al., 2016a; Missiaglia et al., 2014), and *AXIN2*, a critical regulator of Wnt/β-catenin pathway, which is considered an indicator of Wnt pathway activation (Lustig et al., 2002; Thorvaldsen et al., 2017).

OLFM4, a marker of bottom-of-the-crypt cells (van der Flier et al., 2009; Vonlanthen et al., 2014) was overexpressed in HB and LB tumors compared to MSI tumors. Moreover, *OLFM4* was significantly

downregulated in MSI tumors compared to normal tissues. On the contrary, *CD26*, a marker of a subset of colorectal cancer stem cells (Pang et al., 2010), was overexpressed in the MSI group compared to HB and LB tumors.

Both MSI and LB tumors overexpressed *REG4*, although expression was higher in the MSI group. *REG4* is a marker of the deep crypt secretory (DCS) cells, which contribute to the stem cell niche at the bottom of colon crypts (Sasaki et al., 2016). DCS cells are mucous-type cells that can be found intercalated with $LGR5^+$ CBC stem cells at the colon crypt base (Altmann, 1983). DCS cells express EGF and Notch ligands Dll1 and Dll4 – but not Wnt ligand – and are thought to play a role as Paneth cell equivalents contributing to the colon crypt stem cell niche (Sasaki et al., 2016). Rothenberg et al. (2012) found that DCS cells, in addition to *REG4*, also overexpress *AGR2* and *LGR5*. Indeed, MSI tumors and - to a lesser extent - LB tumors, showed upregulation of *AGR2*. Furthermore, DCS cells are known to be KIT^+. *KIT* was not differentially expressed in the MSI cohort of the present work, but this was the case for KIT Ligand (*KITLG*). Interestingly, KITLG can upregulate expression of *PLA2G4A* (Murakami et al., 1995), which was indeed highly expressed in MSI tumors compared to the other two tumor groups and to normal tissues.

Expression of *TFF1* was high in both MSI and LB samples of our cohort, compared to HB tumors and to normal tissues.

Moreover, MSI and – to a lesser extent – LB samples, showed increased expression of *MUC5B*, a gel-forming mucin expressed in the colon by a subset of goblet cells at the crypt base, but not by mature enterocytes (van Klinken et al., 1998; Larsson et al., 2011).

LB tumors showed upregulation of *LYZ* and *MUC2*. Apart from immune cells, Paneth cells of the small intestine also produce lysozyme. However, Paneth cells are absent from the colon (Clevers, 2013; Sasaki et al., 2016; Sato et al., 2011). *MUC2* – encoding the principal component of intestinal mucus - is a marker of goblet cells (Grün et al., 2015).

By looking at tumor pathology, we noticed an enrichment for mucin-producing tumors in the LB and MSI groups. Mucinous adenocarcinomas and signet ring-cell adenocarcinomas are sometimes grouped together when considering colorectal tumors with mucin production, because of the low prevalence of signet-ring cell tumors of the colon and of the possible occurrence of tumors with mixed mucinous and signet-ring cell features (Pande et al., 2008; Wei et al., 2016). In the present study, tumors with intracellular or extracellular mucin production together represented 87.5% of the LB samples, 60% of the MSI tumors and 15% of the HB ones. Both the MSI and the LB group were enriched for mucin-producing tumors.

Expression of colon mucus barrier marker MUC2 is usually negative or strongly reduced in CRC, except for mucinous CRCs and signet-ring cell tumors, which preserve MUC2 expression (Gratchev et al., 2001; Iwase et al., 2005; Walsh et al., 2013). Colon mucus barrier marker MUC5B also associates with mucinous histology (Walsh et al., 2013). Both MUC2 and MUC5B expression are reported to associate with MSI tumors (Perez-Villamil et al., 2012; Walsh et al., 2013), but the number of MSI samples analyzed in the present thesis (5 samples from 5 patients) was probably low to detect an association between MUC2 and MSI status. A study by Perez-Villamil et al. (2012) reported mucinous CRCs to cluster together with MSI tumors and to be characterized by overexpression of *REG4*, *SPP1* and *CD55*, among others. Indeed, *REG4, SPP1* and *CD55* were overexpressed in both the MSI and – to lesser extent - the LB group.

Hugen et al. (2015), by analyzing their own patient cohort and TCGA level 3 SNP6 data, found a reduced rate of copy number aberrations in mucinous CRCs. The same authors also noticed that, among mucinous CRCs, those with a high rate of copy number aberrations had a poorer prognosis. Indeed, in the present thesis, both the MSI and LB groups, which are enriched for mucin-producing tumors, were characterized by a low number of BCNAs.

According to Nitsche et al. (2013), there are no significant differences in survival found between colon adenocarcinomas and mucinous adenocarcinomas, whereas signet-ring cell histology associates with worse survival; among mucinous or signet-ring cell CRCs, those which are microsatellite-stable show a more aggressive behavior.

In conclusion, classifying MSS CRCs on the basis of the number of broad copy number abnormalities, allowed the identification of two distinct groups: on the one hand, HB tumors (with high numbers of BCNAs), showing upregulation of epithelial markers including EGF agonists *AREG* and *EREG*, whose overexpression could confer sensitivity to EGFR blockade agents (Khambata-Ford et al., 2007; Schütte et al., 2017); on the other side, LB tumors (with a low number of BCNAs), showing upregulation of a subset of genes which are also overexpressed – to a greater extent – by MSI tumors: *REG4*, *AGR2*, *SPP1*, *CD55, MUC5B*. Such shared gene expression profile encompassed both markers associated to mucinous/signet-ring cell histology and markers of DCS cells. Indeed, LB and MSI groups were enriched for mucin-producing tumors (87.5% and 60%, respectively).

The number of BCNAs, which can be evaluated by the use of genome-wide techniques, could be helpful for classifying CRC samples along with MSI testing, detection of sequence variants and epigenetic modifications, and provide information of potentially predictive and/or prognostic value.

7. REFERENCES

Aaltonen, L.A., Peltomäki, P., Leach, F.S., Sistonen, P., Pylkkänen, L., Mecklin, J.P., Järvinen, H., Powell, S.M., Jen, J., and Hamilton, S.R. (1993). Clues to the pathogenesis of familial colorectal cancer. Science *260*, 812–816.

Aaltonen, L.A., Salovaara, R., Kristo, P., Canzian, F., Hemminki, A., Peltomäki, P., Chadwick, R.B., Kääriäinen, H., Eskelinen, M., Järvinen, H., et al. (1998). Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. N. Engl. J. Med. *338*, 1481–1487.

Adams, J.C., and Tucker, R.P. (2000). The thrombospondin type 1 repeat (TSR) superfamily: diverse proteins with related roles in neuronal development. Dev. Dyn. *218*, 280–299.

Agnel, M., Vermat, T., and Culouscou, J.M. (1999). Identification of three novel members of the calcium-dependent chloride channel (CaCC) family predominantly expressed in the digestive tract and trachea. FEBS Lett. *455*, 295–301.

Aihara, E., Engevik, K.A., and Montrose, M.H. (2017). Trefoil Factor Peptides and Gastrointestinal Function. Annu. Rev. Physiol. *79*, 357–380.

Albert, T.K., Laubinger, W., Müller, S., Hanisch, F.-G., Kalinski, T., Meyer, F., and Hoffmann, W. (2010). Human Intestinal TFF3 Forms Disulfide-Linked Heteromers with the Mucus-Associated FCGBP Protein and Is Released by Hydrogen Sulfide. J. Proteome Res. *9*, 3108–3117.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). Molecular biology of the cell (Garland Science).

Alper, S.L., Stewart, A.K., Vandorpe, D.H., Clark, J.S., Zachary Horack, R., Simpson, J.E., Walker, N.M., and Clarke, L.L. (2011). Native and recombinant Slc26a3 (downregulated in adenoma, Dra) do not exhibit properties of 2Cl-/1HCOFormula exchange. AJP Cell Physiol. *300*, C276–C286.

Altmann, G.G. (1983). Morphological observations on mucus-secreting nongoblet cells in the deep crypts of the rat ascending colon. Am. J. Anat. *167*, 95–117.

Amanuel, B., Grieu, F., Kular, J., Millward, M., and Iacopetta, B. (2012). Incidence of BRAF p.Val600Glu and p.Val600Lys mutations in a consecutive series of 183 metastatic melanoma patients from a high incidence region. Pathology *44*, 357–359.

Ambort, D., Johansson, M.E. V., Gustafsson, J.K., Nilsson, H.E., Ermund, A., Johansson, B.R., Koeck, P.J.B., Hebert, H., and Hansson, G.C. (2012). Calcium and pH-dependent packing and release of the gel-forming MUC2 mucin. Proc. Natl. Acad. Sci. *109*, 5645–5650.

Aoki, T., and Narumiya, S. (2017). Prostaglandin E 2 -EP2 signaling as a node of chronic inflammation in the colon tumor microenvironment Inflammation and Regeneration. Inflamm. Regen. *37*.

Aoude, L.G., Heitzer, E., Johansson, P., Gartside, M., Wadt, K., Pritchard, A.L., Palmer, J.M., Symmons, J., Gerdes, A.-M., Montgomery, G.W., et al. (2015). POLE mutations in families predisposed to cutaneous melanoma. Fam. Cancer *14*, 621–628.

Babbar, N., and Gerner, E.W. (2011). Targeting polyamines and inflammation for cancer prevention. Recent Results Cancer Res. *188*, 49–64.

Bader, A.G., Kang, S., Zhao, L., and Vogt, P.K. (2005). Oncogenic PI3K deregulates transcription and translation. Nat. Rev. Cancer *5*, 921–929.

Bao, S., Ouyang, G., Bai, X., Huang, Z., Ma, C., Liu, M., Shao, R., Anderson, R.M., Rich, J.N., and Wang, X.-F. (2004). Periostin potently promotes metastatic growth of colon cancer by augmenting cell survival via the Akt/PKB pathway. Cancer Cell *5*, 329–339.

Baril, P., Gangeswaran, R., Mahon, P.C., Caulee, K., Kocher, H.M., Harada, T., Zhu, M., Kalthoff, H., Crnogorac-Jurcevic, T., and Lemoine, N.R. (2007). Periostin promotes invasiveness and resistance of pancreatic cancer cells to hypoxia-induced cell death: role of the β4 integrin and the PI3k pathway. Oncogene *26*, 2082–2094.

Barker, N. (2014). Adult intestinal stem cells: critical drivers of epithelial homeostasis and regeneration. Nat. Rev. Mol. Cell Biol. *15*, 19–33.

Barresi, V., Castorina, S., Musso, N., Capizzi, C., Luca, T., Privitera, G., and Condorelli, D.F. (2017). Chromosomal instability analysis and regional tumor heterogeneity in colon cancer. Cancer Genet. *210*, 9–21.

Bellahcène, A., Castronovo, V., Ogbureke, K.U.E., Fisher, L.W., and Fedarko, N.S. (2008). Small integrin-binding ligand N-linked glycoproteins (SIBLINGs): multifunctional proteins in cancer. Nat. Rev. Cancer *8*, 212–226.

Bellido, F., Pineda, M., Aiza, G., Valdés-Mas, R., Navarro, M., Puente, D.A., Pons, T., González, S., Iglesias, S., Darder, E., et al. (2016). POLE and POLD1 mutations in 529 kindred with familial colorectal cancer and/or polyposis: review of reported cases and recommendations for genetic testing and surveillance. Genet. Med. *18*, 325–332.

Bergstrom, J.H., Berg, K.A., Rodriguez-Pineiro, A.M., Stecher, B., Johansson, M.E. V., and Hansson, G.C. (2014). AGR2, an Endoplasmic Reticulum Protein, Is Secreted into the Gastrointestinal Mucus. PLoS One *9*, e104186.

Beroukhim, R., Getz, G., Nghiemphu, L., Barretina, J., Hsueh, T., Linhart, D., Vivanco, I., Lee, J.C., Huang, J.H., Alexander, S., et al. (2007). Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. Proc. Natl. Acad. Sci. U. S. A. *104*, 20007–20012.

Bettington, M., Walker, N., Clouston, A., Brown, I., Leggett, B., and Whitehall, V. (2013). The serrated pathway to colorectal carcinoma: current concepts and challenges. Histopathology *62*, 367–386.

Bettstetter, M., Woenckhaus, M., Wild, P.J., Rümmele, P., Blaszyk, H., Hartmann, A., Hofstädter, F., and Dietmaier, W. (2005). Elevated nuclear maspin expression is associated with microsatellite instability and high tumour grade in colorectal cancer. J. Pathol. *205*, 606–614.

Beumer, J., and Clevers, H. (2016). Regulation and plasticity of intestinal stem cells during homeostasis and regeneration. Development *143*, 3639–3649.

Birchenough, G.M.H., Johansson, M.E. V, Gustafsson, J.K., Bergström, J.H., and Hansson, G.C. (2015). New developments in goblet cell mucus secretion and function. Mucosal Immunol. *8*, 712–719.

Bishnupuri, K.S., Sainathan, S.K., Bishnupuri, K., Leahy, D.R., Luo, Q., Anant, S., Houchen, C.W., and Dieckgraefe, B.K. (2014). Reg4-induced mitogenesis involves Akt-GSK3?-?-Catenin-TCF-4 signaling in human colorectal cancer. Mol. Carcinog. *53*, E169–E180.

de Bock, C., Ardjmand, A., Molloy, T., Bone, S., Johnstone, D., Campbell, D., Shipman, K., Yeadon, T., Holst, J., Nelmes, G., et al. (2011). The Fat1 cadherin is overexpressed and an independent prognostic factor for survival in paired diagnosis–relapse samples of precursor B-cell acute lymphoblastic leukemia. Leukemia *26*, 918–926.

Boland, C.R., Thibodeau, S.N., Hamilton, S.R., Sidransky, D., Eshleman, J.R., Burt, R.W., Meltzer, S.J., Rodriguez-Bigas, M.A., Fodde, R., Ranzani, G.N., et al. (1998). A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res. *58*, 5248–5257.

Bonnet, N., Standley, K.N., Bianchi, E.N., Stadelmann, V., Foti, M., Conway, S.J., and Ferrari, S.L. (2009). The matricellular protein periostin is required for sost inhibition and the anabolic response to mechanical loading and physical activity. J. Biol. Chem. *284*, 35939–35950.

Briggs, S., and Tomlinson, I. (2013). Germline and somatic polymerase ε and δ mutations define a new class of hypermutated colorectal and endometrial cancers. J. Pathol. *230*, 148–153.

Brunelli, L., Caiola, E., Marabese, M., Broggini, M., and Pastorelli, R. (2014). Capturing the metabolomic diversity of KRAS mutants in non-small-cell lung cancer cells. Oncotarget *5*, 4722–4731.

Buczacki, S.J.A., Zecchini, H.I., Nicholson, A.M., Russell, R., Vermeulen, L., Kemp, R., and Winton, D.J. (2013). Intestinal label-retaining cells are secretory precursors expressing Lgr5. Nature *495*, 65–69.

Burstyn-Cohen, T., Frumkin, A., Xu, Y.T., Scherer, S.S., and Klar, A. (1998). Accumulation of F-spondin in injured peripheral nerve promotes the outgrowth of sensory axons. J. Neurosci. *18*, 8875–8885.

Cancer Genome Atlas Network, Muzny, D.M., Bainbridge, M.N., Chang, K., Dinh, H.H., Drummond, J.A., Fowler, G., Kovar, C.L., Lewis, L.R., Morgan, M.B., et al. (2012). Comprehensive molecular characterization of human colon and rectal cancer. Nature *487*, 330–337.

Cancer Genome Atlas Network, Getz, G., Gabriel, S.B., Cibulskis, K., Lander, E., Sivachenko, A., Sougnez, C., Lawrence, M., Kandoth, C., Dooling, D., et al. (2013). Integrated genomic characterization of endometrial carcinoma. Nature *497*, 67–73.

Cheng, Y.-W., Pincas, H., Bacolod, M.D., Schemmann, G., Giardina, S.F., Huang, J., Barral, S., Idrees, K., Khan, S.A., Zeng, Z., et al. (2008). CpG island methylator phenotype associates with low-degree chromosomal abnormalities in colorectal cancer. Clin. Cancer Res. *14*, 6005–6013.

Chiera, F., Meccia, E., Degan, P., Aquilina, G., Pietraforte, D., Minetti, M., Lambeth, D., and Bignami, M. (2008). Overexpression of human NOX1 complex induces genome instability in mammalian cells. Free Radic. Biol. Med. *44*, 332–342.

Chin, K.L., Aerbajinai, W., Zhu, J., Drew, L., Chen, L., Liu, W., and Rodgers, G.P. (2008). The regulation of *OLFM4* expression in myeloid precursor cells relies on NF-κB transcription factor. Br. J. Haematol. *143*, 421–432.

Ciardiello, F., Kim, N., Saeki, T., Dono, R., Persico, M.G., Plowman, G.D., Garrigues, J., Radke, S., Todaro, G.J., and Salomon, D.S. (1991). Differential expression of epidermal growth factor-related proteins in human colorectal tumors. Proc. Natl. Acad. Sci. U. S. A. *88*, 7792–7796.

Clevers, H. (2013). The Intestinal Crypt, A Prototype Stem Cell Compartment. Cell *154*, 274–284.

Coant, N., Ben Mkaddem, S., Pedruzzi, E., Guichard, C., Tréton, X., Ducroc, R., Freund, J.-N., Cazals-Hatem, D., Bouhnik, Y., Woerther, P.-L., et al. (2010). NADPH Oxidase 1 Modulates WNT and NOTCH1 Signaling To Control the Fate of Proliferative Progenitor Cells in the Colon. Mol. Cell. Biol. *30*, 2636–2650.

Coleman, C.S., Hu, G., and Pegg, A.E. (2004). Putrescine biosynthesis in mammalian tissues. Biochem. J. *379*, 849–855.

Crawford, I., Maloney, P.C., Zeitlin, P.L., Guggino, W.B., Hyde, S.C., Turley, H., Gatter, K.C., Harris, A., and Higgins, C.F. (1991). Immunocytochemical localization of the cystic fibrosis gene product CFTR. Proc. Natl. Acad. Sci. U. S. A. *88*, 9262–9266.

Curtis, L.J., Georgiades, I.B., White, S., Bird, C.C., Harrison, D.J., and Wyllie, A.H. (2000). Specific patterns of chromosomal abnormalities are associated with RER status in sporadic colorectal cancer. J. Pathol. *192*, 440–445.

Dalerba, P., Kalisky, T., Sahoo, D., Rajendran, P.S., Rothenberg, M.E., Leyrat, A.A., Sim, S., Okamoto, J., Johnston, D.M., Qian, D., et al. (2011). Single-cell dissection of transcriptional heterogeneity in human colon tumors. Nat. Biotechnol. *29*, 1120–1127.

Dallol, A., Buhmeida, A., Al-Ahwal, M.S., Al-Maghrabi, J., Bajouh, O., Al-Khayyat, S., Alam, R., Abusanad, A., Turki, R., Elaimi, A., et al. (2016). Clinical significance of frequent somatic mutations detected by high-throughput targeted sequencing in archived colorectal cancer samples. J. Transl. Med. *14*, 118.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes Dev. *25*, 1010–1022.

Deng, G., Chen, A., Hong, J., Chae, H.S., and Kim, Y.S. (1999). Methylation of CpG in a small region of the hMLH1 promoter invariably correlates with the absence of gene expression. Cancer Res. *59*, 2029–2033.

Dienstmann, R., Vermeulen, L., Guinney, J., Kopetz, S., Tejpar, S., and Tabernero, J. (2017). Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. Nat. Rev. Cancer *17*, 79–92.

Dieumegard, B., Grandjouan, S., Sabourin, J.-C., Bihan, M.-L. Le, Lefrère, I., Bellefqih, Pignon, J.-P., Rougier, P., Lasser, P., Bénard, J., et al. (2000). Extensive molecular screening for hereditary non-polyposis colorectal cancer. Br. J. Cancer *82*, 871–880.

Dinjens, W.N., ten Kate, J., van der Linden, E.P., Wijnen, J.T., Khan, P.M., and Bosman, F.T. (1989). Distribution of adenosine deaminase complexing protein (ADCP) in human tissues. J. Histochem. Cytochem. *37*, 1869–1875.

Domingo, E., Freeman-Mills, L., Rayner, E., Glaire, M., Briggs, S., Vermeulen, L., Fessler, E., Medema, J.P., Boot, A., Morreau, H., et al. (2016). Somatic POLE proofreading domain mutation, immune response, and prognosis in colorectal cancer: a retrospective, pooled biomarker study. Lancet. Gastroenterol. Hepatol. *1*, 207–216.

Doublié, S., and Zahn, K.E. (2014). Structural insights into eukaryotic DNA replication. Front. Microbiol. *5*, 444.

Durrant, L.G., Chapman, M.A., Buckley, D.J., Spendlove, I., Robins, R.A., and Armitage, N.C. (2003). Enhanced expression of the complement regulatory protein CD55 predicts a poor prognosis in colorectal cancer patients. Cancer Immunol. Immunother. *52*, 638–642.

Elsayed, F.A., Kets, C.M., Ruano, D., van den Akker, B., Mensenkamp, A.R., Schrumpf, M., Nielsen, M., Wijnen, J.T., Tops, C.M., Ligtenberg, M.J., et al. (2015). Germline variants in POLE are associated with early onset mismatch repair deficient colorectal

cancer. Eur. J. Hum. Genet. *23*, 1080–1084.

Erson-Omay, E.Z., Çağlayan, A.O., Schultz, N., Weinhold, N., Omay, S.B., Özduman, K., Köksal, Y., Li, J., Serin Harmancı, A., Clark, V., et al. (2015). Somatic POLE mutations cause an ultramutated giant cell high-grade glioma subtype with better prognosis. Neuro. Oncol. *17*, 1356–1364.

Faderl, M., Noti, M., Corazza, N., and Mueller, C. (2015). Keeping bugs in check: The mucus layer as a critical component in maintaining intestinal homeostasis. IUBMB Life *67*, 275–285.

Fleming, M., Ravula, S., Tatishchev, S.F., and Wang, H.L. (2012). Colorectal carcinoma: Pathologic aspects. J. Gastrointest. Oncol. *3*, 153–173.

van der Flier, L.G., Haegebarth, A., Stange, D.E., van de Wetering, M., and Clevers, H. (2009). OLFM4 Is a Robust Marker for Stem Cells in Human Intestine and Marks a Subset of Colorectal Cancer Cells. Gastroenterology *137*, 15–17.

Flohr, T., Dai, J.C., Büttner, J., Popanda, O., Hagmüller, E., and Thielmann, H.W. (1999). Detection of mutations in the DNA polymerase delta gene of human sporadic colorectal cancers and colon cancer cell lines. Int. J. Cancer *80*, 919–929.

Forbes, S.A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., Cole, C.G., Ward, S., Dawson, E., Ponting, L., et al. (2017). COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. *45*, D777–D783.

Frängsmyr, L., Baranov, V., and Hammarström, S. (1999). Four carcinoembryonic antigen subfamily members, CEA, NCA, BGP and CGM2, selectively expressed in the normal human colonic epithelium, are integral components of the fuzzy coat. Tumour Biol. *20*, 277–292.

Frizzell, R.A., and Hanrahan, J.W. (2012). Physiology of epithelial chloride and fluid secretion. Cold Spring Harb. Perspect. Med. *2*, a009563.

Fukuyama, M., Rokutan, K., Sano, T., Miyake, H., Shimada, M., and Tashiro, S. (2005). Overexpression of a novel superoxide-producing enzyme, NADPH oxidase 1, in adenoma and well differentiated adenocarcinoma of the human colon. Cancer Lett. *221*, 97–104.

Gajda, A.M., and Storch, J. (2015). Enterocyte fatty acid-binding proteins (FABPs): different functions of liver and intestinal FABPs in the intestine. Prostaglandins. Leukot. Essent. Fatty Acids *93*, 9–16.

Garzon-Muvdi, T., Schiapparelli, P., ap Rhys, C., Guerrero-Cazares, H., Smith, C., Kim, D.-H., Kone, L., Farber, H., Lee, D.Y., An, S.S., et al. (2012). Regulation of brain tumor dispersal by NKCC1 through a novel role in focal adhesion regulation. PLoS Biol. *10*, e1001320.

Geigl, J.B., Obenauf, A.C., Schwarzbraun, T., and Speicher, M.R. (2008). Defining “chromosomal instability”. Trends Genet. *24*, 64–69.

Genschel, J., and Modrich, P. (2009). Functions of MutLalpha, replication protein A (RPA), and HMGB1 in 5’-directed mismatch repair. J. Biol. Chem. *284*, 21536–21544.

Ghersi, G., Zhao, Q., Salamone, M., Yeh, Y., Zucker, S., and Chen, W.-T. (2006). The protease complex consisting of dipeptidyl peptidase IV and seprase plays a role in the migration and invasion of human endothelial cells in collagenous matrices. Cancer Res. *66*, 4652–4661.

Glowacka, W.K., Alberts, P., Ouchida, R., Wang, J.-Y., and Rotin, D. (2012). LAPTM5 protein is a positive regulator of proinflammatory signaling pathways in macrophages. J. Biol. Chem. *287*, 27691–27702.

Goel, A., Nagasaka, T., Arnold, C.N., Inoue, T., Hamilton, C., Niedzwiecki, D., Compton, C., Mayer, R.J., Goldberg, R., Bertagnolli, M.M., et al. (2007). The CpG island methylator phenotype and chromosomal instability are inversely correlated in sporadic colorectal cancer. Gastroenterology *132*, 127–138.

Goodison, S., Urquidi, V., and Tarin, D. (1999). CD44 cell adhesion molecules. Mol. Pathol. *52*, 189–196.

Gratchev, A., Siedow, A., Bumke-Vogt, C., Hummel, M., Foss, H.D., Hanski, M.L., Kobalz, U., Mann, B., Lammert, H., Mansmann, U., et al. (2001). Regulation of the intestinal mucin MUC2 gene expression in vivo: evidence for the role of promoter methylation. Cancer Lett. *168*, 71–80.

Groden, J., Thliveris, A., Samowitz, W., Carlson, M., Gelbert, L., Albertsen, H., Joslyn, G., Stevens, J., Spirio, L., and Robertson, M. (1991). Identification and characterization of the familial adenomatous polyposis coli gene. Cell *66*, 589–600.

Grün, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H., and van Oudenaarden, A. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature *525*, 251–255.

Guinney, J., Dienstmann, R., Wang, X., de Reyni?s, A., Schlicker, A., Soneson, C., Marisa, L., Roepman, P., Nyamundanda, G., Angelino, P., et al. (2015). The consensus molecular subtypes of colorectal cancer. Nat. Med. *21*, 1350–1356.

Gustafsson, J.K., Ermund, A., Ambort, D., Johansson, M.E.V., Nilsson, H.E., Thorell, K., Hebert, H., Sjövall, H., and Hansson, G.C. (2012). Bicarbonate and functional CFTR channel are required for proper mucin secretion and link cystic fibrosis with its mucus phenotype. J. Exp. Med. *209*, 1263–1272.

Hampel, H., Frankel, W.L., Martin, E., Arnold, M., Khanduja, K., Kuebler, P., Nakagawa, H., Sotamaa, K., Prior, T.W., Westman, J., et al. (2005). Screening for the Lynch Syndrome (Hereditary Nonpolyposis Colorectal Cancer). N. Engl. J. Med. *352*, 1851–1860.

Hansen, M.F., Johansen, J., Bjørnevoll, I., Sylvander, A.E., Steinsbekk, K.S., Sætrom, P., Sandvik, A.K., Drabløs, F., and Sjursen, W. (2015). A novel POLE mutation associated with cancers of colon, pancreas, ovaries and small intestine. Fam. Cancer *14*, 437–448.

Hartupee, J.C., Zhang, H., Bonaldo, M.F., Soares, M.B., and Dieckgraefe, B.K. (2001). Isolation and characterization of a cDNA encoding a novel member of the human regenerating protein family: Reg IV. Biochim. Biophys. Acta *1518*, 287–293.

Hauri, H.P., Kappeler, F., Andersson, H., and Appenzeller, C. (2000). ERGIC-53 and traffic in the secretory pathway. J. Cell Sci. *113 (Pt 4)*, 587–596.

Heinimann, K. (2013). Toward a molecular classification of colorectal cancer: the role of microsatellite instability status. Front. Oncol. *3*, 272.

Herman, J.G., Umar, A., Polyak, K., Graff, J.R., Ahuja, N., Issa, J.P., Markowitz, S., Willson, J.K., Hamilton, S.R., Kinzler, K.W., et al. (1998). Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. Proc. Natl. Acad. Sci. U. S. A. *95*, 6870–6875.

Hinoue, T., Weisenberger, D.J., Lange, C.P.E., Shen, H., Byun, H.-M., Van Den Berg, D., Malik, S., Pan, F., Noushmehr, H., van Dijk, C.M., et al. (2012). Genome-scale analysis of aberrant DNA methylation in colorectal cancer. Genome Res. *22*, 271–282.

Höglund, P., Haila, S., Socha, J., Tomaszewski, L., Saarialho-Kere, U., Karjalainen-Lindsberg, M.-L., Airola, K., Holmberg, C., de la Chapelle, A., and Kere, J. (1996). Mutations of the Down–regulated in adenoma (DRA) gene cause congenital chloride diarrhoea. Nat. Genet. *14*, 316–319.

Hong, K.H., Bonventre, J.C., O'Leary, E., Bonventre, J. V, and Lander, E.S. (2001). Deletion of cytosolic phospholipase A(2) suppresses Apc(Min)-induced tumorigenesis. Proc. Natl. Acad. Sci. U. S. A. *98*, 3935–3939.

Horiuchi, K., Amizuka, N., Takeshita, S., Takamatsu, H., Katsuura, M., Ozawa, H., Toyama, Y., Bonewald, L.F., and Kudo, A. (1999). Identification and Characterization of a Novel Protein, Periostin, with Restricted Expression to Periosteum and Periodontal Ligament and Increased Expression by Transforming Growth Factor β. J. Bone Miner. Res. *14*, 1239–1249.

Hu, J., Tian, G., and Zhang, N. (2011). Cytosolic phospholipase A2 and its role in cancer. Clin. Oncol. Cancer Res. *8*, 71–76.

Huang, J., Pan, C., Hu, H., Zheng, S., and Ding, L. (2012). Osteopontin-Enhanced Hepatic Metastasis of Colorectal Cancer Cells. PLoS One *7*, e47901.

Huang, Q., Chen, X., Lu, W., Lai, M., and Lu, B. (2014). Expression of REG4 in Ovarian Mucinous Tumors. Appl. Immunohistochem. Mol. Morphol. *22*, 295–301.

Huang, R., Quan, Y., Huang, A., and Min, Z. (2016). Role of Osteopontin in the Carcinogenesis and Metastasis of Colorectal Cancer. J. Cancer Ther. *7*, 729–740.

Hugen, N., Simmer, F., Mekenkamp, L.J.M., Koopman, M., den Broek, E. van, de Wilt, J.H.W., Punt, C.J.A., Ylstra, B., Meijer, G.A., and Nagtegaal, I.D. (2015). Reduced rate of copy number aberrations in mucinous colorectal carcinoma. Oncotarget *6*, 25715–25725.

Igal, R.A. (2010). Stearoyl-CoA desaturase-1: a novel key player in the mechanisms of cell proliferation, programmed cell death and transformation to cancer. Carcinogenesis *31*, 1509–1515.

Ishiba, T., Nagahara, M., Nakagawa, T., Sato, T., Ishikawa, T., Uetake, H., Sugihara, K., Miki, Y., and Nakanishi, A. (2015). Periostin suppression induces decorin secretion leading to reduced breast cancer cell motility and invasion. Sci. Rep. *4*, 7069.

Ismail, R.S., Baldwin, R.L., Fang, J., Browning, D., Karlan, B.Y., Gasson, J.C., and Chang, D.D. (2000). Differential gene expression between normal and tumor-derived ovarian epithelial cells. Cancer Res. *60*, 6744–6749.

Iwase, T., Kushima, R., Mukaisho, K., Mitsufuji, S., Okanoue, T., and Hattori, T. (2005). Overexpression of CD10 and reduced MUC2 expression correlate with the development and progression of colorectal neoplasms. Pathol. - Res. Pract. *201*, 83–91.

Jacob, P., Rossmann, H., Lamprecht, G., Kretz, A., Neff, C., Lin-Wu, E., Gregor, M., Groneberg, D.A., Kere, J., and Seidler, U. (2002). Down-regulated in adenoma mediates apical Cl-/HCO3- exchange in rabbit, rat, and human duodenum. Gastroenterology *122*, 709–724.

Jansen, A.M., van Wezel, T., van den Akker, B.E., Ventayol Garcia, M., Ruano, D., Tops, C.M., Wagner, A., Letteboer, T.G., Gómez-García, E.B., Devilee, P., et al. (2016). Combined mismatch repair and POLE/POLD1 defects explain unresolved suspected Lynch syndrome cancers. Eur. J. Hum. Genet. *24*, 1089–1092.

Jass, J.R. (2007). Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. Histopathology *50*, 113–130.

Jiricny, J. (2013). Postreplicative Mismatch Repair. Cold Spring Harb. Perspect. Biol. *5*, a012633–a012633.

Johansson, M.E. V., and Hansson, G.C. (2016). Immunological aspects of intestinal mucus and mucins. Nat. Rev. Immunol. *16*, 639–649.

Johansson, M.E. V., Thomsson, K.A., and Hansson, G.C. (2009). Proteomic Analyses of the Two Mucus Layers of the Colon Barrier Reveal That Their Main Component, the Muc2 Mucin, Is Strongly Bound to the Fcgbp Protein. J. Proteome Res. *8*, 3549–3557.

Johansson, M.E. V., Larsson, J.M.H., and Hansson, G.C. (2011). The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. Proc. Natl. Acad. Sci. *108*, 4659–4665.

Johnson, R.E., Klassen, R., Prakash, L., and Prakash, S. (2015). A Major Role of DNA Polymerase δ in Replication of Both the Leading and Lagging DNA Strands. Mol. Cell *59*, 163–175.

Juhasz, A., Markel, S., Gaur, S., Liu, H., Lu, J., Jiang, G., Wu, X., Antony, S., Wu, Y., Melillo, G., et al. (2017). NADPH oxidase 1 supports proliferation of colon cancer cells by modulating reactive oxygen species-dependent signal transduction. J. Biol. Chem. *292*, 7866–7887.

Kang, W., and Reid, K.B.. (2003). DMBT1, a regulator of mucosal homeostasis through the linking of mucosal defense and regeneration? FEBS Lett. *540*, 21–25.

Kaprio, T., Hagström, J., Mustonen, H., Koskensalo, S., Andersson, L.C., and Haglund, C. (2014). REG4 independently predicts better prognosis in non-mucinous colorectal cancer. PLoS One *9*, e109600.

Khambata-Ford, S., Garrett, C.R., Meropol, N.J., Basik, M., Harbison, C.T., Wu, S., Wong, T.W., Huang, X., Takimoto, C.H., Godwin, A.K., et al. (2007). Expression of Epiregulin and Amphiregulin and *K-ras* Mutation Status Predict Disease Control in Metastatic Colorectal Cancer Patients Treated With Cetuximab. J. Clin. Oncol. *25*, 3230–3237.

Kii, I., Nishiyama, T., Li, M., Matsumoto, K., Saito, M., Amizuka, N., and Kudo, A. (2010). Incorporation of Tenascin-C into the Extracellular Matrix by Periostin Underlies an Extracellular Meshwork Architecture. J. Biol. Chem. *285*, 2028–2039.

Kikuchi, Y., Kashima, T.G., Nishiyama, T., Shimazu, K., Morishita, Y., Shimazaki, M., Kii, I., Horie, H., Nagai, H., Kudo, A., et al. (2008). Periostin Is Expressed in Pericryptal Fibroblasts and Cancer-associated Fibroblasts in the Colon. J. Histochem. Cytochem. *56*, 753–764.

Kim, T.-M., Laird, P.W., and Park, P.J. (2013). The Landscape of Microsatellite Instability in Colorectal and Endometrial Cancer Genomes. Cell *155*, 858–868.

Kinzler, K.W., and Vogelstein, B. (1996). Lessons from hereditary colorectal cancer. Cell *87*, 159–170.

van Klinken, B.J., Dekker, J., van Gool, S.A., van Marle, J., Büller, H.A., and Einerhand, A.W. (1998). MUC5B is the prominent mucin in human gallbladder and is also expressed in a subset of colonic goblet cells. Am. J. Physiol. *274*, G871-8.

Ko, S.B.H., Zeng, W., Dorwart, M.R., Luo, X., Kim, K.H., Millen, L., Goto, H., Naruse, S., Soyombo, A., Thomas, P.J., et al. (2004). Gating of CFTR by the STAS domain of SLC26 transporters. Nat. Cell Biol. *6*, 343–350.

Koretz, K., Brüderlein, S., Henne, C., and Möller, P. (1992). Decay-accelerating factor (DAF, CD55) in normal colorectal mucosa, adenomas and carcinomas. Br. J. Cancer *66*, 810–814.

Kudo, A. (2011). Periostin in fibrillogenesis for tissue regeneration: periostin actions inside and outside the cell. Cell. Mol. Life Sci. *68*, 3201–3207.

Kufe, D.W. (2009). Mucins in cancer: function, prognosis and therapy. Nat. Rev. Cancer *9*, 874–885.

Kunkel, T.A. (2004). DNA replication fidelity. J. Biol. Chem. *279*, 16895–16898.

Lam, C.S.-C., Cheung, A.H.-K., Wong, S.K.-M., Wan, T.M.-H., Ng, L., Chow, A.K.-M., Cheng, N.S.-M., Pak, R.C.-H., Li, H.-S., Man, J.H.-W., et al. (2014). Prognostic Significance of CD26 in Patients with Colorectal Cancer. PLoS One *9*, e98582.

Lambeth, J.D., Suh, Y.-A., Arnold, R.S., Lassegue, B., Shi, J., Xu, X., Sorescu, D., Chung, A.B., and Griendling, K.K. (1999). Cell transformation by the superoxide-generating oxidase Mox1. Nature *401*, 79–82.

Landrum, M.J., Lee, J.M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., et al. (2016). ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. *44*, D862–D868.

Larrinaga, G., Perez, I., Sanz, B., Beitia, M., Errarte, P., Fernández, A., Blanco, L., Etxezarraga, M.C., Gil, J., and López, J.I. (2015). Dipeptidyl-Peptidase IV Activity Is Correlated with Colorectal Cancer Prognosis. PLoS One *10*, e0119436.

Larsson, J.M.H., Karlsson, H., Crespo, J.G., Johansson, M.E.V., Eklund, L., Sjövall, H., and Hansson, G.C. (2011). Altered O-glycosylation profile of MUC2 mucin occurs in active ulcerative colitis and is associated with increased inflammation. Inflamm. Bowel Dis. *17*, 2299–2307.

Laurent, E., McCoy, J.W., Macina, R.A., Liu, W., Cheng, G., Robine, S., Papkoff, J., and Lambeth, J.D. (2008). Nox1 is over-expressed in human colon cancers and correlates with activating mutations in K-Ras. Int. J. Cancer *123*, 100–107.

Law, R.H.P., Zhang, Q., McGowan, S., Buckle, A.M., Silverman, G.A., Wong, W., Rosado, C.J., Langendorf, C.G., Pike, R.N., Bird, P.I., et al. (2006). An overview of the serpin superfamily. Genome Biol. *7*, 216.

Lee, M.S., McGuffey, E.J., Morris, J.S., Manyam, G., Baladandayuthapani, V., Wei, W., Morris, V.K., Overman, M.J., Maru, D.M., Jiang, Z.-Q., et al. (2016a). Association of CpG island methylator phenotype and EREG/AREG methylation and expression in colorectal cancer. Br. J. Cancer *114*, 1352–1361.

Lee, S.H., Lee, J.W., Soung, Y.H., Kim, H.S., Park, W.S., Kim, S.Y., Lee, J.H., Park, J.Y., Cho, Y.G., Kim, C.J., et al. (2003). BRAF and KRAS mutations in stomach cancer. Oncogene *22*, 6942–6945.

Lee, Y.-J., Lee, M.-Y., Ruan, A., Chen, C.-K., Liu, H.-P., Wang, C.-J., Chao, W.-R., Han, C.-P., Lee, Y.-J., Lee, M.-Y., et al. (2016b). Multipoint <I>Kras</I> oncogene mutations potentially indicate mucinous carcinoma on the entire spectrum of mucinous ovarian neoplasms. Oncotarget *7*, 82097–82103.

Lehtinen, L., Vesterkvist, P., Roering, P., Korpela, T., Hattara, L., Kaipio, K., Mpindi, J.-P., Hynninen, J., Auranen, A., Davidson, B., et al. (2016). REG4 Is Highly Expressed in Mucinous Ovarian Cancer: A Potential Novel Serum Biomarker. PLoS One *11*, e0151590.

Lengauer, C., Kinzler, K.W., and Vogelstein, B. (1998). Genetic instabilities in human cancers. Nature *396*, 643–649.

Ligtenberg, A.J.M., Bikker, F.J., De Blieck-Hogervorst, J.M.A., Veerman, E.C.I., and Nieuw Amerongen, A. V (2004). Binding of salivary agglutinin to IgA. Biochem. J. *383*, 159–164.

Liu, W., Li, H., Hong, S.-H., Piszczek, G.P., Chen, W., and Rodgers, G.P. (2016). Olfactomedin 4 deletion induces colon adenocarcinoma in ApcMin/+ mice. Oncogene *35*, 5237–5247.

Lublin, D.M., and Atkinson, J.P. (1989). Decay-Accelerating Factor: Biochemistry, Molecular Biology, and Function. Annu. Rev. Immunol. *7*, 35–58.

Lustig, B., Jerchow, B., Sachs, M., Weiler, S., Pietsch, T., Karsten, U., van de Wetering, M., Clevers, H., Schlag, P.M., Birchmeier, W., et al. (2002). Negative feedback loop of Wnt signaling through upregulation of conductin/axin2 in colorectal and liver tumors. Mol. Cell. Biol. *22*, 1184–1193.

Lynch, H.T., Snyder, C.L., Shaw, T.G., Heinen, C.D., and Hitchins, M.P. (2015). Milestones of Lynch syndrome: 1895-2015. Nat. Rev. Cancer *15*, 181–194.

Madsen, J., Sorensen, G.L., Nielsen, O., Tornøe, I., Thim, L., Fenger, C., Mollenhauer, J., and Holmskov, U. (2013). A Variant Form of the Human Deleted in Malignant Brain Tumor 1 (DMBT1) Gene Shows Increased Expression in Inflammatory Bowel Diseases and Interacts with Dimeric Trefoil Factor 3 (TFF3). PLoS One *8*, e64441.

Maruhashi, T., Kii, I., Saito, M., and Kudo, A. (2010). Interaction between Periostin and BMP-1 Promotes Proteolytic Activation of Lysyl Oxidase. J. Biol. Chem. *285*, 13294–13303.

Matthews, J.B., Hassan, I., Meng, S., Archer, S.Y., Hrnjez, B.J., and Hodin, R.A. (1998). Na-K-2Cl cotransporter gene expression and function during enterocyte differentiation. Modulation of Cl- secretory capacity by butyrate. J. Clin. Invest. *101*, 2072–2079.

Melvin, J.E., Park, K., Richardson, L., Schultheis, P.J., and Shull, G.E. (1999). Mouse down-regulated in adenoma (DRA) is an intestinal Cl(-)/HCO(3)(-) exchanger and is up-regulated in colon of mice lacking the NHE3 Na(+)/H(+) exchanger. J. Biol. Chem. *274*, 22855–22861.

Mensenkamp, A.R., Vogelaar, I.P., van Zelst–Stams, W.A.G., Goossens, M., Ouchene, H., Hendriks–Cornelissen, S.J.B., Kwint, M.P., Hoogerbrugge, N., Nagtegaal, I.D., and Ligtenberg, M.J.L. (2014). Somatic Mutations in MLH1 and MSH2 Are a Frequent Cause of Mismatch-Repair Deficiency in Lynch Syndrome-Like Tumors. Gastroenterology *146*, 643–646.e8.

Mentlein, R. (1999). Dipeptidyl-peptidase IV (CD26)--role in the inactivation of regulatory peptides. Regul. Pept. *85*, 9–24.

Missiaglia, E., Jacobs, B., D'Ario, G., Di Narzo, A.F., Soneson, C., Budinska, E., Popovici, V., Vecchione, L., Gerster, S., Yan, P., et al. (2014). Distal and proximal colon cancers differ in terms of molecular, pathological, and clinical features. Ann. Oncol. *25*, 1995–2001.

Mistry, S.K., Burwell, T.J., Chambers, R.M., Rudolph-Owen, L., Spaltmann, F., Cook, W.J., and Morris, S.M. (2002). Cloning of human agmatinase. An alternate path for polyamine synthesis induced in liver by hepatitis B virus. Am. J. Physiol. - Gastrointest. Liver Physiol. *282*, G375–G381.

Mitsushita, J., Lambeth, J.D., and Kamata, T. (2004). The Superoxide-Generating Oxidase Nox1 Is Functionally Required for Ras Oncogene Transformation. Cancer Res. *64*, 3580–3585.

Modrich, P. (2016). Mechanisms in E. coli and Human Mismatch Repair (Nobel Lecture). Angew. Chem. Int. Ed. Engl. *55*, 8490–8501.

Mollenhauer, J., Wiemann, S., Scheurlen, W., Korn, B., Hayashi, Y., Wilgenbus, K.K., von Deimling, A., and Poustka, A. (1997). DMBT1, a new member of the SRCR superfamily, on chromosome 10q25.3–26.1 is deleted in malignant brain tumours. Nat. Genet. *17*, 32–39.

Mollenhauer, J., Herbertz, S., Holmskov, U., Tolnay, M., Krebs, I., Merlo, A., Schrøder, H.D., Maier, D., Breitling, F., Wiemann, S., et al. (2000). DMBT1 encodes a protein involved in the immune defense and in epithelial differentiation and is highly unstable in cancer. Cancer Res. *60*, 1704–1710.

Mollenhauer, J., End, C., Renner, M., Lyer, S., and Poustka, A. (2007). DMBT1 as an archetypal link between infection, inflammation, and cancer. Inmunología *26*, 193–209.

Morra, L., and Moch, H. (2011). Periostin expression and epithelial-mesenchymal transition in cancer: a review and an update. Virchows Arch. *459*, 465–475.

Morris, L.G.T., Ramaswami, D., and Chan, T.A. (2013a). The FAT epidemic: A gene family frequently mutated across multiple human cancer types. Cell Cycle *12*, 1011–1012.

Morris, L.G.T., Kaufman, A.M., Gong, Y., Ramaswami, D., Walsh, L.A., Turcan, Ş., Eng, S., Kannan, K., Zou, Y., Peng, L., et al. (2013b). Recurrent somatic mutation of FAT1 in multiple human cancers leads to aberrant Wnt activation. Nat. Genet. *45*, 253–261.

Munoz, J., Stange, D.E., Schepers, A.G., van de Wetering, M., Koo, B.-K., Itzkovitz, S., Volckmann, R., Kung, K.S., Koster, J., Radulescu, S., et al. (2012). The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent ?+4? cell markers. EMBO J. *31*, 3079–3091.

Murakami, M., Matsumoto, R., Urade, Y., Austen, K.F., and Arm, J.P. (1995). c-kit ligand mediates increased expression of cytosolic phospholipase A2, prostaglandin endoperoxide synthase-1, and hematopoietic prostaglandin D2 synthase and increased IgE-dependent prostaglandin D2 generation in immature mouse mast cells. J. Biol. Chem. *270*, 3239–3246.

Muskett, F.W., May, F.E.B., Westley, B.R., and Feeney, J. (2003). Solution Structure of the Disulfide-Linked Dimer of Human Intestinal Trefoil Factor (TFF3): The Intermolecular Orientation and Interactions Are Markedly Different from Those of Other Dimeric Trefoil Proteins †. Biochemistry *42*, 15139–15147.

Naini, S.M., Choukroun, G.J., Ryan, J.R., Hentschel, D.M., Shah, J. V., and Bonventre, J. V. (2016). Cytosolic phospholipase A2 regulates G1 progression through modulating FOXO1 activity. FASEB J. *30*, 1155–1170.

Nakao, K., Mehta, K.R., Fridlyand, J., Moore, D.H., Jain, A.N., Lafuente, A., Wiencke, J.W., Terdiman, J.P., and Waldman, F.M. (2004). High-resolution analysis of DNA copy number alterations in colorectal cancer by array-based comparative genomic hybridization. Carcinogenesis *25*, 1345–1357.

Nakata, K., Nagai, E., Ohuchida, K., Aishima, S., Hayashi, A., Miyasaka, Y., Yu, J., Mizumoto, K., Tanaka, M., and Tsuneyoshi, M. (2009). REG4 is associated with carcinogenesis in the "intestinal" pathway of intraductal papillary mucinous neoplasms. Mod. Pathol. *22*, 460–468.

Naor, D., Sionov, R. V, and Ish-Shalom, D. (1997). CD44: structure, function, and association with the malignant process. Adv. Cancer Res. *71*, 241–319.

Nilsson, H.E., Ambort, D., Bäckström, M., Thomsson, E., Koeck, P.J.B., Hansson, G.C., and Hebert, H. (2014). Intestinal MUC2 Mucin Supramolecular Topology by Packing and Release Resting on D3 Domain Assembly. J. Mol. Biol. *426*, 2567–2579.

Nitsche, U., Zimmermann, A., Späth, C., Müller, T., Maak, M., Schuster, T., Slotta-Huspenina, J., Käser, S.A., Michalski, C.W., Janssen, K.-P., et al. (2013). Mucinous and Signet-Ring Cell Colorectal Cancers Differ from Classical Adenocarcinomas in Tumor Biology and Prognosis. Ann. Surg. *258*, 775–783.

Nordman, H., Davies, J.R., Lindell, G., de Bolós, C., Real, F., and Carlstedt, I. (2002). Gastric MUC5AC and MUC6 are large oligomeric mucins that differ in size, glycosylation and tissue distribution. Biochem. J. *364*, 191–200.

Norris, A.M., Gore, A., Balboni, A., Young, A., Longnecker, D.S., and Korc, M. (2013). AGR2 is a SMAD4-suppressible gene that modulates MUC1 levels and promotes the initiation and progression of pancreatic intraepithelial neoplasia. Oncogene *32*, 3867–3876.

Norris, R.A., Damon, B., Mironov, V., Kasyanov, V., Ramamurthi, A., Moreno-Rodriguez, R., Trusk, T., Potts, J.D., Goodwin, R.L., Davis, J., et al. (2007). Periostin regulates collagen fibrillogenesis and the biomechanical properties of connective tissues. J. Cell. Biochem. *101*, 695–711.

Oh, H.J., Bae, J.M., Wen, X.-Y., Cho, N.-Y., Kim, J.H., and Kang, G.H. (2017). Overexpression of POSTN in Tumor Stroma Is a Poor Prognostic Indicator of Colorectal Cancer. J. Pathol. Transl. Med. *51*, 306–313.

Orian-Rousseau, V. (2010). CD44, a therapeutic target for metastasising tumours. Eur. J. Cancer *46*, 1271–1277.

Ou, G., Baranov, V., Lundmark, E., Hammarström, S., and Hammarström, M.-L. (2009). Contribution of Intestinal Epithelial Cells to Innate Immunity of the Human Gut - Studies on Polarized Monolayers of Colon Carcinoma Cells. Scand. J. Immunol. *69*, 150–161.

Oue, N., Mitani, Y., Aung, P.P., Sakakura, C., Takeshima, Y., Kaneko, M., Noguchi, T., Nakayama, H., and Yasui, W. (2005). Expression and localization of Reg IV in human neoplastic and non-neoplastic tissues: Reg IV expression is associated with intestinal and neuroendocrine differentiation in gastric adenocarcinoma. J. Pathol. *207*, 185–198.

Oue, N., Kuniyasu, H., Noguchi, T., Sentani, K., Ito, M., Tanaka, S., Setoyama, T., Sakakura, C., Natsugoe, S., and Yasui, W. (2007). Serum concentration of Reg IV in patients with colorectal cancer: overexpression and high serum levels of Reg IV are associated with liver metastasis. Oncology *72*, 371–380.

Palles, C., Cazier, J.-B., Howarth, K.M., Domingo, E., Jones, A.M., Broderick, P., Kemp, Z., Spain, S.L., Guarino, E., Guarino Almeida, E., et al. (2013). Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nat. Genet. *45*, 136–144.

Pande, R., Sunga, A., LeVea, C., Wilding, G.E., Bshara, W., Reid, M., and Fakih, M.G. (2008). Significance of Signet-Ring Cells in Patients with Colorectal Cancer. Dis. Colon Rectum *51*, 50–55.

Pang, R., Law, W.L., Chu, A.C.Y., Poon, J.T., Lam, C.S.C., Chow, A.K.M., Ng, L., Cheung, L.W.H., Lan, X.R., Lan, H.Y., et al. (2010). A Subpopulation of CD26+ Cancer Stem Cells with Metastatic Capacity in Human Colorectal Cancer. Cell Stem Cell *6*, 603–615.

Park, S.-W., Zhen, G., Verhaeghe, C., Nakagami, Y., Nguyenvu, L.T., Barczak, A.J., Killeen, N., and Erle, D.J. (2009). The protein disulfide isomerase AGR2 is essential for production of intestinal mucus. Proc. Natl. Acad. Sci. *106*, 6950–6955.

Pelaseyed, T., Bergström, J.H., Gustafsson, J.K., Ermund, A., Birchenough, G.M.H., Schütte, A., van der Post, S., Svensson, F., Rodríguez-Piñeiro, A.M., Nyström, E.E.L., et al. (2014). The mucus and mucins of the goblet cells and enterocytes provide the first defense line of the gastrointestinal tract and interact with the immune system. Immunol. Rev. *260*, 8–20.

Peltomäki, P. (2014). Epigenetic mechanisms in the pathogenesis of Lynch syndrome. Clin. Genet. *85*, 403–412.

Peltomäki, P. (2016). Update on Lynch syndrome genomics. Fam. Cancer *15*, 385–393.

Peña-Diaz, J., and Jiricny, J. (2012). Mammalian mismatch repair: error-free or error-prone? Trends Biochem. Sci. *37*, 206–214.

Peng, F., Huang, Y., Li, M.-Y., Li, G.-Q., Huang, H.-C., Guan, R., Chen, Z.-C., Liang, S.-P., and Chen, Y.-H. (2016). Dissecting characteristics and dynamics of differentially expressed proteins during multistage carcinogenesis of human colorectal cancer. World J. Gastroenterol. *22*, 4515.

Perez-Villamil, B., Romera-Lopez, A., Hernandez-Prieto, S., Lopez-Campos, G., Calles, A., Lopez-Asenjo, J.-A., Sanz-Ortega, J., Fernandez-Perez, C., Sastre, J., Alfonso, R., et al. (2012). Colon cancer molecular subtypes identified by expression profiling and associated to stroma, mucinous type and different clinical behavior. BMC Cancer *12*, 260.

Pileri, P., Campagnoli, S., Grandi, A., Parri, M., De Camilli, E., Song, C., Ganfini, L., Lacombe, A., Naldi, I., Sarmientos, P., et al. (2016). FAT1: a potential target for monoclonal antibody therapy in colon cancer. Br. J. Cancer *115*, 40–51.

Planck, M., Koul, A., Fernebro, E., Borg, A., Kristoffersson, U., Olsson, H., Wenngren, E., Mangell, P., and Nilbert, M. (1999). hMLH1, hMSH2 and hMSH6 mutations in hereditary non-polyposis colorectal cancer families from southern Sweden. Int. J. Cancer *83*, 197–202.

Playford, R.J. (1997). Trefoil peptides: what are they and what do they do? J. R. Coll. Physicians Lond. *31*, 37–41.

Plazzer, J.P., Sijmons, R.H., Woods, M.O., Peltomäki, P., Thompson, B., Den Dunnen, J.T., and Macrae, F. (2013). The InSiGHT database: utilizing 100 years of insights into Lynch syndrome. Fam. Cancer *12*, 175–180.

Preston, B.D., Albertson, T.M., and Herr, A.J. (2010). DNA replication fidelity and cancer. Semin. Cancer Biol. *20*, 281–293.

Pyle-Chenault, R.A., Stolk, J.A., Molesh, D.A., Boyle-Harlan, D., McNeill, P.D., Repasky, E.A., Jiang, Z., Fanger, G.R., and Xu, J. (2005). VSGP/F-Spondin: A New Ovarian Cancer Marker. Tumor Biol. *26*, 245–257.

Qu, X., Sandmann, T., Frierson, H., Fu, L., Fuentes, E., Walter, K., Okrah, K., Rumpel, C., Moskaluk, C., Lu, S., et al. (2016). Integrated genomic analysis of colorectal cancer progression reveals activation of EGFR through demethylation of the EREG promoter. Oncogene *35*, 6403–6415.

Rafa, L., Dessein, A.-F., Devisme, L., Buob, D., Truant, S., Porchet, N., Huet, G., Buisine, M.-P., and Lesuffleur, T. (2010). REG4 acts as a mitogenic, motility and pro-invasive factor for colon cancer cells. Int. J. Oncol. *36*, 689–698.

Renner, M., Bergmann, G., Krebs, I., End, C., Lyer, S., Hilberg, F., Helmke, B., Gassler, N., Autschbach, F., Bikker, F., et al. (2007). DMBT1 confers mucosal protection in vivo and a deletion variant is associated with Crohn's disease. Gastroenterology *133*, 1499–1509.

Roeckel, N., Woerner, S.M., Kloor, M., Yuan, Y.-P., Patsos, G., Gromes, R., Kopitz, J., and Gebert, J. (2009). High frequency of LMAN1 abnormalities in colorectal tumors with microsatellite instability. Cancer Res. *69*, 292–299.

Rohlin, A., Zagoras, T., Nilsson, S., Lundstam, U., Wahlström, J., Hultén, L., Martinsson, T., Karlsson, G.B., and Nordling, M. (2014). A mutation in POLE predisposing to a multi-tumour phenotype. Int. J. Oncol. *45*, 77–81.

Rothenberg, M.E., Nusse, Y., Kalisky, T., Lee, J.J., Dalerba, P., Scheeren, F., Lobo, N., Kulkarni, S., Sim, S., Qian, D., et al. (2012). Identification of a cKit+ Colonic Crypt Base Secretory Cell That Supports Lgr5+ Stem Cells in Mice. Gastroenterology *142*, 1195–1205.e6.

Rubio, C.A., and Kaufeldt, A. Maspin highlights colorectal serrated polyps: preliminary findings. In Vivo *29*, 391–393.

Sadeqzadeh, E., de Bock, C.E., Zhang, X.D., Shipman, K.L., Scott, N.M., Song, C., Yeadon, T., Oliveira, C.S., Jin, B., Hersey, P., et al. (2011). Dual processing of FAT1 cadherin protein by human melanoma cells generates distinct protein products. J. Biol. Chem. *286*, 28181–28191.

Sahin, U., Weskamp, G., Kelly, K., Zhou, H.-M., Higashiyama, S., Peschon, J., Hartmann, D., Saftig, P., and Blobel, C.P. (2004). Distinct roles for ADAM10 and ADAM17 in ectodomain shedding of six EGFR ligands. J. Cell Biol. *164*.

Sasaki, H., Dai, M., Auclair, D., Fukai, I., Kiriyama, M., Yamakawa, Y., Fujii, Y., and Chen, L.B. (2001). Serum level of the periostin, a homologue of an insect cell adhesion molecule, as a prognostic marker in nonsmall cell lung carcinomas. Cancer *92*, 843–848.

Sasaki, N., Sachs, N., Wiebrands, K., Ellenbroek, S.I.J., Fumagalli, A., Lyubimova, A., Begthel, H., van den Born, M., van Es, J.H., Karthaus, W.R., et al. (2016). Reg4+ deep crypt secretory cells function as epithelial niche for Lgr5+ stem cells in colon. Proc. Natl. Acad. Sci. *113*, E5399–E5407.

Sato, T., van Es, J.H., Snippert, H.J., Stange, D.E., Vries, R.G., van den Born, M., Barker, N., Shroyer, N.F., van de Wetering, M., and Clevers, H. (2011). Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. Nature *469*, 415–418.

Scaglia, N., Chisholm, J.W., and Igal, R.A. (2009). Inhibition of StearoylCoA Desaturase-1 Inactivates Acetyl-CoA Carboxylase and Impairs Proliferation in Cancer Cells: Role of AMPK. PLoS One *4*, e6812.

Schölzel, S., Zimmermann, W., Schwarzkopf, G., Grunert, F., Rogaczewski, B., and Thompson, J. (2000). Carcinoembryonic Antigen Family Members CEACAM6 and CEACAM7 Are Differentially Expressed in Normal Tissues and Oppositely Deregulated in Hyperplastic Colorectal Polyps and Early Adenomas. Am. J. Pathol. *156*, 595–605.

Schroeder, B.W., Verhaeghe, C., Park, S.-W., Nguyenvu, L.T., Huang, X., Zhen, G., and Erle, D.J. (2012). AGR2 is induced in asthma and promotes allergen-induced mucin overproduction. Am. J. Respir. Cell Mol. Biol. *47*, 178–185.

Schubbert, S., Zenker, M., Rowe, S.L., Böll, S., Klein, C., Bollag, G., van der Burgt, I., Musante, L., Kalscheuer, V., Wehner, L.-E., et al. (2006). Germline KRAS mutations cause Noonan syndrome. Nat. Genet. *38*, 331–336.

Schutte, A., Ermund, A., Becker-Pauly, C., Johansson, M.E. V., Rodriguez-Pineiro, A.M., Backhed, F., Muller, S., Lottaz, D., Bond, J.S., and Hansson, G.C. (2014). Microbial-induced meprin β cleavage in MUC2 mucin and a functional CFTR channel are required to release anchored small intestinal mucus. Proc. Natl. Acad. Sci. *111*, 12396–12401.

Schütte, M., Risch, T., Abdavi-Azar, N., Boehnke, K., Schumacher, D., Keil, M., Yildiriman, R., Jandrasits, C., Borodina, T., Amstislavskiy, V., et al. (2017). Molecular dissection of colorectal cancer in pre-clinical models identifies biomarkers predicting sensitivity to EGFR inhibitors. Nat. Commun. *8*, 14262.

Seko, N., Oue, N., Noguchi, T., Sentani, K., Sakamoto, N., Hinoi, T., Okajima, M., and Yasui, W. (2009). Olfactomedin 4 (GW112, hGC-1) is an independent prognostic marker for survival in patients with colorectal cancer. Exp. Ther. Med. *1*, 73–78.

Shang, Y., Chai, N., Gu, Y., Ding, L., Yang, Y., Zhou, J., Ren, G., Hao, X., Fan, D., Wu, K., et al. (2014). Systematic Immunohistochemical Analysis of the Expression of CD46, CD55, and CD59 in Colon Cancer. Arch. Pathol. Lab. Med. *138*, 910–919.

Shao, R., Bao, S., Bai, X., Blanchette, C., Anderson, R.M., Dang, T., Gishizky, M.L., Marks, J.R., and Wang, X.-F. (2004). Acquired expression of periostin by human breast cancers promotes tumor angiogenesis through up-regulation of vascular endothelial growth factor receptor 2 expression. Mol. Cell. Biol. *24*, 3992–4003.

Sharma, M., Predeus, A. V, Kovacs, N., and Feig, M. (2014). Differential mismatch recognition specificities of eukaryotic MutS homologs, MutSα and MutSβ. Biophys. J. *106*, 2483–2492.

Sheng, S., Carey, J., Seftor, E.A., Dias, L., Hendrix, M.J., and Sager, R. (1996). Maspin acts at the cell membrane to inhibit invasion and motility of mammary and prostatic cancer cells. Proc. Natl. Acad. Sci. U. S. A. *93*, 11669–11674.

Shinbrot, E., Henninger, E.E., Weinhold, N., Covington, K.R., Göksenin, A.Y., Schultz, N., Chao, H., Doddapaneni, H., Muzny, D.M., Gibbs, R.A., et al. (2014). Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. Genome Res. *24*, 1740–1750.

Shiozaki, A., Otsuji, E., and Marunaka, Y. (2011). Intracellular chloride regulates the G1/S cell cycle progression in gastric cancer cells. World J. Gastrointest. Oncol. *3*, 119.

Shiozaki, A., Nako, Y., Ichikawa, D., Konishi, H., Komatsu, S., Kubota, T., Fujiwara, H., Okamoto, K., Kishimoto, M., Marunaka, Y., et al. (2014). Role of the $Na^+/K^+/2Cl^-$ cotransporter NKCC1 in cell cycle progression in human esophageal squamous cell carcinoma. World J. Gastroenterol. *20*, 6844–6859.

Sobin, L.H., Gospodarowicz, M.K., and Wittekind, C. (2011). TNM Classification of Malignant Tumours. (Wiley).

Sopko, R., and McNeill, H. (2009). The skinny on Fat: an enormous cadherin that regulates cell adhesion, tissue growth, and planar cell polarity. Curr. Opin. Cell Biol. *21*, 717–723.

Sourrouille, I., Coulet, F., Lefevre, J.H., Colas, C., Eyries, M., Svrcek, M., Bardier-Dupas, A., Parc, Y., and Soubrier, F. (2013). Somatic mosaicism and double somatic hits can lead to MSI colorectal tumors. Fam. Cancer *12*, 27–33.

Stenzinger, A., Pfarr, N., Endris, V., Penzel, R., Jansen, L., Wolf, T., Herpel, E., Warth, A., Klauschen, F., Kloor, M., et al. (2014). Mutations in POLE and survival of colorectal cancer patients--link to disease stage and treatment. Cancer Med. *3*, 1527–1538.

Strong, T. V, Boehm, K., and Collins, F.S. (1994). Localization of cystic fibrosis transmembrane conductance regulator mRNA in the human gastrointestinal tract by in situ hybridization. J. Clin. Invest. *93*, 347–354.

Sung, C.O., Seo, J.W., Kim, K.-M., Do, I.-G., Kim, S.W., and Park, C.-K. (2008). Clinical significance of signet-ring cells in colorectal mucinous adenocarcinoma. Mod. Pathol. *21*, 1533–1541.

Tai, C.-J., Shen, S.-C., Lee, W.-R., Liao, C.-F., Deng, W.-P., Chiou, H.-Y., Hsieh, C.-I., Tung, J.-N., Chen, C.-S., Chiou, J.-F., et al. (2010). Increased cellular apoptosis susceptibility (CSE1L/CAS) protein expression promotes protrusion extension and enhances migration of MCF-7 breast cancer cells. Exp. Cell Res. *316*, 2969–2981.

Tai, C.-J., Su, T.-C., Jiang, M.-C., Chen, H.-C., Shen, S.-C., Lee, W.-R., Liao, C.-F., Chen, Y.-C., Lin, S.-H., Li, L.-T., et al. (2013). Correlations between cytoplasmic CSE1L in neoplastic colorectal glands and depth of tumor penetration and cancer stage. J. Transl. Med. *11*, 29.

Tanabe, H., Takayama, I., Nishiyama, T., Shimazaki, M., Kii, I., Li, M., Amizuka, N., Katsube, K., and Kudo, A. (2010). Periostin Associates with Notch1 Precursor to Maintain Notch1 Expression under a Stress Condition in Mouse Cells. PLoS One *5*, e12234.

Tanaka, T., Ohkubo, S., Tatsuno, I., Prives, C., Kumar-Sinha, C., Sanda, M.G., Ghosh, D., Pienta, K.J., Sewalt, R.G., Otte, A.P., et al. (2007). hCAS/CSE1L associates with chromatin and regulates expression of select p53 target genes. Cell *130*, 638–650.

Thompson, B.A., Spurdle, A.B., Plazzer, J.-P., Greenblatt, M.S., Akagi, K., Al-Mulla, F., Bapat, B., Bernstein, I., Capellá, G., den Dunnen, J.T., et al. (2013). Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. Nat. Genet. *46*, 107–115.

Thorvaldsen, T.E., Pedersen, N.M., Wenzel, E.M., Stenmark, H., Cronin, N., and Beuron, F. (2017). Differential Roles of AXIN1 and AXIN2 in Tankyrase Inhibitor-Induced Formation of Degradasomes and β-Catenin Degradation. PLoS One *12*, e0170508.

Tilman, G., Mattiussi, M., Brasseur, F., van Baren, N., and Decottignies, A. (2007). Human periostin gene expression in normal tissues, tumors and melanoma: evidences for periostin production by both stromal and melanoma cells. Mol. Cancer *6*, 80.

Todaro, M., Gaggianesi, M., Catalano, V., Benfante, A., Iovino, F., Biffoni, M., Apuzzo, T., Sperduti, I., Volpe, S., Cocorullo, G., et al. (2014). CD44v6 Is a Marker of Constitutive and Reprogrammed Cancer Stem Cells Driving Colon Cancer Metastasis. Cell Stem Cell *14*, 342–356.

Tomasetti, C., Li, L., and Vogelstein, B. (2017). Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. Science *355*, 1330–1334.

Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J.G., Baylin, S.B., and Issa, J.P. (1999). CpG island methylator phenotype in colorectal cancer. Proc. Natl. Acad. Sci. U. S. A. *96*, 8681–8686.

Tsuji, T., Satoyoshi, R., Aiba, N., Kubo, T., Yanagihara, K., Maeda, D., Goto, A., Ishikawa, K., Yashiro, M., and Tanaka, M. (2015). Agr2 Mediates Paracrine Effects on Stromal Fibroblasts That Promote Invasion by Gastric Signet-Ring Carcinoma Cells. Cancer Res. *75*, 356–366.

Tzarfati-Majar, V., Burstyn-Cohen, T., and Klar, A. (2001). F-spondin is a contact-repellent molecule for embryonic motor neurons. Proc. Natl. Acad. Sci. U. S. A. *98*, 4722–4727.

Valo, S., Kaur, S., Ristimäki, A., Renkonen-Sinisalo, L., Järvinen, H., Mecklin, J.-P., Nyström, M., and Peltomäki, P. (2015). DNA hypermethylation appears early and shows increased frequency with dysplasia in Lynch syndrome-associated colorectal adenomas and carcinomas. Clin. Epigenetics *7*, 71.

Vasen, H.F., Watson, P., Mecklin, J.P., and Lynch, H.T. (1999). New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. Gastroenterology *116*, 1453–1456.

Vilar, E., and Gruber, S.B. (2010). Microsatellite instability in colorectal cancer—the stable evidence. Nat. Rev. Clin. Oncol. *7*,

153–162.

Vogelstein, B., Fearon, E.R., Hamilton, S.R., Kern, S.E., Preisinger, A.C., Leppert, M., Smits, A.M.M., Bos, J.L., Smits, A.M., and Bos, J.L. (1988). Genetic Alterations during Colorectal-Tumor Development. N. Engl. J. Med. *319*, 525–532.

Vonlanthen, J., Okoniewski, M.J., Menigatti, M., Cattaneo, E., Pellegrini-Ochsner, D., Haider, R., Jiricny, J., Staiano, T., Buffoli, F., and Marra, G. (2014). A comprehensive look at transcription factor gene expression changes in colorectal adenomas. BMC Cancer *14*, 46.

Walsh, M.D., Clendenning, M., Williamson, E., Pearson, S.-A., Walters, R.J., Nagler, B., Packenas, D., Win, A.K., Hopper, J.L., Jenkins, M.A., et al. (2013). Expression of MUC2, MUC5AC, MUC5B, and MUC6 mucins in colorectal cancers and their association with the CpG island methylator phenotype. Mod. Pathol. *26*, 1642–1656.

Wang, D., Fu, L., Sun, H., Guo, L., and DuBois, R.N. (2015). Prostaglandin E2 Promotes Colorectal Cancer Stem Cell Expansion and Metastasis in Mice. Gastroenterology *149*, 1884–1895.e4.

Wei, Q., Wang, X., Gao, J., Li, J., Li, J., Qi, C., Li, Y., Li, Z., and Shen, L. (2016). Clinicopathologic and Molecular Features of Colorectal Adenocarcinoma with Signet-Ring Cell Component. PLoS One *11*, e0156659.

Weisenberger, D.J., Siegmund, K.D., Campan, M., Young, J., Long, T.I., Faasse, M.A., Kang, G.H., Widschwendter, M., Weener, D., Buchanan, D., et al. (2006). CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. Nat. Genet. *38*, 787–793.

Welsh, M.J., Smith, P.L., Fromm, M., and Frizzell, R.A. (1982). Crypts are the site of intestinal fluid and electrolyte secretion. Science *218*, 1219–1221.

van de Wetering, M., Sancho, E., Verweij, C., de Lau, W., Oving, I., Hurlstone, A., van der Horn, K., Batlle, E., Coudreuse, D., Haramis, A.-P., et al. (2002). The β-Catenin/TCF-4 Complex Imposes a Crypt Progenitor Phenotype on Colorectal Cancer Cells. Cell *111*, 241–250.

Weyn, C., Van Raemdonck, S., Dendooven, R., Maes, V., Zwaenepoel, K., Lambin, S., and Pauwels, P. (2017). Clinical performance evaluation of a sensitive, rapid low-throughput test for KRAS mutation analysis using formalin-fixed, paraffin-embedded tissue samples. BMC Cancer *17*, 139.

Wojtalewicz, N., Sadeqzadeh, E., Weiß, J. V., Tehrani, M.M., Klein-Scory, S., Hahn, S., Schmiegel, W., Warnken, U., Schnölzer, M., de Bock, C.E., et al. (2014). A Soluble Form of the Giant Cadherin Fat1 Is Released from Pancreatic Cancer Cells by ADAM10 Mediated Ectodomain Shedding. PLoS One *9*, e90461.

Wood, S.M., Gill, A.J., Brodsky, A.S., Lu, S., Friedman, K., Karashchuk, G., Lombardo, K., Yang, D., and Resnick, M.B. (2017). Fatty acid-binding protein 1 is preferentially lost in microsatellite instable colorectal carcinomas and is immune modulated via the interferon γ pathway. Mod. Pathol. *30*, 123–133.

Wu, Z.-Q., Brabletz, T., Fearon, E., Willis, A.L., Hu, C.Y., Li, X.-Y., and Weiss, S.J. (2012). Canonical Wnt suppressor, Axin2, promotes colon carcinoma oncogenic activity. Proc. Natl. Acad. Sci. *109*, 11312–11317.

Yao, M., Xie, C., Kiang, M.-Y., Teng, Y., Harman, D., Tiffen, J., Wang, Q., Sved, P., Bao, S., Witting, P., et al. (2015). Targeting of cytosolic phospholipase A2α impedes cell cycle re-entry of quiescent prostate cancer cells. Oncotarget *6*, 34458–34474.

Yin, S., Li, X., Meng, Y., Finley, R.L., Sakr, W., Yang, H., Reddy, N., and Sheng, S. (2005). Tumor-suppressive maspin regulates cell response to oxidative stress by direct interaction with glutathione S-transferase. J. Biol. Chem. *280*, 34985–34996.

Ying, H., Kimmelman, A.C., Lyssiotis, C.A., Hua, S., Chu, G.C., Fletcher-Sananikone, E., Locasale, J.W., Son, J., Zhang, H., Coloff, J.L., et al. (2012). Oncogenic Kras maintains pancreatic tumors through regulation of anabolic glucose metabolism. Cell *149*, 656–670.

Yoshida, R., Miyashita, K., Inoue, M., Shimamoto, A., Yan, Z., Egashira, A., Oki, E., Kakeji, Y., Oda, S., and Maehara, Y. (2011). Concurrent genetic alterations in DNA polymerase proofreading and mismatch repair in human colorectal cancer. Eur. J. Hum. Genet. *19*, 320–325.

Yuen, S.T., Davies, H., Chan, T.L., Ho, J.W., Bignell, G.R., Cox, C., Stephens, P., Edkins, S., Tsui, W.W., Chan, A.S., et al. (2002). Similarity of the phenotypic patterns associated with BRAF and KRAS mutations in colorectal neoplasia. Cancer Res. *62*, 6451–6455.

Zeilstra, J., Joosten, S.P.J., van Andel, H., Tolg, C., Berns, A., Snoek, M., van de Wetering, M., Spaargaren, M., Clevers, H., and Pals, S.T. (2014). Stem cell CD44v isoforms promote intestinal cancer formation in Apc(min) mice downstream of Wnt signaling. Oncogene *33*, 665–670.

Zhang, Y., Yuan, F., Presnell, S.R., Tian, K., Gao, Y., Tomkinson, A.E., Gu, L., and Li, G.-M. (2005). Reconstitution of 5′-Directed Human Mismatch Repair in a Purified System. Cell *122*, 693–705.

Zhao, L., and Vogt, P.K. (2008). Class I PI3K in oncogenic cellular transformation. Oncogene *27*, 5486–5496.

Zucman-Rossi, J., Jeannot, E., Nhieu, J.T. Van, Scoazec, J.-Y., Guettier, C., Rebouissou, S., Bacq, Y., Leteurtre, E., Paradis, V., Michalak, S., et al. (2006). Genotype-phenotype correlation in hepatocellular adenoma: new classification and relationship with HCC. Hepatology *43*, 515–524.

8. SUPPLEMENTARY MATERIAL

Supplementary Table 1: Details on the custom MMR/Polymerases Panel.

Gene Name	Chromosome	Number of Amplicons	Total Bases	Covered Bases	Missed Bases	Overall Coverage	Number of Exons
MSH2	chr2	53	4405	4140	265	0.940	17
MSH6	chr2	54	5076	4780	296	0.942	12
MLH1	chr3	52	4171	4126	45	0.989	21
MSH3	chr5	72	5814	5544	270	0.954	24
PMS2	chr7	41	4089	3198	891	0.782	15
POLE	chr12	132	11684	11051	633	0.946	49
MLH3	chr14	62	5562	5285	277	0.950	12
POLD1	chr19	59	5773	4746	1027	0.822	26

Supplementary Table 2: Genes targeted by the Ion AmpliSeq™ Cancer Hotspot Panel v2.

Ion AmpliSeq™ Cancer Hotspot Panel v2 (CHS Panel)

#	Gene	RefSeq NM	Description (GeneCards)	Ensemble Transcript ID	Chromosome (Ensembl)
1	*ABL1*	NM_007313.2	ABL Proto-Oncogene 1, Non-Receptor Tyrosine Kinase	ENST00000372348.6	9q34.12
2	*AKT1*	NM_001014432	AKT Serine/Threonine Kinase 1	ENST00000349310.7	14q32.33
3	*ALK*	NM_004304.4	Anaplastic Lymphoma Receptor Tyrosine Kinase	ENST00000389048.7	2p23.1
4	*APC*	NM_000038.5	APC, WNT Signaling Pathway Regulator	ENST00000257430.8	5q22.2
5	*ATM*	NM_000051.3	ATM Serine/Threonine Kinase	ENST00000278616.8	11q22.3
6	*BRAF*	NM_004333.4	B-Raf Proto-Oncogene, Serine/Threonine Kinase	ENST00000288602.10	7q34
7	*CDH1*	NM_004360.3	Cadherin 1	ENST00000261769.9	16q22.1
8	*CDKN2A*	NM_000077.4	Cyclin Dependent Kinase Inhibitor 2A	ENST00000304494.9	9p21.3
9	*CSF1R*	NM_005211.3	Colony Stimulating Factor 1 Receptor	ENST00000286301.7	5q32
10	*CTNNB1*	NM_001904.3	Catenin Beta 1	ENST00000349496.9	3p22.1
11	*EGFR*	NM_005228.3	Epidermal Growth Factor Receptor	ENST00000275493.6	7p11.2
12	*ERBB2*	NM_004448	Erb-B2 Receptor Tyrosine Kinase 2	ENST00000269571.9	17q12
13	*ERBB4*	NM_005235.2	Erb-B2 Receptor Tyrosine Kinase 4	ENST00000342788.8	2q34
14	*EZH2*	NM_152998	Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit	ENST00000350995.6	7q36.1
15	*FBXW7*	NM_033632.3	F-Box And WD Repeat Domain Containing 7	ENST00000281708.8	4q31.3
16	*FGFR1*	NM_000604/ NM_023110	Fibroblast Growth Factor Receptor 1	ENST00000447712.6	8p11.23
17	*FGFR2*	NM_022970.3	Fibroblast Growth Factor Receptor 2	ENST00000457416.6	10q26.13
18	*FGFR3*	NM_001163213.1	Fibroblast Growth Factor Receptor 3	ENST00000340107.8	4p16.3
19	*FLT3*	NM_004119.2	Fms Related Tyrosine Kinase 3	ENST00000241453.11	13q12.2
20	*GNA11*	NM_002067.1	G Protein Subunit Alpha 11	ENST00000078429.8	19p13.3
21	*GNAQ*	NM_002072.4	G Protein Subunit Alpha Q	ENST00000286548.8	9q21.2
22	*GNAS*	NM_001077490/	GNAS Complex Locus	ENST00000371100.8	20q13.32

		NM_080425.3			
23	*HNF1A*	NM_000545.5	HNF1 Homeobox A	ENST00000257555.10	12q24.31
24	*HRAS*	NM_001130442.1	HRas Proto-Oncogene, GTPase	ENST00000451590.5	11p15.5
25	*IDH1*	NM_005896.3	Isocitrate Dehydrogenase (NADP(+)) 1, Cytosolic	ENST00000345146.6	2q34
26	*IDH2*	NM_002168.2	Isocitrate Dehydrogenase (NADP(+)) 2, Mitochondrial	ENST00000330062.7	15q26.1
27	*JAK2*	NM_004972.3	Janus Kinase 2	ENST00000381652.3	9p24.1
28	*JAK3*	NM_000215	Janus Kinase 3	ENST00000458235.5	19p13.11
29	*KDR*	NM_002253.2	Kinase Insert Domain Receptor	ENST00000263923.4	4q12
30	*KIT*	NM_000222.2	KIT Proto-Oncogene Receptor Tyrosine Kinase	ENST00000288135.5	4q12
31	*KRAS*	NM_033360.3	KRAS Proto-Oncogene, GTPase	ENST00000311936.7	12p12.1
32	*MET*	NM_001127500.1	MET Proto-Oncogene, Receptor Tyrosine Kinase	ENST00000318493.10	7q31.2
33	*MLH1*	NM_000249.3	MutL Homolog 1	ENST00000231790.6	3p22.2
34	*MPL*	NM_005373.1	MPL Proto-Oncogene, Thrombopoietin Receptor	ENST00000372470.8	1p34.2
35	*NOTCH1*	NM_017617.3	Notch 1	ENST00000277541.6	9q34.3
36	*NPM1*	NM_002520.4	Nucleophosmin	ENST00000296930.9	5q35.1
37	*NRAS*	NM_002524	Neuroblastoma RAS Viral Oncogene Homolog	ENST00000369535.4	1p13.2
38	*PDGFRA*	NM_006206.4	Platelet Derived Growth Factor Receptor Alpha	ENST00000257290.9	4q12
39	*PIK3CA*	NM_006218.2	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha	ENST00000263967.3	3q26.32
40	*PTEN*	NM_000314.4	Phosphatase And Tensin Homolog	ENST00000371953.7	10q23.31
41	*PTPN11*	NM_002834.3	Protein Tyrosine Phosphatase, Non-Receptor Type 11	ENST00000351677.6	12q24.13
42	*RB1*	NM_000321	RB Transcriptional Corepressor 1	ENST00000267163.4	13q14.2
43	*RET*	NM_020975.4	Ret Proto-Oncogene	ENST00000355710.7	10q11.21
44	*SMAD4*	NM_005359.5	SMAD Family Member 4	ENST00000342988.7	18q21.2
45	*SMARCB1*	NM_003073.3	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily B, Member 1	ENST00000263121.11	22q11.23
46	*SMO*	NM_005631.4	Smoothened, Frizzled Class Receptor	ENST00000249373.7	7q32.1
47	*SRC*	NM_005417	SRC Proto-Oncogene, Non-Receptor Tyrosine Kinase	ENST00000373578.6	20q11.23
48	*STK11*	NM_000455.4	Serine/Threonine Kinase 11	ENST00000326873.11	19p13.3
49	*TP53*	NM_000546.5	Tumor Protein P53	ENST00000269305.8	17p13.1
50	*VHL*	NM_000551.3	Von Hippel-Lindau Tumor Suppressor	ENST00000256474.2	3p25.3

Supplementary Table 3: Detailed list of germline mutations in CRC samples. *Both MSI and MSS samples are listed in table according to their BCNA number (from lowest to highest).*

Sample	BCNA	Microsatellite Status	Gene	cDNA Change	Protein Change	rs/MAF (Exac)	Percentage of variant reads with PGM	InSiGHT Class	Clinical significance (ClinVar)
P3_T1	0	MSI	***MLH1***	c.546-2A>G	p.Arg182Serfs*6[b]	rs267607759/NA[a]	Tu: 58.88%, Mu: 47.82%	Class 5	Pathogenic. Variation ID: 90267
			POLD1	c.2017G>A	p.Glu673Lys	rs61751955/A = 0.0001582	Tu: 50.17%, Mu: 46.56%	-	Uncertain significance. Variation ID: 220938
P29_T1	1	MSI	***MSH2***	c.2536C>T	p.Gln846*	rs63750857/NA[a]	Tu: 44.55%, Mu: 45.70%	Class 5	Pathogenic. Variation ID: 90996
			POLD1	c.2275G>A	p.Val759Ile	rs145473716/A = 0.001741	Tu: 48.11%, Mu: 47.70%	-	Conflicting interpretations of pathogenicity. Variation ID: 221038
			POLE	c.2083T>A	p.Phe695Ile	rs5744799/T = 0.01097	Tu: 49.05%, Mu: 49.72%	-	Benign. Variation ID: 221179
P75_T1	1	MSI	***MSH6***	c.4001+1_4001+2ins TAAC	-	rs587779302; rs587782538; rs576893678	Tu: 52.81%, Mu: 50.71%	Not present	Benign/Likely benign. Variation ID: 182672[c]
P13_T1	3	MSI	-	-	-	-	-	-	-
P41_T1	3	MSI	-	-	-	-	-	-	-
P59_T2	5	MSI	-	-	-	-	-	-	-
PAA3_T2	6	MSI	-	-	-	-	-	-	-
PCC3_T1	0	MSS	-	-	-	-	-	-	-
P65_T1	0	MSS	***PMS2***	c.1789A>T	p.Thr597Ser	rs1805318/A = 0.008209	Tu: 51.56%, Mu: 58.82%	Class 1	Likely benign. Variation ID: 41707
P73_T1	0	MSS	-	-	-	-	-	-	-
P85_T1	0	MSS	-	-	-	-	-	-	-
P85_T2	2	MSS	-	-	-	-	-	-	-
P47_T1	5	MSS	***MSH6***	c.4001+1_4001+2ins TAAC	-	rs587779302; rs587782538; rs576893678	Tu: 47.71%, Mu: 54.04	Not present	Benign/Likely benign. Variation ID: 182672[c]
P97_T2	10	MSS	-	-	-	-	-	-	-
P19_T1	16	MSS	***POLE***	c.1007A>G	p.Asn336Ser	rs5744760/C = 0.002692	Tu: 48.89%, Mu: 51.49%	-	Benign/Likely benign. Variation ID: 240370

[a]This variant is neither recorded in the 1000 Genomes Project Phase 3. [b]*MLH1* c.546-2A>G is predicted to cause skipping of exon 7 leading to a frameshift mutation p.Arg182Serfs*6. [c]This variant is reported in ClinVar as *MSH6* c.4001+12_4001+15dupACTA. Actually, the nomenclature c.4001+1_4001+2insTAAC, which we used after the alignment made by the Ion Reporter software, can be found at the following link: https://www.ncbi.nlm.nih.gov/clinvar/11240596/#clinical-assertions, but not in the last version of ClinVar database records, which instead has *MSH6* c.4001+12_4001+15dupACTA.

Supplementary Table 4: Detailed list of somatic mutations of MMR/Polymerases Panel genes in CRC samples. Both MSI and MSS samples are listed in table according to their BCNA number (from lowest to highest).

Sample	BCNA	Microsatellite Status	Gene	cDNA Change	Protein Change	rs/MAF (Exac)	Percentage of variant reads with PGM	InSiGHT Class	COSMIC	ClinVar
P3_T1	0	MSI	***MSH6***	c.3163G>A	p.Ala1055Thr	rs587779254/A = 0.000008428	8.74%	Class 3	not present	Uncertain significance. Variation ID: 89342
			POLE	c.1630G>A	p.Val544Met	NA/NA[a]	11.02%	-	not present	not present
			POLE	c.2780A>G	p.Asn927Ser	NA/NA[a]	8.57%	-	not present	not present
			POLE	c.4570C>T	p.Pro1524Ser	NA/NA[a]	9.46%	-	not present	not present
P29_T1	1	MSI	***MSH6***	c.1082G>A	p.Arg361His	rs63750440/A = 0.000008243	45.05%	Class 3	COSM190062 (FATHMM: Pathogenic, score 0.95)	Uncertain significance. Variation ID: 89168
P75_T1	1	MSI	-	-	-	-	-	-	-	-
P13_T1	3	MSI	***MLH1***	c.1459C>T	p.Arg487*	rs63749795/NA[a]	45.63%; CN-LOH[b]	Class 5	not present	Pathogenic. Variation ID: 89744
			MSH3	c.554A>T	p.Asp185Val	NA/NA[a]	18.08%	-	not present	not present
			MSH6	c.1082G>A	p.Arg361His	rs63750440/A = 0.000008243	19.70%	Class 3	COSM190062 (FATHMM: Pathogenic, score 0.95)	Uncertain significance. Variation ID: 89168
			POLD1	c.1732G>A	p.Gly578Ser	rs753850419/A = 0.00002826	25.23%	-	not present	not present
P41_T1	3	MSI	***MLH3***	c.4011G>T	p.Glu1337Asp	NA/NA[a]	10.23%	-	not present	not present
			POLE	c.2132C>T	p.Ser711Phe	NA/NA[a]	8.72%	-	not present	not present
			POLE	c.3455A>G	p.Gln1152Arg	NA/NA[a]	9.09%	-	not present	not present
P59_T2	5	MSI	***MSH2***	c.2327_2328insT	p.Cys778Leufs*9	NA/NA[a]	40.49%	-	Not present	Not present
			POLE	c.4603G>A	p.Gly1535Ser	rs138564205/T = 0.00006691	35.48%	-	COSM5959179 (FATHMM: none, score 0.00)	Uncertain significance. Variation ID: 240530
PAA3_T2	6	MSI	***MLH1***	c.1276C>T	p.Gln426*	rs63750316/NA[a]	15.2%	Class 5	Not present	Pathogenic. Variation ID: 89691
PCC3_T1	0	MSS	***MLH3***	c.55A>C	p.Ile19Leu	NA/NA[a]	62%	-	Not present	Not present

			POLE	c.4652A>C	p.His1551Pro	NA/NA[a]	8.33%	-	Not present	Not present
P65_T1	0	MSS	-	-	-	-	-	-	-	-
P73_T1	0	MSS	-	-	-	-	-	-	-	-
P85_T1	0	MSS	-	-	-	-	-	-	-	-
P85_T2	2	MSS	-	-	-	-	-	-	-	-
P47_T1	5	MSS	-	-	-	-	-	-	-	-
P97_T2	10	MSS	-	-	-	-	-	-	-	-
P19_T1	16	MSS	-	-	-	-	-	-	-	-

[a]This variant is neither recorded in the 1000 Genomes Project Phase 3. [b]Copy Neutral-Loss Of Heterozygosity (CN-LOH) involving a region of chromosome 3 including MLH1 as well.

Supplementary Table 5: Detailed list of somatic mutations of CHS Panel genes in CRC samples. Both MSI and MSS samples are listed in table according to their BCNA number (from lowest to highest).

Sample	BCNA	Microsatellite Status	Gene	cDNA Change	Protein Change	rs/MAF (Exac)	Percentage of variant reads with PGM	COSMIC	ClinVar
P3_T1	0	MSI	***CDH1***	c.1115C>A	p.Pro372His	NA/NA[a]	9.46	Not present	Not present
P29_T1	1	MSI	***KRAS***	c.35G>A	p.Gly12Asp	rs121913529/T = 0.00001976	51.03	COSM521 (FATHMM: Pathogenic, score 0.98)	Pathogenic/Likely pathogenic. Variation ID: 12582
			KRAS	c.175G>A	p.Ala59Thr	rs121913528 /NA[a]	42.88	COSM546 (FATHMM: Pathogenic, score 0.98)	Pathogenic/Likely pathogenic. Variation ID: 12581
			PIK3CA	c.1633G>A	p.Glu545Lys	rs104886003/A = 0.00000834	46.12	COSM763 (FATHMM: Pathogenic, score 0.97)	Pathogenic/Likely Pathogenic. Variation ID: 13655
P75_T1	1	MSI	***KRAS***	c.35G>T	p.Gly12Val	rs121913529/NA[a]	46.92	COSM520 (FATHMM: Pathogenic, score 0.98).	Pathogenic/Likely pathogenic. Variation ID: 12583
			TP53	c.631_632delAC	p.Thr211Phefs*4	NA/NA[a]	26.81	COSM45222 (FATHMM: none, score 0.00)	Not present
			TP53	c.746G>T	p.Arg249Met	NA/NA[a]	28.62	COSM43871 (FATHMM: Pathogenic, score 0.99)	Likely pathogenic. Variation ID: 376653

P13_T1	3	MSI	***TP53***	c.524G>A	p.Arg175His	rs28934578/T = 0.000008243	35.48	COSM10648 (FATHMM: Pathogenic, score 0.99).	Pathogenic/Likely pathogenic. Variation ID: 12374
P41_T1	3	MSI	***ATM***	c.1010G>A	p.Arg337His	rs202160435/A = 0.0000909	17.14	COSM21301 (FATHMM: Pathogenic, score 0.99)	Uncertain significance. Variation ID: 127328
			BRAF	c.1798_1799insAGA	p.Val600delinsGluMet	NA/NA[a]	17.85	Not present	Not present
			KRAS	c.40G>A	p.Val14Ile	rs104894365/T = 0.000009814	13.91	COSM12722 (FATHMM Pathogenic, score 0.98)	Pathogenic. Variation ID: 12589
P59_T2	5	MSI	***FBXW7***	c.1177C>T	p.Arg393*	NA/NA[a]	32.72	COSM22973 (FTAHMM: Pathogenic, score 0.99)	Not present
			PIK3CA	c.3073A>G	p.Thr1025Ala	rs397517202 /NA[a]	39.75	COSM771 (FATHMM: Pathogenic, score 1.00)	Pathogenic. Variation ID: 45467
PAA3_T2	6	MSI	***KRAS***	c.35G>A	p.Gly12Asp	rs121913529/T = 0.00001976	15.33	COSM521 (FATHMM: Pathogenic, score 0.98)	Pathogenic/Likely pathogenic. Variation ID: 12582.
PCC3_T1	0	MSS	-	-	-	-	-	-	-
P65_T1	0	MSS	-	-	-	-	-	-	-
P73_T1	0	MSS	-	-	-	-	-	-	-
P85_T1	0	MSS	***HNF1A***	c.833G>A	p.Arg278Gln	rs760640415/A = 0.00002775	21.78	COSM1359416 (FATHMM: none, score 0.00)	Not present
			PTEN	c.343G>T	p.Asp115Tyr	NA/NA[a]	17.91	not present	Not present
			STK11	c.1069G>A	p.Glu357Lys	rs759473833/A = 0.000009815	23.15	Not present	Uncertain significance. Variation ID: 198732
P85_T2	2	MSS	***HNF1A***	c.833G>A	p.Arg278Gln	rs760640415/A = 0.00002775	32.12	COSM1359416 (FATHMM: none, score 0.00)	Not present
			PTEN	c.343G>T	p.Asp115Tyr	NA/NA[a]	25.16	Not present	Not present
			RET	c.2767C>T	p.Leu923Phe	NA/NA[a]	8.70	Not present	Not present
			STK11	c.1069G>A	p.Glu357Lys	rs759473833/A = 0.000009815	27.74	Not present	Uncertain significance. Variation ID: 198732
P47_T1	5	MSS	***ALK***	c.3599C>T	p.Ala1200Val	rs200585833/A = 0.0001158	19.41	COSM317003 (Pathogenic, score 0.98)	Uncertain significance. Variation ID: 239825

			APC	c.4104_4105insG	p.Pro1369Alafs*6	NA/NA[a]	37.25	Not present	Not present
			BRAF	c.1790T>A	p.Leu597Gln	rs121913366/NA[a]	20.83	COSM1125 (FATHMM: Pathogenic, score 0.99)	Pathogenic/Likely pathogenic. Variation ID: 76687
			SMAD4	c.1009G>A	p.Glu337Lys	NA/NA[a]	23.17	COSM417827 (FATHMM: Pathogenic, score 0.99)	Not present
P97_T2	10	MSS	***APC***	c.2626C>T	p.Arg876*	rs121913333/NA[a]	64.62	COSM18852 (FATHMM: Pathogenic, score 0.95)	Pathogenic/Likely pathogenic. Variation ID: 216014
			TP53	c.473G>A	p.Arg158His	rs587782144/T = 0.000008249	70.45	COSM10690 (FATHMM: Pathogenic, score 0.99)	Pathogenic/Likely pathogenic. Variation ID: 141963
P19_T1	16	MSS	***KRAS***	c.35G>T	p.Gly12Val	rs121913529/NA[a]	27.50	COSM520 (FATHMM: Pathogenic, score 0.98).	Pathogenic/Likely pathogenic. Variation ID: 12583.

[a]This variant is neither recorded in the 1000 Genomes Project Phase

Supplementary Table 6: *Top genes (up to 50 entries) which have significant differential expression in HB tumors compared to normal tissue (FC <-2 or FC >2) and are upregulated (FC >2) in the HB group compared to LB and MSI tumors. Genes listed in descending order on the bases of HB FC increase in gene expression compared to LB.*

#	HB RMA	LB RMA	MSI RMA	N RMA	HB SD	LB SD	MSI SD	N SD	FDR p-value (All Conditions)	FC HB vs LB	FC HB vs MSI	FC HB vs N	Gene Symbol	Description	Chr
1	12.12	7.72	8.20	7.18	3.61	3.70	1.73	0.94	1.90E-05	21.04	15.10	30.59	NOX1	NADPH oxidase 1;	chrX
2	12.85	9.96	11.37	7.21	3.07	2.44	2.65	1.13	7.10E-07	7.43	2.80	50.04	SCD	stearoyl-CoA desaturase (delta-9-desaturase);	chr10
3	10.47	7.61	7.92	7.96	2.25	0.75	2.49	1.37	1.22E-04	7.24	5.83	5.70	AREG; AREGB	amphiregulin; amphiregulin B;	chr4
4	10.49	7.65	7.76	8.07	2.22	0.70	2.53	1.33	1.27E-04	7.18	6.65	5.36	AREG; AREGB	amphiregulin; amphiregulin B;	chr4
5	10.80	8.34	9.02	7.74	2.07	1.32	0.98	0.95	7.00E-06	5.49	3.43	8.32	TSPAN6	tetraspanin 6;	chrX
6	9.56	7.27	5.73	18.35	3.59	3.19	2.52	4.26	3.00E-06	4.89	14.28	-443.28	SLC26A3	solute carrier family 26, member 3;	chr7
7	16.35	14.40	13.47	13.17	2.76	2.27	2.49	2.78	9.75E-03	3.88	7.39	9.11	CEACAM6	carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen);	chr19
8	11.20	9.38	5.97	8.03	2.20	1.69	2.75	1.05	8.00E-06	3.54	37.57	9.00	CFTR	cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7);	chr7
9	12.46	10.67	11.32	9.69	1.55	1.19	1.61	1.02	2.00E-06	3.47	2.21	6.82	PRDX5	peroxiredoxin 5	chr11
10	11.37	9.64	8.47	9.49	1.62	1.51	2.16	1.23	1.41E-04	3.31	7.48	3.68	VIL1	villin 1;	chr2
11	8.01	6.30	6.94	6.46	1.41	1.39	1.21	0.18	5.20E-05	3.27	2.10	2.93	HUNK	hormonally up-regulated Neu-associated kinase;	chr21
12	12.81	11.12	11.45	7.60	2.28	1.98	2.31	1.35	1.00E-06	3.22	2.58	36.95	IPO7	importin 7	chr11

13	8.47	6.81	5.58	6.49	2.21	2.42	0.74	1.12	2.68E-03	3.14	7.37	3.92	ACE2	angiotensin I converting enzyme 2;	chrX
14	12.82	11.17	11.37	7.56	2.43	2.34	1.91	1.86	2.00E-06	3.14	2.72	38.31	LINC00657	long intergenic non-protein coding RNA 657	chr20
15	9.70	8.06	8.22	6.97	1.63	1.94	1.30	0.49	3.71E-07	3.13	2.80	6.66	AXIN2	axin 2	chr17
16	7.98	6.34	6.54	6.69	1.73	0.86	1.91	0.79	7.44E-04	3.10	2.71	2.44	EREG	epiregulin;	chr4
17	9.98	8.36	8.94	8.08	1.30	0.88	1.13	0.97	2.80E-05	3.08	2.05	3.74	ATP9A	ATPase, class II, type 9A;	chr20
18	9.36	7.77	7.07	7.24	1.26	1.33	1.20	0.61	7.00E-06	3.00	4.89	4.33	FARP1; FARP1-IT1	FERM, RhoGEF (ARHGEF) and pleckstrin domain protein 1 (chondrocyte-derived); FARP1 intronic transcript 1 (non-protein coding);	chr13
19	9.15	7.56	8.08	7.02	1.48	0.88	0.89	0.52	1.00E-06	2.99	2.10	4.35	POFUT1; MIR1825	protein O-fucosyltransferase 1; microRNA 1825;	chr20
20	12.47	10.91	9.92	9.87	2.09	1.72	2.20	1.85	6.10E-03	2.94	5.86	6.06	ACSL5	acyl-CoA synthetase long-chain family member 5;	chr10
21	8.10	6.60	6.24	6.24	1.44	1.27	1.23	0.61	4.90E-05	2.82	3.63	3.65	FAM111B	family with sequence similarity 111, member B	chr11
22	8.59	7.12	6.89	7.23	1.50	0.97	0.45	0.43	3.00E-05	2.76	3.23	2.57	NKD1	naked cuticle homolog 1 (Drosophila);	chr16
23	13.51	12.08	12.28	9.61	2.34	1.64	2.69	1.77	4.94E-04	2.69	2.35	14.97	TM9SF3	transmembrane 9 superfamily member 3;	chr10
24	12.34	10.92	10.01	8.11	2.07	1.83	1.85	0.92	2.77E-07	2.67	5.04	18.77	HSPH1	heat shock 105kDa/110kDa protein 1;	chr13
25	9.43	8.06	7.20	6.80	1.40	1.11	1.89	0.44	7.61E-08	2.57	4.68	6.17	MACC1	metastasis associated in colon cancer 1;	chr7
26	6.85	5.52	5.18	5.19	1.55	1.28	0.62	0.35	6.70E-05	2.52	3.18	3.17	CCL24	chemokine (C-C motif) ligand 24;	chr7
27	8.09	6.80	6.05	7.00	1.43	1.30	1.63	0.58	1.68E-03	2.45	4.11	2.13	CDHR1	cadherin-related family member 1;	chr10
28	11.69	10.46	10.01	7.29	1.79	1.54	1.32	0.58	1.09E-09	2.35	3.20	21.16	RNF43	ring finger protein 43	chr17
29	7.30	6.07	6.23	6.29	1.27	1.22	1.11	0.25	9.16E-04	2.34	2.10	2.02	RNF157	ring finger protein 157;	chr17
30	9.64	8.44	7.60	6.99	1.89	1.66	2.06	1.22	4.10E-04	2.30	4.10	6.25	TM4SF1	transmembrane 4 L six family member 1;	chr3
31	12.29	11.19	11.18	9.21	2.05	1.66	2.54	1.68	4.96E-04	2.15	2.16	8.48	TFRC	transferrin receptor (p90, CD71);	chr3
32	9.21	8.16	8.19	7.76	0.86	0.61	0.79	0.44	9.10E-08	2.06	2.03	2.72	PRPF6	pre-mRNA processing factor 6;	chr20
33	7.78	6.77	6.01	6.45	1.15	0.73	1.90	0.36	1.26E-04	2.01	3.40	2.51	PROX1	prospero homeobox 1;	chr1

HB, high-BCNA tumors; LB, low-BCNA tumors; MSI, microsatellite-instability tumors; N, normal tissue; RMA, bi-weight average signal (Robust Multi-array Average); SD, standard deviation; FDR, false discovery rate; FC, fold change, Chr, chromosome.

***Supplementary Table 7**: Top genes (up to 50 entries) which have significant differential expression in LB tumors compared to normal tissue (FC <-2 or FC >2) and are upregulated (FC >2) in the LB group compared to HB and MSI tumors. Genes listed in descending order on the bases of LB FC increase in gene expression compared to HB.*

#	HB RMA	LB RMA	MSI RMA	N RMA	HB SD	LB SD	MSI SD	N SD	FDR p-value (All Conditions)	FC LB vs HB	FC LB vs MSI	FC LB vs N	Gene Symbol	Description	Chr
1	12.66	16.90	10.70	11.03	3.44	2.31	3.77	1.60	3.21E-04	18.96	73.61	58.75	LYZ	lysozyme	chr12
2	8.58	12.72	8.12	10.46	2.31	3.42	3.90	2.72	2.48E-03	17.58	24.20	4.79	MUC2	mucin 2, oligomeric mucus/gel-forming;	chr11
3	7.56	9.82	8.20	7.63	1.63	3.10	1.57	1.55	3.88E-02	4.79	3.08	4.57	LAPTM5	lysosomal protein transmembrane 5;	chr1
4	10.18	12.41	6.70	7.39	3.57	2.24	4.86	2.22	2.19E-03	4.71	52.50	32.36	DMBT1	deleted in malignant brain tumors 1;	chr10
5	6.60	8.64	7.56	7.12	0.84	0.95	3.83	1.18	4.86E-04	4.11	2.11	2.87	CTSE	cathepsin E;	chr1
6	9.60	11.42	7.49	12.45	2.06	2.16	1.67	1.98	5.80E-04	3.54	15.33	-2.03	FABP1	fatty acid binding protein 1, liver;	chr2
7	8.95	10.65	8.47	9.16	1.08	2.42	1.14	1.01	2.18E-02	3.24	4.53	2.82	SRGN	serglycin;	chr10
8	7.59	9.03	7.29	7.20	1.29	1.89	0.64	1.03	1.03E-02	2.72	3.36	3.55	ARHGDIB	Rho GDP dissociation inhibitor (GDI) beta	chr12
9	5.85	7.24	6.06	6.13	0.63	1.77	1.34	0.56	7.89E-04	2.62	2.27	2.15	TCN1	transcobalamin I (vitamin B12 binding protein, R binder family)	chr11
10	6.96	8.28	7.18	9.64	1.95	2.46	1.26	2.35	3.27E-03	2.50	2.13	-2.57	NXPE1	neurexophilin and PC-esterase domain family, member 1	chr11

HB, high-BCNA tumors; LB, low-BCNA tumors; MSI, microsatellite-instability tumors; N, normal tissue; RMA, bi-weight average signal (Robust Multi-array Average); SD, standard deviation; FDR, false discovery rate; FC, fold change, Chr, chromosome.

***Supplementary Table 8:** Top genes (up to 50 entries) which have significant differential expression in MSI tumors compared to normal tissue (FC <-2 or FC >2) and are upregulated (FC >2) in the MSI group compared to HB and LB tumors. Genes listed in descending order on the bases of MSI FC increase in gene expression compared to HB.*

#	HB RMA	LB RMA	MSI RMA	N RMA	HB SD	LB SD	MSI SD	N SD	FDR p-value (All Conditions)	FC MSI vs HB	FC MSI vs LB	FC MSI vs N	Gene Symbol	Description	Chr
1	5.75	8.34	13.65	5.67	1.06	2.97	4.68	0.93	2.73E-07	238.78	39.64	252.85	REG4	regenerating islet-derived family, member 4;	chr1
2	6.35	7.39	11.41	6.29	1.07	1.29	2.60	0.57	4.00E-06	33.28	16.19	34.72	SERPINB5	serpin peptidase inhibitor, clade B (ovalbumin), member 5;	chr18
3	5.99	6.32	10.29	6.61	1.96	2.07	3.83	0.44	1.24E-03	19.66	15.64	12.86	PLA2G4A	phospholipase A2, group IVA (cytosolic, calcium-dependent);	chr1
4	7.79	8.70	11.94	6.65	1.32	1.85	2.41	0.79	5.00E-06	17.81	9.50	39.21	AGR2	anterior gradient 2 homolog (Xenopus laevis);	chr7
5	7.20	9.12	11.10	7.09	1.74	2.67	2.84	0.84	2.85E-03	14.90	3.94	16.11	MUC5B	mucin 5B, oligomeric mucus/gel-forming	chr11
6	5.83	6.47	9.26	6.38	0.66	0.93	2.23	0.62	1.60E-05	10.76	6.93	7.33	SPON1	spondin 1, extracellular matrix protein	chr11
7	9.01	10.31	12.28	8.22	0.99	1.42	1.83	1.02	1.40E-05	9.67	3.93	16.72	CD55	CD55 molecule, decay accelerating factor for complement (Cromer blood group);	chr1
8	7.16	8.37	10.42	6.62	2.48	4.89	3.50	0.54	2.90E-04	9.59	4.15	13.95	SPP1	secreted phosphoprotein 1;	chr4
9	7.12	8.05	10.24	7.01	1.82	2.62	3.56	0.60	6.22E-03	8.73	4.56	9.40	DPP4	dipeptidyl-peptidase 4;	chr2
10	11.46	11.41	14.56	8.94	2.09	2.09	3.35	1.53	5.10E-05	8.63	8.88	49.22	FAT1	FAT atypical cadherin 1; FAT tumor suppressor homolog 1 (Drosophila);	chr4
11	11.00	10.66	14.11	7.99	2.01	1.44	3.68	1.29	7.00E-06	8.63	10.95	69.60	LMAN1	lectin, mannose-binding, 1;	chr18

12	6.07	7.39	8.95	5.89	1.28	1.31	1.60	0.42	2.04E-04	7.37	2.95	8.39	SAMD5	sterile alpha motif domain containing 5;	chr6
13	6.19	6.62	8.89	7.02	0.89	1.53	2.16	0.95	1.17E-03	6.50	4.84	3.66	AFAP1-AS1	AFAP1 antisense RNA 1	chr4
14	8.31	9.18	10.98	8.59	1.58	1.56	2.09	1.38	3.39E-02	6.37	3.47	5.24	IQGAP2	IQ motif containing GTPase activating protein 2;	chr5
15	14.09	14.59	16.74	11.04	2.47	2.18	3.30	1.75	1.73E-04	6.27	4.44	51.88	ANXA2P2	annexin A2 pseudogene 2	chr9
16	10.93	11.57	13.40	9.30	1.82	2.17	2.20	1.75	1.43E-03	5.52	3.55	17.09	MIR614; GPRC5A	microRNA 614; G protein-coupled receptor, family C, group 5, member A	chr12
17	8.98	9.62	11.41	7.52	1.53	1.70	2.35	1.13	5.54E-04	5.41	3.45	14.81	GALNT1	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 1 (GalNAc-T1);	chr18
18	6.49	7.55	8.88	5.85	1.33	1.40	1.86	0.43	5.80E-05	5.24	2.51	8.14	HSPA4L	heat shock 70kDa protein 4-like;	chr4
19	7.02	8.06	9.40	6.86	0.69	1.38	2.13	0.61	9.00E-06	5.21	2.53	5.81	CREB3L1	cAMP responsive element binding protein 3-like 1	chr11
20	9.51	9.13	11.81	8.11	1.62	1.38	1.85	0.87	6.79E-04	4.93	6.42	13.00	UGT8	UDP glycosyltransferase 8;	chr4
21	11.86	11.90	14.14	10.06	1.70	1.66	2.77	1.44	5.71E-04	4.83	4.72	16.92	NARS	asparaginyl-tRNA synthetase	chr18
22	6.52	6.69	8.70	6.21	0.68	1.41	1.67	0.62	6.50E-05	4.54	4.03	5.62	KITLG	KIT ligand	chr12
23	10.21	10.33	12.37	8.15	1.71	1.39	2.06	0.83	1.30E-05	4.46	4.10	18.66	ADAM9	ADAM metallopeptidase domain 9;	chr8
24	10.56	11.10	12.69	8.68	1.47	1.55	1.92	0.78	1.00E-05	4.37	3.01	16.13	AHR	aryl hydrocarbon receptor;	chr7
25	8.00	7.79	10.09	7.56	0.78	1.88	1.54	0.74	1.47E-02	4.25	4.92	5.80	DUSP6	dual specificity phosphatase 6	chr12
26	11.51	11.42	13.58	8.53	1.91	2.23	2.92	1.58	3.90E-05	4.21	4.46	33.24	HSPA1B; HSPA1A	heat shock 70kDa protein 1B; heat shock 70kDa protein 1A;	chr6
27	6.23	6.83	8.28	6.54	0.50	0.48	1.60	0.41	4.00E-06	4.14	2.73	3.35	DAPK1	death-associated protein kinase 1;	chr9
28	9.78	10.54	11.80	7.68	2.14	2.02	2.26	0.98	2.43E-04	4.07	2.40	17.39	IFI30	interferon, gamma-inducible protein 30	chr19
29	7.12	6.86	9.14	5.95	0.91	1.29	1.17	0.35	7.43E-07	4.05	4.85	9.12	PBK	PDZ binding kinase;	chr8
30	10.01	9.69	12.00	7.67	1.29	0.99	2.47	0.86	1.00E-06	3.96	4.96	20.12	SUPT4H1	suppressor of Ty 4 homolog 1 (S. cerevisiae)	chr17
31	5.09	4.99	7.07	4.88	1.18	2.20	1.26	0.72	3.01E-03	3.92	4.21	4.55	HTR1D	5-hydroxytryptamine (serotonin) receptor 1D, G protein-coupled;	chr1
32	9.30	9.92	11.27	8.18	1.28	1.46	2.28	0.59	1.02E-04	3.90	2.54	8.53	P4HA1	prolyl 4-hydroxylase, alpha polypeptide I;	chr10
33	6.17	6.02	8.09	5.98	1.26	1.81	2.00	0.52	1.10E-02	3.79	4.20	4.32	SLC6A14	solute carrier family 6 (amino acid transporter), member 14;	chrX
34	7.95	7.91	9.87	6.40	1.12	0.78	1.81	0.55	6.80E-07	3.79	3.89	11.08	SLC39A6	solute carrier family 39 (zinc transporter), member 6	chr18
35	9.52	10.27	11.43	8.60	1.48	1.21	1.93	1.08	3.98E-03	3.75	2.23	7.09	TMEM66	transmembrane protein 66;	chr8
36	6.97	7.56	8.85	7.24	1.07	0.98	2.24	0.73	5.31E-03	3.68	2.45	3.06	PLA2G2A	phospholipase A2, group IIA (platelets, synovial fluid);	chr1
37	8.72	9.13	10.58	7.98	1.01	1.03	1.49	0.82	1.92E-03	3.65	2.73	6.08	SS18	synovial sarcoma translocation, chromosome 18	chr18
38	8.67	9.27	10.50	7.98	0.98	0.90	1.33	0.68	1.73E-04	3.56	2.34	5.74	ASAH1	N-acylsphingosine amidohydrolase (acid ceramidase) 1;	chr8
39	6.82	6.71	8.63	6.66	1.32	1.40	1.46	0.43	9.13E-03	3.51	3.78	3.93	PLA2G16	phospholipase A2, group XVI	chr11
40	8.92	8.89	10.73	7.44	1.09	1.44	1.49	0.51	4.00E-06	3.50	3.56	9.77	CXCL16	chemokine (C-X-C motif) ligand 16;	chr17
41	8.45	8.87	10.24	7.14	1.03	0.84	1.66	0.56	4.00E-06	3.47	2.58	8.59	ARID5B	AT rich interactive domain 5B (MRF1-like);	chr10
42	9.89	8.95	11.68	6.61	1.47	1.77	2.33	0.54	2.73E-08	3.46	6.66	33.66	ANLN	anillin, actin binding protein;	chr7
43	8.94	9.66	10.71	6.30	3.31	2.97	4.28	1.00	6.48E-04	3.43	2.07	21.29	IFI6	interferon, alpha-inducible protein 6;	chr1

44	7.87	8.05	9.64	7.01	1.19	0.57	1.13	0.48	1.15E-04	3.42	3.02	6.20	LY6E	lymphocyte antigen 6 complex, locus E;	chr8
45	7.19	6.98	8.96	6.59	1.25	1.56	2.39	0.89	2.84E-03	3.40	3.94	5.19	IRS2	insulin receptor substrate 2;	chr13
46	11.23	11.57	13.00	8.84	1.68	1.53	2.48	1.47	7.40E-05	3.40	2.70	17.87	TMED10	transmembrane emp24-like trafficking protein 10 (yeast)	chr14
47	8.41	8.59	10.17	7.46	0.97	0.66	1.32	0.49	1.40E-05	3.39	2.98	6.53	GATA6	GATA binding protein 6;	chr18
48	7.79	8.26	9.54	7.82	1.20	0.59	1.86	0.91	2.41E-02	3.37	2.43	3.29	NT5E	5'-nucleotidase, ecto (CD73);	chr6
49	8.77	8.33	10.51	7.04	1.41	1.17	1.63	0.46	8.00E-06	3.33	4.53	11.05	NUSAP1	nucleolar and spindle associated protein 1	chr15
50	7.02	7.17	8.75	6.66	0.79	0.56	0.99	0.34	2.20E-05	3.32	3.00	4.25	CRYZ	crystallin, zeta (quinone reductase);	chr1

HB, high-BCNA tumors; LB, low-BCNA tumors; MSI, microsatellite-instability tumors; N, normal tissue; RMA, bi-weight average signal (Robust Multi-array Average); SD, standard deviation; FDR, false discovery rate; FC, fold change, Chr, chromosome.

Supplementary Table 9: *Top genes (up to 50 entries) which have significant differential expression in both HB and LB tumors compared to normal tissue (FC <-2 or FC >2) and are upregulated (FC >2) in both groups compared to MSI CRCs. Genes listed in descending order on the bases of HB FC increase in gene expression compared to MSI.*

#	HB RMA	LB RMA	MSI RMA	N RMA	HB SD	LB SD	MSI SD	N SD	FDR p-value (All Conditions)	FC HB vs MSI	FC HB vs N	FC LB vs MSI	FC LB vs N	Gene Symbol	Description	Chr
1	14.71	15.24	6.25	8.76	4.00	3.39	5.72	1.24	7.00E-06	351.68	61.88	506.51	89.12	OLFM4	olfactomedin 4;	chr13
2	12.24	11.43	6.90	15.47	2.77	2.99	3.46	3.73	6.73E-04	40.54	-9.37	23.06	-16.48	KRT20	keratin 20;	chr17
3	11.20	9.38	5.97	8.03	2.20	1.69	2.75	1.05	8.00E-06	37.57	9.00	10.63	2.55	CFTR	cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7);	chr7
4	9.56	7.27	5.73	18.35	3.59	3.19	2.52	4.26	3.00E-06	14.28	-443.28	2.92	-2169.20	SLC26A3	solute carrier family 26, member 3;	chr7
5	10.18	12.41	6.70	7.39	3.57	2.24	4.86	2.22	2.19E-03	11.16	6.88	52.50	32.36	DMBT1	deleted in malignant brain tumors 1;	chr10
6	8.63	7.98	5.50	6.29	1.39	0.74	2.65	0.39	3.00E-06	8.77	5.05	5.58	3.21	DACH1	dachshund homolog 1 (Drosophila);	chr13
7	9.32	8.51	6.83	6.80	2.15	1.66	2.54	0.46	1.26E-04	5.62	5.73	3.20	3.25	KIAA1324	KIAA1324;	chr1
8	9.40	8.59	6.99	7.16	1.70	1.40	1.87	0.41	5.00E-06	5.31	4.75	3.02	2.70	DPEP1	dipeptidase 1 (renal);	chr16
9	9.25	8.33	6.99	6.56	1.19	1.76	0.68	0.82	8.37E-08	4.77	6.42	2.53	3.41	CCAT1;	colon cancer associated transcript 1 (non-protein coding);	chr8
10	8.69	8.81	6.52	7.60	1.47	0.97	1.23	0.99	3.27E-04	4.51	2.14	4.90	2.32	CYP2B6	cytochrome P450, family 2, subfamily B, polypeptide 6	chr19
11	8.71	9.16	6.56	15.51	3.19	2.03	2.61	3.84	6.00E-05	4.45	-111.07	6.07	-81.40	CEACAM7	carcinoembryonic antigen-related cell adhesion molecule 7;	chr19
12	9.60	11.42	7.49	12.45	2.06	2.16	1.67	1.98	5.80E-04	4.33	-7.20	15.33	-2.03	FABP1	fatty acid binding protein 1, liver;	chr2
13	16.70	17.63	14.66	19.17	2.00	1.36	2.71	0.97	1.20E-05	4.12	-5.55	7.87	-2.91	MIR4461	microRNA 4461	chr5
14	10.80	10.54	8.76	7.10	2.54	2.47	1.66	0.39	3.00E-06	4.12	13.02	3.45	10.89	TGFBI; LOC100653157; LOC100652886	transforming growth factor, beta-induced, 68kDa; uncharacterized LOC100653157; uncharacterized LOC100652886;	chr5
15	12.66	16.90	10.70	11.03	3.44	2.31	3.77	1.60	3.21E-04	3.88	3.10	73.61	58.75	LYZ	lysozyme	chr12
16	9.41	8.93	7.62	6.93	1.97	1.73	1.64	0.82	6.00E-05	3.48	5.59	2.49	4.00	SPINK1	serine peptidase inhibitor, Kazal type 1;	chr5

17	8.58	8.89	6.92	10.39	2.17	1.27	1.08	2.14	5.15E-03	3.15	-3.50	3.90	-2.82	MUC12	mucin 12, cell surface associated;	chr7
18	10.86	10.76	9.35	7.52	2.57	2.55	2.72	1.55	8.96E-04	2.86	10.17	2.67	9.50	CCL20	chemokine (C-C motif) ligand 20;	chr2
19	10.80	10.59	9.42	9.13	1.28	0.81	1.60	1.28	1.60E-03	2.59	3.17	2.24	2.75	USP12	ubiquitin specific peptidase 12;	chr13
20	9.68	10.24	8.31	12.50	2.24	2.05	0.97	1.62	4.40E-05	2.58	-7.09	3.81	-4.81	IGHD3-16	immunoglobulin heavy diversity 3-16;	chr14
21	10.60	10.86	9.24	8.64	2.06	1.38	2.71	1.70	4.54E-02	2.56	3.87	3.08	4.66	PRSS8	protease, serine, 8	chr16
22	8.54	8.39	7.19	10.31	1.59	1.30	1.82	0.69	2.10E-04	2.54	-3.41	2.30	-3.77	EYA3-IT1	EYA3 intronic transcript 1 (non-protein coding)	chr1
23	7.07	6.96	5.89	9.36	1.91	1.19	1.16	2.84	2.98E-04	2.27	-4.89	2.09	-5.30	LYPD8	LY6/PLAUR domain containing 8	chr1
24	9.40	9.26	8.23	7.96	1.09	0.68	1.69	0.39	2.10E-05	2.25	2.71	2.04	2.45	HOXB9	homeobox B9;	chr17
25	10.70	10.58	9.55	12.36	1.72	1.56	1.34	1.62	3.76E-03	2.22	-3.17	2.04	-3.44	IGHD2-21	immunoglobulin heavy diversity 2-21;	chr14
26	9.58	10.13	8.47	12.06	1.91	1.50	1.86	1.25	4.70E-05	2.16	-5.57	3.15	-3.82	TRAJ59	T cell receptor alpha joining 59 (non-functional)	chr14
27	9.66	9.66	8.59	7.54	1.39	1.43	0.98	0.94	1.30E-04	2.10	4.35	2.10	4.36	TGFBR2	transforming growth factor, beta receptor II (70/80kDa);	chr3
28	5.24	5.41	4.18	6.51	1.09	0.79	0.55	0.99	4.90E-05	2.08	-2.43	2.34	-2.15	MIR514B	microRNA 514b	chrX
29	8.11	8.24	7.07	9.37	1.04	1.45	0.88	1.49	8.02E-03	2.07	-2.40	2.25	-2.20	CDHR5	cadherin-related family member 5;	chr11
30	5.38	5.37	4.33	6.91	0.94	0.82	0.54	0.93	5.66E-07	2.07	-2.89	2.05	-2.91	TRAJ2	T cell receptor alpha joining 2 (non-functional)	chr14
31	8.85	9.63	7.81	10.87	2.37	1.59	0.95	1.58	5.45E-03	2.06	-4.05	3.51	-2.37	IGHD3-10	immunoglobulin heavy diversity 3-10;	chr14
32	8.09	8.72	7.07	10.09	1.67	1.93	1.99	1.41	9.71E-04	2.03	-4.00	3.14	-2.59	IGHV3-64	immunoglobulin heavy variable 3-64;	chr14

HB, high-BCNA tumors; LB, low-BCNA tumors; MSI, microsatellite-instability tumors; N, normal tissue; RMA, bi-weight average signal (Robust Multi-array Average); SD, standard deviation; FDR, false discovery rate; FC, fold change, Chr, chromosome.

Supplementary Table 10*: Top genes (up to 50 entries) which have significant differential expression in both HB and MSI tumors compared to normal tissue (FC <-2 or FC >2) and are upregulated (FC >2) in both groups compared to LB CRCs. Genes listed in descending order on the bases of HB FC increase in gene expression compared to LB.*

#	HB RMA	LB RMA	MSI RMA	N RMA	HB SD	LB SD	MSI SD	N SD	FDR p-value (All Conditions)	FC HB vs LB	FC HB vs N	FC MSI vs LB	FC MSI vs N	Gene Symbol	Description	Chr
1	12.85	9.96	11.37	7.21	3.07	2.44	2.65	1.13	7.10E-07	7.43	50.04	2.66	17.90	SCD	stearoyl-CoA desaturase (delta-9-desaturase);	chr10
2	12.76	10.59	12.02	7.19	2.33	1.90	2.21	1.20	6.30E-08	4.51	47.78	2.69	28.51	CSE1L	CSE1 chromosome segregation 1-like (yeast);	chr20
3	16.06	14.09	16.69	8.87	3.22	2.64	3.81	1.16	3.40E-08	3.91	146.18	6.06	226.13	SLC12A2	solute carrier family 12 (sodium/potassium/chloride transporters), member 2;	chr5
4	11.81	9.88	11.87	8.07	1.70	1.51	1.77	0.71	1.39E-08	3.82	13.43	3.97	13.96	TPX2	TPX2, microtubule-associated, homolog (Xenopus laevis);	chr20
5	13.12	11.27	12.87	9.22	1.78	1.48	2.06	0.99	2.00E-06	3.61	14.90	3.04	12.52	PPA1	pyrophosphatase (inorganic) 1;	chr10
6	9.54	7.73	10.47	6.71	1.72	1.55	2.17	0.55	4.59E-07	3.50	7.11	6.70	13.62	SLC7A5	solute carrier family 7 (amino acid transporter light chain, L system), member 5;	chr16
7	8.95	7.18	8.78	6.65	1.50	1.16	1.39	0.56	7.00E-06	3.41	4.93	3.03	4.38	AGMAT	agmatine ureohydrolase (agmatinase);	chr1
8	11.64	9.90	11.60	7.28	1.94	1.60	2.12	1.16	2.00E-06	3.34	20.42	3.26	19.93	LRPPRC	leucine-rich	chr2

															pentatricopeptide repeat containing;	
9	9.51	7.80	8.83	6.80	1.70	1.34	1.25	0.66	5.00E-06	3.28	6.55	2.04	4.07	PM20D2	peptidase M20 domain containing 2;	chr6
10	9.67	7.99	9.69	6.22	1.54	1.18	1.31	0.49	5.07E-09	3.22	10.92	3.25	11.02	CDCA7	cell division cycle associated 7;	chr2
11	11.79	10.10	11.52	9.03	2.07	1.51	2.30	1.53	2.29E-04	3.22	6.74	2.68	5.61	SDC4	syndecan 4;	chr20
12	9.50	7.84	10.78	6.94	2.15	1.53	2.08	0.61	1.20E-05	3.15	5.90	7.65	14.33	CLDN1	claudin 1;	chr3
13	8.90	7.31	8.51	6.47	1.23	1.38	1.96	0.54	5.00E-06	3.01	5.38	2.30	4.11	UTP20	UTP20, small subunit (SSU) processome component, homolog (yeast)	chr12
14	14.47	12.89	14.46	11.40	2.39	1.67	2.81	2.33	6.67E-03	2.99	8.45	2.96	8.36	CDH1	cadherin 1, type 1, E-cadherin (epithelial);	chr16
15	9.03	7.45	8.47	7.35	1.09	0.70	0.99	0.47	2.00E-06	2.99	3.20	2.03	2.18	GSS	glutathione synthetase;	chr20
16	11.57	10.00	11.53	9.16	1.46	1.18	1.54	0.93	7.20E-05	2.96	5.29	2.89	5.15	ARPC1A	actin related protein 2/3 complex, subunit 1A, 41kDa;	chr7
17	9.75	8.19	9.61	7.03	1.68	1.30	1.19	0.42	8.27E-07	2.95	6.63	2.67	6.01	EPHB2	EPH receptor B2;	chr1
18	12.18	10.62	11.81	8.38	1.76	1.61	1.94	0.58	6.48E-08	2.95	13.90	2.29	10.79	SOX9	SRY (sex determining region Y)-box 9	chr17
19	12.45	10.90	12.19	8.31	2.09	1.69	2.51	1.20	6.19E-07	2.94	17.70	2.45	14.72	EIF1AX	eukaryotic translation initiation factor 1A, X-linked; eukaryotic translation initiation factor 1A, X-chromosomal-like;	chrX
20	12.55	11.04	13.60	8.23	2.26	2.56	3.38	1.04	2.00E-06	2.85	19.94	5.93	41.49	RCN1P2	reticulocalbin 1, EF-hand calcium binding domain pseudogene 2	chr13
21	11.15	9.73	10.99	8.39	0.97	0.95	1.12	0.80	5.01E-08	2.68	6.75	2.39	6.02	RPS21	ribosomal protein S21;	chr20
22	11.07	9.65	11.49	7.17	1.69	1.34	2.00	0.69	4.10E-08	2.67	14.93	3.58	20.06	IARS	isoleucyl-tRNA synthetase;	chr9
23	11.10	9.69	10.73	8.11	1.59	0.94	1.88	0.76	2.00E-06	2.66	7.98	2.04	6.14	SMC1A	structural maintenance of chromosomes 1A;	chrX
24	11.78	10.37	12.17	8.04	1.84	1.26	2.24	1.04	2.00E-06	2.65	13.36	3.48	17.58	NUDT21	nudix (nucleoside diphosphate linked moiety X)-type motif 21;	chr16
25	12.21	10.82	12.91	7.66	1.76	1.42	2.46	1.00	2.73E-08	2.63	23.52	4.25	37.96	IFITM1	interferon induced transmembrane protein 1	chr11
26	12.08	10.69	12.05	7.19	2.26	1.68	2.26	1.25	2.30E-07	2.62	29.61	2.56	28.99	PRKDC	protein kinase, DNA-activated, catalytic polypeptide;	chr8
27	12.21	10.82	11.96	9.07	1.51	1.13	1.99	0.84	7.07E-08	2.60	8.78	2.20	7.41	EIF2S2	eukaryotic translation initiation factor 2, subunit 2 beta, 38kDa;	chr20
28	12.26	10.89	12.07	8.47	1.95	1.61	2.20	1.70	1.41E-04	2.59	13.89	2.27	12.15	NCKAP1	NCK-associated protein 1;	chr2
29	12.48	11.11	12.93	8.60	2.15	1.64	2.25	1.26	1.70E-05	2.59	14.78	3.52	20.11	STT3B	STT3B, subunit of the oligosaccharyltransferase complex (catalytic);	chr3
30	10.31	8.94	11.03	7.18	1.52	1.26	1.83	0.48	6.44E-08	2.59	8.81	4.26	14.51	MTHFD1	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1, methenyltetrahydrofolate cyclohydrolase, formyltetrahydrofolate synthetase;	chr14
31	11.72	10.36	12.64	8.51	1.42	1.55	2.28	0.97	4.08E-07	2.57	9.26	4.85	17.46	SUPT16H	suppressor of Ty 16 homolog (S. cerevisiae);	chr14
32	10.29	8.92	10.98	8.12	2.22	1.51	1.70	1.15	2.54E-03	2.57	4.49	4.16	7.25	ECH1	enoyl CoA hydratase 1, peroxisomal	chr19
33	11.32	9.99	11.55	7.94	1.53	1.16	2.02	0.78	1.23E-07	2.51	10.35	2.95	12.18	SSRP1	structure specific recognition protein 1	chr11
34	11.02	9.70	11.05	7.77	1.92	1.36	1.75	0.68	4.00E-06	2.49	9.50	2.54	9.71	PSMB3	proteasome (prosome, macropain) subunit, beta	chr17

															type, 3;	
35	16.87	15.55	17.31	11.31	2.85	1.89	3.89	1.92	1.60E-05	2.49	47.12	3.37	63.80	EEF2	eukaryotic translation elongation factor 2	chr19
36	9.63	8.34	10.26	6.75	1.71	1.45	1.79	0.94	3.00E-06	2.45	7.35	3.79	11.36	GPSM2	G-protein signaling modulator 2;	chr1
37	11.04	9.75	11.47	7.10	1.86	1.91	1.98	0.75	1.75E-07	2.44	15.32	3.30	20.71	XPOT	exportin, tRNA	chr12
38	8.61	7.32	9.00	6.72	1.52	0.93	1.39	0.28	1.00E-05	2.44	3.72	3.20	4.88	EXOSC5	exosome component 5	chr19
39	9.35	8.08	9.50	6.63	1.31	0.90	1.49	0.45	1.62E-07	2.40	6.57	2.67	7.30	NUP205	nucleoporin 205kDa;	chr7
40	9.97	8.72	9.73	7.58	1.96	1.46	1.43	0.70	3.30E-05	2.39	5.25	2.01	4.42	SQLE	squalene epoxidase;	chr8
41	11.44	10.18	11.56	7.29	1.90	1.97	2.15	0.86	4.26E-08	2.39	17.71	2.61	19.35	TOP2A	topoisomerase (DNA) II alpha 170kDa;	chr17
42	10.42	9.16	11.35	6.27	3.33	2.69	3.90	0.68	2.20E-05	2.38	17.76	4.56	33.99	LGR5	leucine-rich repeat containing G protein-coupled receptor 5	chr12
43	10.52	9.28	11.30	7.07	2.28	2.14	2.28	1.15	4.20E-05	2.37	10.96	4.06	18.78	CYP51A1	cytochrome P450, family 51, subfamily A, polypeptide 1;	chr7
44	10.12	8.87	10.51	6.34	1.83	1.56	1.63	0.73	1.78E-07	2.37	13.70	3.13	18.05	CDK4	cyclin-dependent kinase 4	chr12
45	10.96	9.72	11.15	7.85	2.49	1.84	1.94	0.76	2.40E-05	2.37	8.64	2.68	9.80	CYP2S1	cytochrome P450, family 2, subfamily S, polypeptide 1	chr19
46	11.59	10.35	12.17	7.65	1.80	1.48	2.73	1.07	2.00E-06	2.36	15.36	3.51	22.82	HSP90AB3P	heat shock protein 90kDa alpha (cytosolic), class B member 3, pseudogene	chr4
47	8.53	7.30	8.64	6.99	1.11	0.89	1.22	0.50	1.30E-05	2.35	2.91	2.53	3.13	HIBADH	3-hydroxyisobutyrate dehydrogenase;	chr7
48	10.49	9.28	10.45	7.17	1.59	1.13	1.50	0.66	9.10E-09	2.32	10.02	2.25	9.75	MYC	v-myc myelocytomatosis viral oncogene homolog (avian);	chr8
49	9.01	7.83	10.00	6.48	1.28	1.43	1.76	0.48	1.58E-07	2.27	5.79	4.52	11.53	CENPF	centromere protein F, 350/400kDa; centromere protein F, 350/400kDa (mitosin);	chr1
50	12.64	11.49	12.50	8.05	2.42	2.01	2.55	1.18	4.30E-07	2.22	24.01	2.01	21.72	RPN2	ribophorin II;	chr20

HB, high-BCNA tumors; LB, low-BCNA tumors; MSI, microsatellite-instability tumors; N, normal tissue; RMA, bi-weight average signal (Robust Multi-array Average); SD, standard deviation; FDR, false discovery rate; FC, fold change, Chr, chromosome.

Supplementary Table 11: *Top genes (up to 50 entries) which have significant differential expression in both LB and MSI tumors compared to normal tissue (FC <-2 or FC >2) and are upregulated (FC >2) in both groups compared to HB CRCs. Genes listed in descending order on the bases of LB FC increase in gene expression compared to HB.*

#	HB RMA	LB RMA	MSI RMA	N RMA	HB SD	LB SD	MSI SD	N SD	FDR p-value (All Conditions)	FC LB vs HB	FC LB vs N	FC MSI vs HB	FC MSI vs N	Gene Symbol	Description	Chr
1	7.58	10.32	11.18	6.98	1.82	2.71	3.74	1.22	2.17E-04	6.67	10.16	12.13	18.49	TFF1	trefoil factor 1	7.58
2	5.75	8.34	13.65	5.67	1.06	2.97	4.68	0.93	2.73E-07	6.02	6.38	238.78	252.85	REG4	regenerating islet-derived family, member 4	5.75
3	7.74	10.21	10.61	6.89	2.65	3.20	3.83	1.23	4.60E-03	5.52	9.92	7.29	13.11	POSTN	periostin, osteoblast specific factor	7.74
4	7.06	9.46	8.72	6.83	2.28	2.02	3.01	1.35	1.57E-02	5.27	6.21	3.16	3.73	PI3	peptidase inhibitor 3, skin-derived	7.06
5	7.20	9.12	11.10	7.09	1.74	2.67	2.84	0.84	2.85E-03	3.78	4.08	14.90	16.11	MUC5B	mucin 5B, oligomeric mucus/gel-forming	7.20
6	7.01	8.86	8.50	5.97	2.67	2.82	3.63	0.28	4.89E-04	3.60	7.43	2.80	5.78	BGN	biglycan	7.01
7	10.52	11.97	12.42	8.54	1.63	1.70	2.53	1.15	6.60E-05	2.73	10.80	3.72	14.74	HIF1A	hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)	10.52

8	6.07	7.39	8.95	5.89	1.28	1.31	1.60	0.42	2.04E-04	2.49	2.84	7.37	8.39	SAMD5	sterile alpha motif domain containing 5	6.07
9	9.01	10.31	12.28	8.22	0.99	1.42	1.83	1.02	1.40E-05	2.46	4.25	9.67	16.72	CD55	CD55 molecule, decay accelerating factor for complement (Cromer blood group)	9.01
10	8.61	9.91	10.40	7.22	2.44	2.99	3.14	0.76	5.63E-04	2.45	6.47	3.45	9.08	COL3A1; MIR3606	collagen, type III, alpha 1; microRNA 3606	8.61
11	7.75	9.04	9.73	7.63	1.32	1.51	1.54	1.05	1.20E-02	2.44	2.66	3.94	4.31	EMP3	epithelial membrane protein 3	7.75
12	7.16	8.37	10.42	6.62	2.48	4.89	3.50	0.54	2.90E-04	2.31	3.36	9.59	13.95	SPP1	secreted phosphoprotein 1	7.16
13	6.76	7.81	7.81	6.71	1.06	1.63	1.17	0.30	1.10E-02	2.08	2.14	2.07	2.13	FCER1G	Fc fragment of IgE, high affinity I, receptor for; gamma polypeptide	6.76
14	6.49	7.55	8.88	5.85	1.33	1.40	1.86	0.43	5.80E-05	2.08	3.24	5.24	8.14	HSPA4L	heat shock 70kDa protein 4-like	6.49
15	8.71	9.77	10.53	6.84	2.38	2.65	2.98	0.39	5.80E-05	2.07	7.58	3.52	12.89	COL1A1	collagen, type I, alpha 1	8.71
16	6.35	7.39	11.41	6.29	1.07	1.29	2.60	0.57	4.00E-06	2.06	2.14	33.28	34.72	SERPINB5	serpin peptidase inhibitor, clade B (ovalbumin), member 5	6.35
17	7.02	8.06	9.40	6.86	0.69	1.38	2.13	0.61	9.00E-06	2.06	2.30	5.21	5.81	CREB3L1	cAMP responsive element binding protein 3-like 1	7.02
18	9.52	10.56	11.06	8.46	1.11	1.17	2.04	1.24	5.15E-03	2.06	4.31	2.90	6.06	ROCK1	Rho-associated, coiled-coil containing protein kinase 1	9.52
19	10.01	11.04	11.32	8.88	1.59	1.73	1.87	1.00	6.86E-04	2.04	4.47	2.48	5.42	CD68	CD68 molecule	10.01

HB, high-BCNA tumors; LB, low-BCNA tumors; MSI, microsatellite-instability tumors; N, normal tissue; RMA, bi-weight average signal (Robust Multi-array Average); SD, standard deviation; FDR, false discovery rate; FC, fold change, Chr, chromosome.

Supplementary Table 12: *Top genes (up to 50 entries) which have significant differential expression in HB tumors compared to normal tissue (FC <-2 or FC >2) and are downregulated (FC <-2) in the HB group compared to LB and MSI tumors. Genes listed in ascending order on the bases of HB FC decrease in gene expression compared to LB.*

#	HB RMA	LB RMA	MSI RMA	N RMA	HB SD	LB SD	MSI SD	N SD	FDR p-value (All Conditions)	FC HB vs LB	FC HB vs MSI	FC HB vs N	Gene Symbol	Description	Chr
1	7.01	8.86	8.50	5.97	2.67	2.82	3.63	0.28	4.89E-04	-3.60	-2.80	2.06	BGN	biglycan;	chrX
2	10.52	11.97	12.42	8.54	1.63	1.70	2.53	1.15	6.60E-05	-2.73	-3.72	3.96	HIF1A	hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor);	chr14
3	6.70	8.14	8.54	9.32	2.01	1.30	1.40	2.37	2.79E-02	-2.70	-3.57	-6.14	GCNT3	glucosaminyl (N-acetyl) transferase 3, mucin type;	chr15
4	8.61	9.91	10.40	7.22	2.44	2.99	3.14	0.76	5.63E-04	-2.45	-3.45	2.64	COL3A1; MIR3606	collagen, type III, alpha 1; microRNA 3606;	chr2
5	8.71	9.77	10.53	6.84	2.38	2.65	2.98	0.39	5.80E-05	-2.07	-3.52	3.66	COL1A1	collagen, type I, alpha 1;	chr17
6	9.52	10.56	11.06	8.46	1.11	1.17	2.04	1.24	5.15E-03	-2.06	-2.90	2.09	ROCK1	Rho-associated, coiled-coil containing protein kinase 1;	chr18
7	10.01	11.04	11.32	8.88	1.59	1.73	1.87	1.00	6.86E-04	-2.04	-2.48	2.18	CD68	CD68 molecule;	chr17

HB, high-BCNA tumors; LB, low-BCNA tumors; MSI, microsatellite-instability tumors; N, normal tissue; RMA, bi-weight average signal (Robust Multi-array Average); SD, standard deviation; FDR, false discovery rate; FC, fold change, Chr, chromosome.

Supplementary Table 13: *Top genes (up to 50 entries) which have significant differential expression in LB tumors compared to normal tissue (FC <-2 or FC >2) and are downregulated (FC <-2) in the LB group compared to HB and MSI tumors. Genes listed in ascending order on the bases of LB FC decrease in gene expression compared to HB.*

#	HB RMA	LB RMA	MSI RMA	N RMA	HB SD	LB SD	MSI SD	N SD	FDR p-value (All Conditions)	FC LB vs HB	FC LB vs MSI	FC LB vs N	Gene Symbol	Description	Chr
1	12.85	9.96	11.37	7.21	3.07	2.44	2.65	1.13	7.10E-07	-7.43	-2.66	6.73	SCD	stearoyl-CoA desaturase (delta-9-desaturase);	chr10
2	12.76	10.59	12.02	7.19	2.33	1.90	2.21	1.20	6.30E-08	-4.51	-2.69	10.60	CSE1L	CSE1 chromosome segregation 1-like (yeast);	chr20
3	16.06	14.09	16.69	8.87	3.22	2.64	3.81	1.16	3.40E-08	-3.91	-6.06	37.34	SLC12A2	solute carrier family 12 (sodium/potassium/chloride transporters), member 2;	chr5
4	11.81	9.88	11.87	8.07	1.70	1.51	1.77	0.71	1.39E-08	-3.82	-3.97	3.52	TPX2	TPX2, microtubule-associated, homolog (Xenopus laevis);	chr20
5	13.12	11.27	12.87	9.22	1.78	1.48	2.06	0.99	2.00E-06	-3.61	-3.04	4.12	PPA1	pyrophosphatase (inorganic) 1;	chr10
6	9.54	7.73	10.47	6.71	1.72	1.55	2.17	0.55	4.59E-07	-3.50	-6.70	2.03	SLC7A5	solute carrier family 7 (amino acid transporter light chain, L system), member 5;	chr16
7	11.64	9.90	11.60	7.28	1.94	1.60	2.12	1.16	2.00E-06	-3.34	-3.26	6.11	LRPPRC	leucine-rich pentatricopeptide repeat containing;	chr2
8	9.67	7.99	9.69	6.22	1.54	1.18	1.31	0.49	5.07E-09	-3.22	-3.25	3.39	CDCA7	cell division cycle associated 7;	chr2
9	11.79	10.10	11.52	9.03	2.07	1.51	2.30	1.53	2.29E-04	-3.22	-2.68	2.09	SDC4	syndecan 4;	chr20
10	14.47	12.89	14.46	11.40	2.39	1.67	2.81	2.33	6.67E-03	-2.99	-2.96	2.82	CDH1	cadherin 1, type 1, E-cadherin (epithelial);	chr16
11	9.75	8.19	9.61	7.03	1.68	1.30	1.19	0.42	8.27E-07	-2.95	-2.67	2.25	EPHB2	EPH receptor B2;	chr1
12	12.18	10.62	11.81	8.38	1.76	1.61	1.94	0.58	6.48E-08	-2.95	-2.29	4.72	SOX9	SRY (sex determining region Y)-box 9	chr17
13	12.45	10.90	12.19	8.31	2.09	1.69	2.51	1.20	6.19E-07	-2.94	-2.45	6.01	EIF1AX; LOC101060318	eukaryotic translation initiation factor 1A, X-linked; eukaryotic translation initiation factor 1A, X-chromosomal-like;	chrX
14	12.55	11.04	13.60	8.23	2.26	2.56	3.38	1.04	2.00E-06	-2.85	-5.93	7.00	RCN1P2	reticulocalbin 1, EF-hand calcium binding domain pseudogene 2	chr13
15	11.15	9.73	10.99	8.39	0.97	0.95	1.12	0.80	5.01E-08	-2.68	-2.39	2.52	RPS21	ribosomal protein S21;	chr20
16	11.07	9.65	11.49	7.17	1.69	1.34	2.00	0.69	4.10E-08	-2.67	-3.58	5.60	IARS	isoleucyl-tRNA synthetase;	chr9
17	11.10	9.69	10.73	8.11	1.59	0.94	1.88	0.76	2.00E-06	-2.66	-2.04	3.00	SMC1A	structural maintenance of chromosomes 1A;	chrX
18	11.78	10.37	12.17	8.04	1.84	1.26	2.24	1.04	2.00E-06	-2.65	-3.48	5.04	NUDT21	nudix (nucleoside diphosphate linked moiety X)-type motif 21;	chr16
19	12.21	10.82	12.91	7.66	1.76	1.42	2.46	1.00	2.73E-08	-2.63	-4.25	8.94	IFITM1	interferon induced transmembrane protein 1	chr11
20	12.08	10.69	12.05	7.19	2.26	1.68	2.26	1.25	2.30E-07	-2.62	-2.56	11.31	PRKDC	protein kinase, DNA-activated, catalytic polypeptide;	chr8
21	12.21	10.82	11.96	9.07	1.51	1.13	1.99	0.84	7.07E-08	-2.60	-2.20	3.37	EIF2S2	eukaryotic translation initiation factor 2, subunit 2 beta, 38kDa;	chr20
22	12.26	10.89	12.07	8.47	1.95	1.61	2.20	1.70	1.41E-04	-2.59	-2.27	5.35	NCKAP1	NCK-associated protein 1;	chr2
23	12.48	11.11	12.93	8.60	2.15	1.64	2.25	1.26	1.70E-05	-2.59	-3.52	5.71	STT3B	STT3B, subunit of the oligosaccharyltransferase complex (catalytic);	chr3
24	10.31	8.94	11.03	7.18	1.52	1.26	1.83	0.48	6.44E-08	-2.59	-4.26	3.40	MTHFD1	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1,	chr14

														methenyltetrahydrofolate cyclohydrolase, formyltetrahydrofolate synthetase;	
25	11.72	10.36	12.64	8.51	1.42	1.55	2.28	0.97	4.08E-07	-2.57	-4.85	3.60	SUPT16H	suppressor of Ty 16 homolog (S. cerevisiae);	chr14
26	11.32	9.99	11.55	7.94	1.53	1.16	2.02	0.78	1.23E-07	-2.51	-2.95	4.12	SSRP1	structure specific recognition protein 1	chr11
27	11.02	9.70	11.05	7.77	1.92	1.36	1.75	0.68	4.00E-06	-2.49	-2.54	3.82	PSMB3	proteasome (prosome, macropain) subunit, beta type, 3;	chr17
28	16.87	15.55	17.31	11.31	2.85	1.89	3.89	1.92	1.60E-05	-2.49	-3.37	18.91	EEF2	eukaryotic translation elongation factor 2	chr19
29	9.63	8.34	10.26	6.75	1.71	1.45	1.79	0.94	3.00E-06	-2.45	-3.79	3.00	GPSM2	G-protein signaling modulator 2;	chr1
30	11.04	9.75	11.47	7.10	1.86	1.91	1.98	0.75	1.75E-07	-2.44	-3.30	6.27	XPOT	exportin, tRNA	chr12
31	9.35	8.08	9.50	6.63	1.31	0.90	1.49	0.45	1.62E-07	-2.40	-2.67	2.73	NUP205	nucleoporin 205kDa;	chr7
32	9.97	8.72	9.73	7.58	1.96	1.46	1.43	0.70	3.30E-05	-2.39	-2.01	2.19	SQLE	squalene epoxidase;	chr8
33	11.44	10.18	11.56	7.29	1.90	1.97	2.15	0.86	4.26E-08	-2.39	-2.61	7.41	TOP2A	topoisomerase (DNA) II alpha 170kDa;	chr17
34	10.42	9.16	11.35	6.27	3.33	2.69	3.90	0.68	2.20E-05	-2.38	-4.56	7.46	LGR5	leucine-rich repeat containing G protein-coupled receptor 5	chr12
35	10.52	9.28	11.30	7.07	2.28	2.14	2.28	1.15	4.20E-05	-2.37	-4.06	4.63	CYP51A1	cytochrome P450, family 51, subfamily A, polypeptide 1;	chr7
36	10.12	8.87	10.51	6.34	1.83	1.56	1.63	0.73	1.78E-07	-2.37	-3.13	5.77	CDK4	cyclin-dependent kinase 4	chr12
37	10.96	9.72	11.15	7.85	2.49	1.84	1.94	0.76	2.40E-05	-2.37	-2.68	3.65	CYP2S1	cytochrome P450, family 2, subfamily S, polypeptide 1	chr19
38	11.59	10.35	12.17	7.65	1.80	1.48	2.73	1.07	2.00E-06	-2.36	-3.51	6.50	HSP90AB3P	heat shock protein 90kDa alpha (cytosolic), class B member 3, pseudogene	chr4
39	10.49	9.28	10.45	7.17	1.59	1.13	1.50	0.66	9.10E-09	-2.32	-2.25	4.33	MYC	v-myc myelocytomatosis viral oncogene homolog (avian);	chr8
40	9.01	7.83	10.00	6.48	1.28	1.43	1.76	0.48	1.58E-07	-2.27	-4.52	2.55	CENPF	centromere protein F, 350/400kDa; centromere protein F, 350/400kDa (mitosin);	chr1
41	12.64	11.49	12.50	8.05	2.42	2.01	2.55	1.18	4.30E-07	-2.22	-2.01	10.83	RPN2	ribophorin II;	chr20
42	11.75	10.61	12.93	9.35	1.48	1.54	2.16	1.17	6.80E-05	-2.21	-5.00	2.39	MIR622	microRNA 622	chr13
44	9.64	8.50	10.83	7.40	1.52	1.03	2.09	0.55	7.00E-06	-2.21	-5.03	2.15	IDH2	isocitrate dehydrogenase 2 (NADP+), mitochondrial;	chr15
44	13.55	12.41	14.73	8.93	2.45	1.56	3.12	1.28	4.00E-06	-2.20	-5.00	11.17	PLP2	proteolipid protein 2 (colonic epithelium-enriched);	chrX
45	10.76	9.63	10.95	7.40	1.59	1.25	1.81	0.57	2.04E-07	-2.20	-2.50	4.66	TRIM28	tripartite motif containing 28	chr19
46	10.48	9.36	11.30	8.08	1.52	1.20	1.88	0.77	1.70E-05	-2.18	-3.85	2.41	PSMC2	proteasome (prosome, macropain) 26S subunit, ATPase, 2;	chr7
47	9.98	8.86	10.02	7.50	1.50	1.11	1.34	0.68	4.00E-06	-2.18	-2.24	2.55	BRCC3	BRCA1/BRCA2-containing complex, subunit 3;	chrX
48	12.41	11.28	13.42	9.14	1.79	1.74	2.57	1.26	5.30E-05	-2.18	-4.41	4.43	KRT18	keratin 18	chr12
49	10.84	9.71	11.02	8.22	1.60	1.29	1.67	1.07	4.60E-05	-2.18	-2.47	2.83	VPS35	vacuolar protein sorting 35 homolog (S. cerevisiae);	chr16
50	11.13	10.02	11.10	7.75	1.99	1.52	2.06	0.82	1.00E-06	-2.16	-2.11	4.82	PDZD8	PDZ domain containing 8;	chr10

HB, high-BCNA tumors; LB, low-BCNA tumors; MSI, microsatellite-instability tumors; N, normal tissue; RMA, bi-weight average signal (Robust Multi-array Average); SD, standard deviation; FDR, false discovery rate; FC, fold change, Chr, chromosome.

Supplementary Table 14: *Top genes (up to 50 entries) which have significant differential expression in MSI tumors compared to normal tissue (FC <-2 or FC >2) and are downregulated (FC <-2) in the MSI group compared to HB and LB tumors. Genes listed in ascending order on the bases of MSI FC decrease in gene expression compared to HB.*

#	HB RMA	LB RMA	MSI RMA	N RMA	HB SD	LB SD	MSI SD	N SD	FDR p-value (All Conditions)	FC MSI vs HB	FC MSI vs LB	FC MSI vs N	Gene Symbol	Description	Chr
1	14.71	15.24	6.25	8.76	4.00	3.39	5.72	1.24	7.00E-06	-351.68	-506.51	-5.68	OLFM4	olfactomedin 4;	chr13
2	12.24	11.43	6.90	15.47	2.77	2.99	3.46	3.73	6.73E-04	-40.54	-23.06	-380.04	KRT20	keratin 20;	chr17
3	11.20	9.38	5.97	8.03	2.20	1.69	2.75	1.05	8.00E-06	-37.57	-10.63	-4.17	CFTR	cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7);	chr7
4	9.56	7.27	5.73	18.35	3.59	3.19	2.52	4.26	3.00E-06	-14.28	-2.92	-6330.50	SLC26A3	solute carrier family 26, member 3;	chr7
5	9.97	9.56	6.68	8.58	1.48	1.29	1.25	1.44	2.35E-03	-9.81	-7.37	-3.72	ATP10B	ATPase, class V, type 10B;	chr5
6	11.37	9.64	8.47	9.49	1.62	1.51	2.16	1.23	1.41E-04	-7.48	-2.26	-2.03	VIL1	villin 1;	chr2
7	8.06	7.26	5.18	8.98	2.01	2.78	1.68	2.57	1.19E-02	-7.32	-4.21	-13.85	MEP1A	meprin A, alpha (PABA peptide hydrolase);	chr6
8	9.83	9.30	7.49	9.32	1.81	1.69	0.90	1.54	2.65E-02	-5.08	-3.52	-3.57	IL2RG	interleukin 2 receptor, gamma;	chrX
9	8.26	7.86	6.00	8.95	1.70	1.63	1.26	2.45	8.82E-03	-4.76	-3.63	-7.69	LRRC19	leucine rich repeat containing 19;	chr9
10	8.69	8.81	6.52	7.60	1.47	0.97	1.23	0.99	3.27E-04	-4.51	-4.90	-2.11	CYP2B6	cytochrome P450, family 2, subfamily B, polypeptide 6	chr19
11	8.71	9.16	6.56	15.51	3.19	2.03	2.61	3.84	6.00E-05	-4.45	-6.07	-494.03	CEACAM7	carcinoembryonic antigen-related cell adhesion molecule 7;	chr19
12	9.60	11.42	7.49	12.45	2.06	2.16	1.67	1.98	5.80E-04	-4.33	-15.33	-31.14	FABP1	fatty acid binding protein 1, liver;	chr2
13	8.03	7.05	5.98	7.75	1.14	1.31	0.96	0.94	3.13E-03	-4.12	-2.09	-3.41	TFCP2L1	transcription factor CP2-like 1;	chr2
14	16.70	17.63	14.66	19.17	2.00	1.36	2.71	0.97	1.20E-05	-4.12	-7.87	-22.87	MIR4461	microRNA 4461	chr5
15	10.80	10.54	8.76	7.10	2.54	2.47	1.66	0.39	3.00E-06	-4.12	-3.45	3.16	TGFBI	transforming growth factor, beta-induced, 68kDa;	chr5
16	9.83	9.22	7.84	10.28	1.32	1.24	1.03	1.20	9.70E-03	-3.97	-2.61	-5.42	IGHD2-15	immunoglobulin heavy diversity 2-15;	chr14
17	7.39	6.78	5.53	7.16	1.05	1.35	1.34	0.85	9.62E-03	-3.61	-2.38	-3.10	A1CF	APOBEC1 complementation factor;	chr10
18	8.18	7.60	6.37	8.94	1.58	1.86	0.92	1.84	1.44E-02	-3.50	-2.34	-5.93	SULT1B1	sulfotransferase family, cytosolic, 1B, member 1;	chr4
19	10.24	9.71	8.49	11.08	1.49	1.65	0.91	1.85	4.98E-03	-3.37	-2.34	-6.05	RNY5	RNA, Ro-associated Y5	chr7
20	8.58	8.89	6.92	10.39	2.17	1.27	1.08	2.14	5.15E-03	-3.15	-3.90	-11.03	MUC12	mucin 12, cell surface associated;	chr7
21	6.57	8.57	4.99	9.16	2.73	2.77	1.41	3.24	1.19E-02	-2.99	-11.93	-18.04	UGT2B17	UDP glucuronosyltransferase 2 family, polypeptide B17;	chr4
22	10.86	10.76	9.35	7.52	2.57	2.55	2.72	1.55	8.96E-04	-2.86	-2.67	3.56	CCL20	chemokine (C-C motif) ligand 20;	chr2
23	7.79	7.60	6.30	8.68	1.05	1.12	1.40	0.96	2.65E-04	-2.82	-2.46	-5.23	MIR4437	microRNA 4437	chr2
24	7.08	7.00	5.64	6.98	0.57	0.71	0.81	0.46	1.02E-02	-2.71	-2.57	-2.53	HMGN5	high mobility group nucleosome binding domain 5;	chrX
25	7.32	7.28	5.89	6.97	0.33	0.22	0.90	0.20	5.00E-06	-2.70	-2.63	-2.12	MLH1	mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli);	chr3
26	9.68	10.24	8.31	12.50	2.24	2.05	0.97	1.62	4.40E-05	-2.58	-3.81	-18.33	IGHD3-16	immunoglobulin heavy diversity 3-16;	chr14
27	8.54	8.39	7.19	10.31	1.59	1.30	1.82	0.69	2.10E-04	-2.54	-2.30	-8.67	EYA3-IT1	EYA3 intronic transcript 1 (non-protein coding)	chr1
28	5.91	5.70	4.57	6.46	0.93	0.70	1.02	1.07	4.82E-03	-2.52	-2.19	-3.71	MIR4782	microRNA 4782	chr2
29	5.83	5.61	4.56	6.50	1.09	0.84	0.88	0.98	1.07E-02	-2.41	-2.06	-3.83	MIR376A2	microRNA 376a-2	chr14

30	7.24	7.60	6.05	7.11	0.88	0.62	0.70	0.52	2.46E-03	-2.28	-2.93	-2.08	PPP1R9A	protein phosphatase 1, regulatory subunit 9A;	chr7
31	9.56	9.40	8.37	9.92	1.10	0.87	0.72	1.19	2.33E-02	-2.28	-2.05	-2.93	MIR196A1	microRNA 196a-1	chr17
32	7.07	6.96	5.89	9.36	1.91	1.19	1.16	2.84	2.98E-04	-2.27	-2.09	-11.10	LYPD8	LY6/PLAUR domain containing 8	chr1
33	10.70	10.58	9.55	12.36	1.72	1.56	1.34	1.62	3.76E-03	-2.22	-2.04	-7.02	IGHD2-21	immunoglobulin heavy diversity 2-21;	chr14
34	8.95	8.99	7.82	9.84	0.87	0.94	1.13	0.67	1.21E-04	-2.19	-2.25	-4.05	TRAJ25	T cell receptor alpha joining 25 (non-functional)	chr14
35	6.65	7.72	5.53	8.37	1.63	1.50	1.04	1.60	2.67E-03	-2.17	-4.56	-7.16	UGT2B15	UDP glucuronosyltransferase 2 family, polypeptide B15;	chr4
36	9.58	10.13	8.47	12.06	1.91	1.50	1.86	1.25	4.70E-05	-2.16	-3.15	-12.04	TRAJ59	T cell receptor alpha joining 59 (non-functional)	chr14
37	9.66	9.66	8.59	7.54	1.39	1.43	0.98	0.94	1.30E-04	-2.10	-2.10	2.08	TGFBR2	transforming growth factor, beta receptor II (70/80kDa);	chr3
38	5.24	5.41	4.18	6.51	1.09	0.79	0.55	0.99	4.90E-05	-2.08	-2.34	-5.03	MIR514B	microRNA 514b	chrX
39	8.11	8.24	7.07	9.37	1.04	1.45	0.88	1.49	8.02E-03	-2.07	-2.25	-4.96	CDHR5	cadherin-related family member 5;	chr11
40	5.38	5.37	4.33	6.91	0.94	0.82	0.54	0.93	5.66E-07	-2.07	-2.05	-5.98	TRAJ2	T cell receptor alpha joining 2 (non-functional)	chr14
41	8.85	9.63	7.81	10.87	2.37	1.59	0.95	1.58	5.45E-03	-2.06	-3.51	-8.33	IGHD3-10	immunoglobulin heavy diversity 3-10;	chr14
42	8.09	8.72	7.07	10.09	1.67	1.93	1.99	1.41	9.71E-04	-2.03	-3.14	-8.12	IGHV3-64	immunoglobulin heavy variable 3-64;	chr14

HB, high-BCNA tumors; LB, low-BCNA tumors; MSI, microsatellite-instability tumors; N, normal tissue; RMA, bi-weight average signal (Robust Multi-array Average); SD, standard deviation; FDR, false discovery rate; FC, fold change, Chr, chromosome.

Supplementary Table 15: *Top genes (up to 50 entries) which have significant differential expression in both HB and LB tumors compared to normal tissue (FC <-2 or FC >2) and are downregulated (FC <-2) in both groups compared to MSI CRCs. Genes listed in ascending order on the bases of HB FC decrease in gene expression compared to MSI.*

#	HB RMA	LB RMA	MSI RMA	N RMA	HB SD	LB SD	MSI SD	N SD	FDR p-value (All Conditions)	FC HB vs MSI	FC HB vs N	FC LB vs MSI	FC LB vs N	Gene Symbol	Description	Chr
1	7.79	8.70	11.94	6.65	1.32	1.85	2.41	0.79	5.00E-06	-17.81	2.20	-9.50	4.13	AGR2	anterior gradient 2 homolog (Xenopus laevis);	chr7
2	11.46	11.41	14.56	8.94	2.09	2.09	3.35	1.53	5.10E-05	-8.63	5.70	-8.88	5.54	FAT1	FAT atypical cadherin 1; FAT tumor suppressor homolog 1 (Drosophila);	chr4
3	11.00	10.66	14.11	7.99	2.01	1.44	3.68	1.29	7.00E-06	-8.63	8.06	-10.95	6.36	LMAN1	lectin, mannose-binding, 1;	chr18
4	14.09	14.59	16.74	11.04	2.47	2.18	3.30	1.75	1.73E-04	-6.27	8.27	-4.44	11.68	ANXA2P2	annexin A2 pseudogene 2	chr9
5	10.93	11.57	13.40	9.30	1.82	2.17	2.20	1.75	1.43E-03	-5.52	3.09	-3.55	4.81	MIR614; GPRC5A	microRNA 614; G protein-coupled receptor, family C, group 5, member A	chr12
6	8.98	9.62	11.41	7.52	1.53	1.70	2.35	1.13	5.54E-04	-5.41	2.74	-3.45	4.29	GALNT1	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 1 (GalNAc-T1);	chr18
7	9.51	9.13	11.81	8.11	1.62	1.38	1.85	0.87	6.79E-04	-4.93	2.64	-6.42	2.02	UGT8	UDP glycosyltransferase 8;	chr4
8	11.86	11.90	14.14	10.06	1.70	1.66	2.77	1.44	5.71E-04	-4.83	3.50	-4.72	3.59	NARS	asparaginyl-tRNA synthetase	chr18
9	10.21	10.33	12.37	8.15	1.71	1.39	2.06	0.83	1.30E-05	-4.46	4.18	-4.10	4.55	ADAM9	ADAM metallopeptidase domain 9;	chr8
10	10.56	11.10	12.69	8.68	1.47	1.55	1.92	0.78	1.00E-05	-4.37	3.69	-3.01	5.35	AHR	aryl hydrocarbon receptor;	chr7
11	11.51	11.42	13.58	8.53	1.91	2.23	2.92	1.58	3.90E-05	-4.21	7.90	-4.46	7.46	HSPA1B; HSPA1A	heat shock 70kDa protein 1B; heat shock 70kDa protein 1A;	chr6
12	6.05	6.43	8.09	8.72	1.11	0.96	2.27	1.95	3.00E-06	-4.11	-6.37	-3.15	-4.88	CD177P1	CD177 molecule pseudogene 1	chr19
13	9.78	10.54	11.80	7.68	2.14	2.02	2.26	0.98	2.43E-04	-4.07	4.28	-2.40	7.26	IFI30	interferon, gamma-inducible protein 30	chr19

14	10.01	9.69	12.00	7.67	1.29	0.99	2.47	0.86	1.00E-06	-3.96	5.08	-4.96	4.06	SUPT4H1	suppressor of Ty 4 homolog 1 (S. cerevisiae)	chr17
15	9.30	9.92	11.27	8.18	1.28	1.46	2.28	0.59	1.02E-04	-3.90	2.19	-2.54	3.36	P4HA1	prolyl 4-hydroxylase, alpha polypeptide I;	chr10
16	7.95	7.91	9.87	6.40	1.12	0.78	1.81	0.55	6.80E-07	-3.79	2.92	-3.89	2.85	SLC39A6	solute carrier family 39 (zinc transporter), member 6	chr18
17	8.92	8.89	10.73	7.44	1.09	1.44	1.49	0.51	4.00E-06	-3.50	2.79	-3.56	2.74	CXCL16	chemokine (C-X-C motif) ligand 16;	chr17
18	8.45	8.87	10.24	7.14	1.03	0.84	1.66	0.56	4.00E-06	-3.47	2.48	-2.58	3.33	ARID5B	AT rich interactive domain 5B (MRF1-like);	chr10
19	9.89	8.95	11.68	6.61	1.47	1.77	2.33	0.54	2.73E-08	-3.46	9.73	-6.66	5.05	ANLN	anillin, actin binding protein;	chr7
20	8.94	9.66	10.71	6.30	3.31	2.97	4.28	1.00	6.48E-04	-3.43	6.22	-2.07	10.28	IFI6	interferon, alpha-inducible protein 6;	chr1
21	5.70	5.88	7.48	7.74	0.80	0.74	2.06	1.64	4.00E-06	-3.43	-4.13	-3.02	-3.63	CD177	CD177 molecule	chr19
22	11.23	11.57	13.00	8.84	1.68	1.53	2.48	1.47	7.40E-05	-3.40	5.26	-2.70	6.63	TMED10	transmembrane emp24-like trafficking protein 10 (yeast)	chr14
23	8.77	8.33	10.51	7.04	1.41	1.17	1.63	0.46	8.00E-06	-3.33	3.32	-4.53	2.44	NUSAP1	nucleolar and spindle associated protein 1	chr15
24	12.29	12.54	14.02	9.79	1.98	1.34	3.02	1.61	3.66E-04	-3.32	5.66	-2.80	6.72	ANXA2	annexin A2;	chr15
25	7.69	7.87	9.42	6.67	0.82	0.55	1.11	0.29	8.77E-08	-3.31	2.03	-2.92	2.30	FUT8	fucosyltransferase 8 (alpha (1,6) fucosyltransferase);	chr14
26	8.07	8.11	9.75	6.34	1.52	1.47	2.05	0.77	4.00E-05	-3.20	3.32	-3.13	3.40	ME1	malic enzyme 1, NADP(+)-dependent, cytosolic;	chr6
27	10.90	10.96	12.55	8.73	1.49	2.01	2.48	1.29	1.38E-04	-3.14	4.47	-3.01	4.67	HSPA1A; HSPA1B	heat shock 70kDa protein 1A; heat shock 70kDa protein 1B;	chr6
28	12.67	12.39	14.31	8.80	1.98	1.86	3.20	1.11	1.00E-06	-3.13	14.57	-3.79	12.04	MLEC	malectin	chr12
29	11.01	11.23	12.65	7.70	2.14	1.47	2.42	0.90	3.00E-06	-3.12	9.88	-2.68	11.50	RPL36	ribosomal protein L36	chr19
30	8.81	8.91	10.44	7.71	1.26	0.89	1.87	0.64	3.25E-04	-3.10	2.13	-2.89	2.29	JAG1	jagged 1;	chr20
31	10.42	10.37	12.05	7.75	1.51	1.33	2.13	0.83	5.02E-07	-3.08	6.39	-3.19	6.17	PKM	pyruvate kinase, muscle	chr15
32	10.41	10.56	12.03	8.67	1.44	1.10	1.75	0.78	2.20E-05	-3.07	3.33	-2.77	3.70	TSPAN13	tetraspanin 13;	chr7
33	11.83	11.03	13.44	7.37	2.44	1.77	2.98	1.17	2.00E-06	-3.05	21.94	-5.31	12.59	RPS5	ribosomal protein S5	chr19
34	13.69	14.01	15.30	9.84	2.41	1.57	3.38	1.59	8.00E-06	-3.04	14.47	-2.43	18.10	PDIA3	protein disulfide isomerase family A, member 3;	chr15
35	9.23	8.96	10.83	7.37	1.42	1.10	2.06	0.56	1.20E-05	-3.02	3.63	-3.66	3.00	TRIM2	tripartite motif containing 2;	chr4
36	7.74	7.47	9.34	6.40	0.98	0.95	1.30	0.23	2.00E-06	-3.02	2.53	-3.66	2.10	KIAA1244	KIAA1244;	chr6
37	10.18	10.44	11.76	8.26	1.43	0.94	1.78	0.69	8.00E-06	-2.99	3.79	-2.50	4.53	NSA2	NSA2 ribosome biogenesis homolog (S. cerevisiae);	chr5
38	6.78	6.84	8.35	5.55	1.11	1.16	1.60	0.64	3.80E-05	-2.98	2.35	-2.84	2.46	ANXA2P1	annexin A2 pseudogene 1	chr4
39	8.55	8.40	10.12	7.18	0.86	0.75	1.53	0.45	9.09E-07	-2.98	2.58	-3.29	2.33	CHORDC1	cysteine and histidine-rich domain (CHORD) containing 1	chr11
40	6.43	6.13	7.99	9.36	1.07	0.95	1.24	2.19	2.00E-06	-2.96	-7.61	-3.65	-9.39	SLC4A4	solute carrier family 4, sodium bicarbonate cotransporter, member 4;	chr4
41	11.71	11.64	13.27	10.03	1.55	1.28	2.69	1.70	1.72E-03	-2.96	3.19	-3.11	3.04	KLF5	Kruppel-like factor 5 (intestinal);	chr13
42	9.70	8.67	11.26	7.19	1.95	1.51	1.93	0.52	2.00E-06	-2.95	5.70	-6.03	2.79	PSAT1	phosphoserine aminotransferase 1;	chr9
43	9.65	10.05	11.20	8.52	1.34	0.87	1.58	0.72	1.27E-04	-2.92	2.20	-2.22	2.89	GSR	glutathione reductase;	chr8
44	8.58	9.04	10.11	6.70	1.86	2.11	2.32	0.89	2.05E-04	-2.88	3.69	-2.11	5.05	SLPI	secretory leukocyte peptidase inhibitor;	chr20
45	9.70	9.53	11.22	7.91	1.50	0.95	1.98	0.92	1.47E-04	-2.87	3.46	-3.24	3.06	MBNL2	muscleblind-like splicing regulator 2;	chr13
46	10.01	10.49	11.52	7.64	1.89	1.60	2.56	1.03	1.97E-04	-2.86	5.16	-2.05	7.19	DNAJC15	DnaJ (Hsp40) homolog, subfamily C, member 15;	chr13
47	7.50	7.26	9.01	6.21	1.05	0.83	1.36	0.29	2.00E-06	-2.84	2.45	-3.35	2.08	GPR180	G protein-coupled receptor 180;	chr13
48	11.35	11.29	12.85	9.35	1.60	1.21	2.14	1.11	5.40E-05	-2.83	4.00	-2.95	3.83	SOD1	superoxide dismutase 1, soluble;	chr21

49	15.00	14.14	16.50	9.66	2.96	1.99	4.06	2.00	1.50E-05	-2.82	40.63	-5.12	22.39	RPS16	ribosomal protein S16	chr19
50	6.33	6.41	7.82	9.48	1.23	1.47	1.40	1.87	6.00E-06	-2.81	-8.84	-2.67	-8.40	CA2	carbonic anhydrase II;	chr8

HB, high-BCNA tumors; LB, low-BCNA tumors; MSI, microsatellite-instability tumors; N, normal tissue; RMA, bi-weight average signal (Robust Multi-array Average); SD, standard deviation; FDR, false discovery rate; FC, fold change, Chr, chromosome.

Supplementary Table 16: *Top genes (up to 50 entries) which have significant differential expression in both HB and MSI tumors compared to normal tissue (FC <-2 or FC >2) and are downregulated (FC <-2) in both groups compared to LB CRCs. Genes listed in ascending order on the bases of HB FC decrease in gene expression compared to LB.*

#	HB RMA	LB RMA	MSI RMA	N RMA	HB SD	LB SD	MSI SD	N SD	FDR p-value (All Conditions)	FC HB vs LB	FC HB vs N	FC MSI vs LB	FC MSI vs N	Gene Symbol	Description	Chr
1	8.58	12.72	8.12	10.46	2.31	3.42	3.90	2.72	2.48E-03	-17.58	-3.67	-24.20	-5.05	MUC2	mucin 2, oligomeric mucus/gel-forming;	chr11
2	16.14	18.63	16.50	18.49	2.20	2.00	2.35	1.44	5.21E-03	-5.60	-5.07	-4.38	-3.97	MIR548I4	microRNA 548i-4	chr7
3	6.57	8.57	4.99	9.16	2.73	2.77	1.41	3.24	1.19E-02	-3.98	-6.03	-11.93	-18.04	UGT2B17	UDP glucuronosyltransferase 2 family, polypeptide B17;	chr4
4	6.82	8.80	7.19	9.79	2.56	2.23	2.09	2.57	3.30E-02	-3.94	-7.81	-3.07	-6.08	FCGBP	Fc fragment of IgG binding protein	chr19
5	9.60	11.42	7.49	12.45	2.06	2.16	1.67	1.98	5.80E-04	-3.54	-7.20	-15.33	-31.14	FABP1	fatty acid binding protein 1, liver;	chr2
6	6.96	8.28	7.18	9.64	1.95	2.46	1.26	2.35	3.27E-03	-2.50	-6.42	-2.13	-5.49	NXPE1	neurexophilin and PC-esterase domain family, member 1	chr11
7	6.55	7.86	6.02	7.66	1.07	1.41	1.51	1.21	4.35E-03	-2.49	-2.15	-3.59	-3.11	MUC4	mucin 4, cell surface associated;	chr3
8	7.20	8.40	7.04	10.98	2.97	4.32	1.43	3.25	1.80E-02	-2.30	-13.74	-2.58	-15.40	ANPEP	alanyl (membrane) aminopeptidase;	chr15
9	5.88	7.02	5.51	6.92	1.95	3.15	1.15	1.18	4.40E-02	-2.20	-2.06	-2.85	-2.66	SI	sucrase-isomaltase (alpha-glucosidase);	chr3
10	6.65	7.72	5.53	8.37	1.63	1.50	1.04	1.60	2.67E-03	-2.10	-3.29	-4.56	-7.16	UGT2B15	UDP glucuronosyltransferase 2 family, polypeptide B15;	chr4

HB, high-BCNA tumors; LB, low-BCNA tumors; MSI, microsatellite-instability tumors; N, normal tissue; RMA, bi-weight average signal (Robust Multi-array Average); SD, standard deviation; FDR, false discovery rate; FC, fold change, Chr, chromosome.

Supplementary Table 17: *Top genes (up to 50 entries) which have significant differential expression in both LB and MSI tumors compared to normal tissue (FC <-2 or FC >2) and are downregulated (FC <-2) in both groups compared to HB CRCs. Genes listed in ascending order on the bases of LB FC decrease in gene expression compared to HB.*

#	HB RMA	LB RMA	MSI RMA	N RMA	HB SD	LB SD	MSI SD	N SD	FDR p-value (All Conditions)	FC LB vs HB	FC LB vs N	FC MSI vs HB	FC MSI vs N	Gene Symbol	Description	Chr
1	12.85	9.96	11.37	7.21	3.07	2.44	2.65	1.13	7.10E-07	-7.43	6.73	-2.80	17.90	SCD	stearoyl-CoA desaturase (delta-9-desaturase);	12.85
2	11.20	9.38	5.97	8.03	2.20	1.69	2.75	1.05	8.00E-06	-3.54	2.55	-37.57	-4.17	CFTR	cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7)	11.20
3	9.56	7.27	5.73	18.35	3.59	3.19	2.52	4.26	3.00E-06	-4.89	-2169.20	-14.28	-6330.50	SLC26A3	solute carrier family 26, member 3;	9.56
4	12.81	11.12	11.45	7.60	2.28	1.98	2.31	1.35	1.00E-06	-3.22	11.47	-2.58	14.35	IPO7	importin 7	12.81
5	12.82	11.17	11.37	7.56	2.43	2.34	1.91	1.86	2.00E-06	-3.14	12.19	-2.72	14.07	LINC00657	long intergenic non-protein coding RNA 657	12.82
6	9.70	8.06	8.22	6.97	1.63	1.94	1.30	0.49	3.71E-07	-3.13	2.13	-2.80	2.37	AXIN2	axin 2	9.70
7	13.51	12.08	12.28	9.61	2.34	1.64	2.69	1.77	4.94E-04	-2.69	5.56	-2.35	6.37	TM9SF3	transmembrane 9	13.51

															superfamily member 3	
8	12.34	10.92	10.01	8.11	2.07	1.83	1.85	0.92	2.77E-07	-2.67	7.04	-5.04	3.73	HSPH1	heat shock 105kDa/110kDa protein 1	12.34
9	11.69	10.46	10.01	7.29	1.79	1.54	1.32	0.58	1.09E-09	-2.35	9.02	-3.20	6.61	RNF43	ring finger protein 43	11.69
10	12.29	11.19	11.18	9.21	2.05	1.66	2.54	1.68	4.96E-04	-2.15	3.95	-2.16	3.93	TFRC	transferrin receptor (p90, CD71)	12.29

HB, high-BCNA tumors; LB, low-BCNA tumors; MSI, microsatellite-instability tumors; N, normal tissue; RMA, bi-weight average signal (Robust Multi-array Average); SD, standard deviation; FDR, false discovery rate; FC, fold change, Chr, chromosome.