Characterisation of genomic alterations in proximal and distal colorectal cancer patients

Ryia-Illani Mohd Yunos [Corresp., 1], **Nurul-Syakima Ab Mutalib** [1], **Sheau Sean Khor** [2], **Sazuita Saidin** [1], **Norshahidah Mahamad Nadzir** [1], **Zuraini Abd. Razak** [1], **Isa Md. Rose** [3], **Ismail Sagap** [4], **Luqman Mazlan** [4], **Norfilza Mohd Mokhtar** [5], **Rahman Jamal** [Corresp. 1]

[1] UKM Medical Molecular Biology Institute, Cheras, Kuala Lumpur, Malaysia

[2] Life Sciences Solutions, Thermo Fisher Scientific Incorporation, Singapore, Singapore

[3] Department of Clinical Oral Biology, Faculty of Dentistry, Universiti Kebangsaan Malaysia, Kuala Lumpur, Kuala Lumpur, Malaysia

[4] Department of Surgery, Faculty of Medicine, Universiti Kebangsaan Malaysia, Cheras, Kuala Lumpur, Malaysia

[5] Department of Physiology, Faculty of Medicine, Universiti Kebangsaan Malaysia, Cheras, Kuala Lumpur, Malaysia

Corresponding Authors: Ryia-Illani Mohd Yunos, Rahman Jamal
Email address: ryia.yunos@ppukm.ukm.edu.my, rahmanj@ppukm.ukm.edu.my

Background: Colorectal cancer (CRC) is one of the commonest cancers in Malaysia and majority of the patients will present with advanced disease at diagnosis particularly in cases of proximal CRCs. Little is known about the relationship between the genetic landscape and the anatomical location of the tumour as well as the prognostication in CRC patients. **Objectives:** The objectives of this study were to determine the somatic single nucleotide variants (SNV) and the cellular pathways between the proximal and distal CRCs. **Methods:** Whole exome sequencing was performed using the Ion Proton platform on 10 pairs of normal and CRC samples. The sequencing results were analysed using the Torrent Suite Software and the variants were annotated using ANNOVAR followed by validation with Sanger sequencing. **Results:** The commonly altered genes in CRCs are KRAS, APC, TP53 and ATM. APC is the most frequently altered gene in both proximal and distal CRCs. KRAS and ATM genes were exclusively altered in the proximal CRCs with a frequency of 60% and 40%, respectively. On the other hand, *TP53* mutations did not show any CRC anatomical predominance. There were five recurrent novel variants in proximal CRCs and one recurrent novel variant in distal CRC. Wnt signalling pathway was the most frequently altered pathway in both proximal and distal CRCs whereas TGF-β and PI3K signalling pathways were predominantly altered in the proximal CRCs. **Conclusion:** We found that proximal CRCs presented with more variants and altered pathways as compared to distal CRCs. We also discovered that the TGF-Beta signalling (four mutations) and PI3K signalling (two mutations) pathways were exclusively altered in proximal CRCs. However, further study in larger series of samples coupled with functional studies will be needed to confirm the identified variants and determine their role in the genesis of proximal and distal CRCs.

Characterisation of genomic alterations in proximal and distal colorectal cancer patients

Ryia Illani Mohd Yunos[1], Nurul-Syakima Ab Mutalib[1], Janice Khor Sheau Sean[2], Sazuita Saidin[1], Norshahidah Mahamad Nadzir[1], Zuraini Abd Razak[1], Isa Md. Rose[3], Ismail Sagap[4], Luqman Mazlan[4], Norfilza Mohd Mokhtar[5], Rahman Jamal[1]

[1] UKM Medical Molecular Biology Institute (UMBI), [2]Life Sciences Solutions, Thermo Fisher Scientific Inc., [3]Department of Clinical Oral Biology, Faculty of Dentistry, [4]Department of Surgery, [5]Department of Physiology, Faculty of Medicine, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia.

Abstract

Background: Colorectal cancer (CRC) is one of the common cancers in Malaysia. Majority of the patients will present with advanced disease at diagnosis particularly in cases of proximal CRCs. Little are known about the relationship between the genetic landscape and the anatomical location of the tumour; as well as the prognostication in CRC patients.

Objectives: The objectives of this study were to determine the somatic single nucleotide variants (SNV) and the cellular pathways between the proximal and distal CRCs.

Methods: Whole exome sequencing was performed on the Ion Proton platform on 10 pairs of normal and CRC samples. The sequencing results were analysed using the Torrent Suite Software and the variants were annotated using ANNOVAR; followed by validation with Sanger sequencing.

Results: The common altered genes in CRCs are KRAS, APC, TP53 and ATM. APC is the most frequently altered gene in both proximal and distal CRCs. KRAS and ATM genes were particularly altered in the proximal CRCs with a frequency of 60% and 40%, respectively. On the other hand, *TP53* mutations did not show any CRC anatomical predominance. There were five recurrent novel variants in

proximal CRCs and one recurrent novel variant in distal CRC. Wnt signalling pathway was the most frequently altered pathway in both proximal and distal CRCs whereas TGF-β and PI3K signalling pathways were predominantly altered in the proximal CRCs.

Conclusion: We found that proximal CRCs presented with more variants and different altered pathways as compared to distal CRCs. We also discovered that the TGF-Beta signalling (four mutations) and PI3K signalling (two mutations) pathways were exclusively altered in proximal CRCs. However, further study in larger series of samples coupled with functional studies will be required to confirm the identified variants and determine their roles in the genesis of proximal and distal CRCs.

Introduction

Colorectal cancer (CRC) is the third most common diagnosed cancer worldwide (Ferlay *et. al*., 2015) and is ranked as the second most common cancer in Malaysia (Zainal Arrifin and Nor Salehah, 2011). According to the National Cancer Registry Report, CRC is the most common cancer among men and the third most common among women in Malaysia (Lim *et al*., 2008). In year 2012 there were 1,360,602 new cases and 693,933 deaths from CRC worldwide (Ferlay *et. al*., 2013).

Anatomically, CRCs are classified into three subsets named, proximal, distal and rectum. Proximal CRC is located on the right side of the colon, which includes cecum, ascending colon, hepatic flexure, transverse colon and splenic flexure; while distal CRC is located on the left side of the colon, including sigmoid and descending colon (Gonzalez *et al*., 2001). It is postulated that both the proximal and distal CRCs are anatomically differed. Thus, these arise different biological pathways, suggesting different molecular mechanisms involved (Bufill, 1990). Biological differences between the normal proximal and distal colon suggest that the carcinogenesis in these locations may be mediated via different molecular pathways (Minoo *et. al*., 2010; Missiaglia *et al*., 2014). This may have profound prognostic and therapeutic implications. Differences in gene expression between adenocarcinoma of the proximal and

distal CRC do exist and these information are ought to be taken into consideration when investigating new predictive and prognostic biomarkers (Maus *et al*., 2015). Among the molecular characteristics that distinguishing between proximal and distal CRCs are the low frequency of TP53 and KRAS gene mutations, lower c-MYC expression and DNA mismatch repair (MMR) deficiency in proximal CRC (Bogaert and Prenen, 2014). Distal CRC, on the other hand, shows higher frequency of TP53 and KRAS gene mutations and c-MYC expression (Bogaert and Prenen, 2014).

Several large cohort studies in the last 20 years have demonstrated that the proximal and distal CRCs differed in their susceptibility to screening tests, the stage at which they were diagnosed, patient characteristics, pathology and prognosis (Benedix *et al*., 2010; Myer *et al*., 2012). A few studies find that the distribution of distal and proximal CRCs varies according to ethnicity (Li and Lai., 2009). Individuals with African ancestry are more likely to develop proximal cancer rather than distal CRCs (Wu *et al*., 2004; Irby, *et al*., 2006); whereas Asians and Pacific Islanders are more prone to distal colon and rectal cancers (Wu *et al*., 2004). A study by Goh and colleagues supported this fact, whereby 77% of CRC cases in Malaysia were diagnosed as distal CRCs (Goh *et al*., 2005). However, despite the different distributions, the actual causative molecular events that lead to the different prognosis between proximal and distal CRCs remains poorly understood.

Next generation sequencing technologies have revolutionized cancer research and management, particularly in diagnostic and treatment strategies (Meldrum *et. al*., 2011; Roychowdury *et. al*., 2011). Using the whole exome sequencing approach, we examined the exomes of Malaysian patients diagnosed with proximal and distal CRCs in order to identify the differences and the distinct signatures based on their anatomical distribution. We hope that this can contribute to a better management and treatment of CRC patients in the future.

Methods and Materials

Clinical material

A total of ten paired colorectal carcinoma and their corresponding adjacent normal tissues (five proximal and five distal), were collected from patients at Universiti Kebangsaan Malaysia Medical Centre, Kuala Lumpur. The written informed consents were provided by patients to participate in the study. This study was approved by the UKM Research Ethics Committee (Reference no: UKM 1.5.3.5/244/UMBI-004-2012). The tissues were subjected to cryosectioning; followed by haematoxylin and eosin (H&E) staining. Our pathologist confirmed the presence of at least 80% cancer cells in the tumour specimens and less than 20% necrosis in paired unaffected colorectal tissue adjacent to the tumour site. DNA extraction was performed using the QIAamp DNA mini kit (QIAGEN, Valencia, CA) according to the manufacturer's protocol. Quality and quantity of the extracted DNA were assessed using Qubit Flourometer (Invitrogen, Carlsbad, CA, USA), NanoDrop 2000 spectrophotometer (NanoDrop Technologies, Wilmington, DE); as well as agarose gel electrophoresis. To confirm the identity of each tumour and normal paired samples, genotyping was performed by multiplex PCR based on microsatellite polymorphisms using Coriell Identity Mapping Kit (Coriell Institute, New Jersey, USA).

Exome capture, library construction and next generation sequencing

One microgram (1 μg) of genomic DNA from each pair of tumour and normal sample was mechanically sheared by ultrasonic fragmentation using the Covaris® System to achieve fragments of about 50 – 500bp. The fragmentation profile was assessed by the Bioanalyzer High Sensitivity DNA Analysis Kit (Agilent Technologies, Carlsbad, CA, USA). The fragmented DNA was used to construct a

fragment library using the Ion Plus Fragment Library Kit (Life Technologies, Guilford, CT) according to the manufacturer's instructions for ligation, end repair, purification, size selection and final amplification prior to exome capture. For multiplexing the samples, adapters with short stretches of index sequences from Ion Xpress™ Barcode Adapters 1 – 16 kit was used and thus allowing the sequencing of two samples in a single Ion PI™ chip run. Five hundred nanograms (500 ng) of the amplified, size selected library (~285 bp) from each sample was subjected to exome capture procedure using the Ion TargetSeq™ Exome kit (Life Technologies, Guilford, CT) according to manufacturer's protocol. Exomes were captured using ~2 million TargetSeq™ capture probes with biotinylated oligos that range from ~50 bases to ~120 bases. The captured DNA fragments were isolated using streptavidin-coated Dynabeads paramagnetic beads, and then were amplified and purified. Finally, the samples were quantitated and qualitatively assessed on the Bioanalyzer High Sensitivity DNA chip (Agilent Technologies, Carlsbad, CA, USA). The purified, 10 pM of exome-enriched libraries were used for template preparation on Ion PI™ Ion Sphere™ Particles (ISPs) for sequencing on an Ion PI™ Chip using Ion Proton sequencer (Life Technologies, Guilford, CT).

Bioinformatics analyses

The data was processed using the Torrent Suite software v4.2.1. The Torrent Suite Software automated the generation of sequence reads, trimming of adapter sequences, removal of poor quality reads; as well as sequence alignment to the hg19 human genome reference. Variants were called using the Torrent Variant Caller plugin, configured for somatic mutation detection with low stringency setting. Variant filtering and calculation of transition to transversion ratio (Ti/Tv) was performed using SnpSift tool in SnpEff (Cingolani *et. al*., 2012).The variants were then subjected to annotationusing ANNOVAR (Wang and Hakonarson, 2010). Gene based annotation was performed against RefSeq Gene, UCSC Known Gene and ENSEMBL Gene. The variants were further annotated against the conserved region

(phastConsElements46way), segmental duplication region (genomicSuperDups), alternative allele frequency in all subjects in the National Heart Lung and Blood Institute Exome Sequencing Project (NHLBI-ESP) project with 6500 exomes (esp6500si_all), alternative allele frequency data in 1000 Genomes Project (1000g2014oct_all), the Exome Aggregation Consortium (ExAC 01), dbSNP version 138 (snp138), CLINVAR (clinvar_20140929) and COSMIC version 70 (cosmic70). Protein impact prediction was also performed on ANNOVAR using SIFT, PolyPhen2 HDIV, PolyPhen2 HVAR, LRT, MutationTaster, MutationAssessor, FATHMM, GERP++, PhyloP and SiPhy (using command ljb26_all). In order to identify potential druggable variants, the drug gene interaction database, DGIdb 2.0 was utilized to annotate the variants against drugs genes interaction dataset (http://dgidb.genome.wustl.edu/).

Variants prioritization

Variants exclusion criterias included those with base quality less than Q30, not resulting in amino acid changes, identified in unannotated genes (unknown), called in both tumour and normal exomes, representing probable mapping ambiguities, and have minimal allele frequency (MAF) of more than 5% in the 1000 Genomes Project, ExAC and ESP6500 database. A variant in a tumour was considered to be a novel true candidate somatic mutation if the corresponding normal sample has at least 10 reads covering this position, zero variant reads and has not been reported in dbSNP138 or the 1000 Genomes data set (October 2014). For the resulting candidate of somatic mutations, the alignment of each sample was manually examined to check for sequencing artefacts and alignment errors using the Integrated Genomic Viewer (IGV) (Robinson *et. al*., 2011; Thorvaldsdóttir *et. al*., 2013).

Gene pathway analysis

The Ingenuity Pathway Analysis (IPA) (Qiagen, Valencia, CA) was used for core analysis to identify the affected canonical pathways in proximal and distal CRCs.

Variant confirmation

Variants were validated using the Sanger sequencing method. Primers corresponding to the selected locations were designed using IDT-DNA Primer Quest (Coralville, IA). The primer sequences were included in Supplementary Table 1. PCR products were generated and cycle sequencing was performed using the Big Dye Terminator V3.1 reagent (Life Technologies, Guilford, CT). The cycle sequencing products were then processed using ethanol precipitation and sequencing was carried out using the ABI 3130xl capillary electrophoresis (Life Technologies, Guilford, CT). The results were analysed using the Basic Local Alignment System Tool (BLAST) and Sequence Scanner (Applied Biosystem, Foster City, CA).

Results

Clinicopathological characteristic of patients

The characteristics of all ten patients were listed in Table 1. With regards to cancer stage, 20% (n=2) of the patients were of Dukes' A, 40% (n=4) were Dukes' B and the remaining 40% (n=4) were Dukes' C (Dukes 1932). The average age of patients was approximately 69 years old (range 58–75 years). The samples comprised of three well differentiated adenocarcinoma, three moderate differentiated adenocarcinoma and four poorly differentiated adenocarcinoma. From these ten samples, four patients were positives for lymph nodes metastasis and six were negative.

Exome sequencing analysis and coverage

The capture regions covered by Ion TargetSeq was about 37.3 Mb and we managed to obtain an average of 39.6 million reads. An average coverage of about 70X for each sample was obtained and the coverage of the target region at 20X was more than 85%. This was comparatively higher than what were obtained by a study on pancreatic cancer (Wang *et al*., 2012) and a study on colon and rectal cancers (The Cancer Genome Atlas 2012). On average, the number of variants detected at Q30 for each sample was 35, 713 (32, 804 - 37,487 variants) (Table 2). To assess the quality of variants, the ratio of the number of transitions to the number of transversions was determined. The expected Ti/Tv ratio for exome target regions is 2.8 (DePristo *et al*., 2011). However, the target regions of exome capture kits often covered more than just exons. For SNPs resided in these target regions, Ti/Tv ratios between 2.0 and 3.0 were observed (Guo *et al*., 2014). The target regions of Ion TargetSeq kit covered both exonic and non exonic regions, such as 3' UTR and 5' UTR, therefore, we obtained an average Ti/Tv ratio of 2.7.

Upon variants prioritizations, we obtained a total of 4,737 and 4,178 variants in proximal and distal CRC, respectively. Among all the variants found in proximal CRC, 539 were protein-altering mutations in 508 genes located in the conserved regions. On the other hand, the distal CRCs had 246 protein-altering mutations in 180 genes located in the conserved regions.

Profile of mutated CRC-related genes and altered pathways

Thirty three (33) genes that have been reported in the tumorigenesis of CRCs; compiled from a list of ten publications (Perreault *et al*., 2001; Moreno-Bueno *et al*., 2003; Segditsas *et al*., 2003; Reya *et al*., 2005; Van es *et al*., 2005; Femia *et al*., 2007; Nagel *et al*., 2008; Clevers *et al*., 2012; Diaz *et al*., 2012; Vogelstein *et al*., 2013). We plotted each altered gene which detected in our patients in figure 1A. Altered genes were defined as any gene that has at least one or more protein-altering mutations. Fifteen (15) out of the 33 CRC-related genes were altered in proximal cancers (29 mutations) and five CRC-

related genes were altered in distal cancers (eight mutations) as listed in Table 3. To explore the affected pathways and the differences between the two subsites of CRC, we focused on major pathways involved in the CRC tumorigenesis (Moreno-Bueno *et al*.,2003; Segditsas *et al*., 2003; Femia *et al*., 2007; Nagel *et al*.,2008; Clevers *et al*., 2012; Diaz *et al*., 2012; Vogelstein *et al*., 2013). This approach was adapted from Ashktorab *et al*. (2014). An affected pathway is defined when at least one or more genes are altered in any pathway (Figure 1B). We observed that 90% (n=9) of the CRC patients shared an affected Wnt signalling pathway with five genes being altered (12 mutations). RTK-RAS and TP53 signalling pathways were also found to be altered in both proximal and distal CRCs with six mutations in both pathways. We also discovered that the TGF-Beta signalling (four mutations) and PI3K signalling (two mutations) pathways were exclusively altered in proximal CRCs. Overall, there were more altered pathways in proximal as compared to distal CRCs. Analysis using Ingenuity Pathway Analysis (IPA) software identified the Wnt signalling and growth factor signalling pathways as the most commonly affected (Figure 3).

Most frequently altered genes in proximal and distal CRCs

We discovered that the APC gene (six mutations) was the most frequently mutated gene with a frequency of 60% (n=3) in both proximal and distal CRCs. The second most frequently mutated gene in both proximal and distal CRC was TP53 gene (six mutations) with the frequency of 40% (n=2). This suggested that *APC* and *TP53* did not have predominance for either side of the colon. Interestingly, we found that the KRAS gene (three mutations) and ATM gene (two mutations) were uniquely altered in proximal CRCs with the frequency of 60% and 40% respectively (Figure 2). Among all of the frequent altered genes in this set of analysis, we identified three novel mutations in APC gene, namely, a nonframeshift substitution (APC g.112176559TT>GC) in tumour R5T, stopgain (APC g.112162896T>G) in tumour L1T, stopgain (APC g. 112175322C>G) in tumour L4T and two novel mutations in ATM gene of tumour R3T (ATM g.108141828A>G) and R5T (ATM g.108106477G>T)(Table 4). Sanger sequencing successfully confirmed each of the 17 somatic mutations we had detected in these four genes.

214

215 **Recurrent variants and mutated genes**

216 In this discovery set of ten Malaysian CRC patients, five genes with one mutation were
217 demonstrated in at least two individuals of proximal CRC patients. On the other hand, only one gene,
218 with one mutation, was identified to be recurrent in at least two individuals of distal CRCs. Four novel
219 variants in C9orf50, ZNF337, ZNF783 and CFAP74 genes were found; which have not been reported in
220 dbSNP and COSMIC with minimum allele frequency less than 5% in three databases (1000 genome,
221 ExAC and ESP6500) (Table 5).

222 Detailed analysis revealed that the minimum reads covering the mutated allele was 35 reads with
223 a minimum of 1% of the total reads containing the alternate base. This was found in the variant presented
224 in C9orf50 gene. The variant in the KRAS gene (KRAS g.25398284C>T) has the highest variant
225 coverage with 121 reads and 52 reads (63%) contained the variant in sample R2T. The same variant was
226 detected in sample R5T with 96 reads and 48 of the reads (50%) contained the alternate base.

227

228 **Actionable alterations in proximal and distal CRC**

229 We looked at the drug-gene interaction by annotating against the Drug Gene Interaction Database
230 (DGIdb) and identified ten genes implicated in CRC tumorigenesis that are clinically relevant, including
231 targets of new and existing therapies and genes. Twenty one (21) variants in ten genes and eight variants
232 in three genes were identified in proximal and distal CRCs, respectively. Notably, 80% (8/10) of CRC
233 patients harboured at least one actionable alteration (range one to seven alterations) that has been linked
234 to a clinical treatment option or is currently being investigated in clinical trials for novel targeted
235 therapies. For example, in the present study, 5FU-based chemotherapy is considered to target patients
236 with wild type *TP53*, which includes patients R2, R3, R5, L1, L2 and L4 (Figure 1A).

Discussions

CRC is a heterogeneous disease with the genetic landscape and clinical outcomes dependent on the anatomic location of the cancer. Many efforts have been made to unveil the genetic alterations and molecular features of colorectal cancer (The Cancer Genome Atlas 2012; Ashktorab *et al*., 2014). Studies shows increased evidences on these alterations would determine the prognosis and response to treatment (Nikhil *et al*., 2012; Fang *et al*., 2014; Maus et., al 2015). Nevertheless, how the anatomic subsite has an impact on the molecular features and prognosis remained not well-studied.

In this study we analysed somatic alterations between proximal and distal CRC in the Malaysian patients. It has provided an insight into the identification of known and novel somatic mutations that suggest relationship between the genomic alterations, cellular pathways, actionable genes and anatomical location of the tumour. Overall, we discovered that proximal CRCs exhibited more somatic mutations and altered pathways as opposed to distal CRCs. This suggested that these may account for the poor prognosis of proximal CRC patients.

The Wnt signalling and EGFR pathways are among the leading pathways in colorectal cancer tumorigenesis (Kandoth *et al*., 2013; Vogelstein *et al*., 2013). To identify possible pathway differences in proximal and distal, we performed pathway analysis using IPA. We discovered that the most enriched pathways are the Wnt and the growth factor signalling. Ninety percent (90%) of our patients in this discovery set, irrespective of their anatomical location, had a mutation in one or more members of the Wnt signalling pathway, predominantly in *APC*. The frequency of mutated APC gene across proximal and distal CRC patients in our study is 60% (Figure 2). This is in concordance with recent findings where they discovered over 50% of the recruited patients in the study exhibited altered Wnt signalling pathway with APC being the most significantly mutated gene (Pamplona *et al*., 2015; The Cancer Genome Atlas 2012).

Wnt signalling pathway activation is required for maintenance of colorectal tumours harbouring *APC* mutations (Scholer-Dahirel *et al*., 2011). Inactivating *APC* mutations occur in about 85% of CRCs, resulting in β-catenin stabilization and increased signalling through the Tcf/Lef transcription factors. Mutant β-catenin is free to enter the nucleus and constitutively activates transcription through Tcfs (Gregorieff and Clevers 2005). β-catenin inhibition *in vivo* strongly inhibited the growth of established *APC*-mutant colorectal tumour xenografts (Scholer-Dahirel *et al*., 2011). In our small set of patients, those with wild type APC gene (R3T and L5T) harboured at least one mutation in CTNNB1 gene (Figure 1A), resulting in altered Wnt signalling pathway. It has been shown that, up to 50% of CRC with wild type APC gene were found to have *CTNNB1* mutations (Sparks *et al*. 1998).

From our analyses, we also found that the TP53 signalling pathway was altered in both proximal and distal CRCs. With a frequency of 40%, *TP53* is the second commonest altered gene in our study. TP53 gene mutations in an Iranian cohort of CRC patients occurred as frequent as in other studies (Mahdavinia *et al*., 2009) with equal distribution across different anatomic location. However, different mutation spectra of *TP53* was observed depending on proximal or distally located tumour (Malekzadeh *et al*., 2009). Alteration of *TP53* may have different prognostic significance depending on the ethnic group (Manne *et al*., 1998), anatomic subsite of the colon (Manne *et al*., 1998; Samowitz *et al*., 2002), and stage of disease (Soong *et al*., 1997; Adrover *et al*., 1999). There is a convincing evidence that patients with wild-type *TP53* gained a survival benefit from the use of 5-fluorouracil (5FU)-based chemotherapy (Iacopetta 2003). Patients with mutant p53 do not gain this survival benefit (Daniel *et al*., 2003). We identified six patients with wild type *TP53* and these patients could potentially benefit from 5FU based chemotherapy. However, we postulated that four *TP53* mutated patients, in our study, might not gain survival benefit from 5FU treatment. In addition, the ATM gene that plays a significant role in the TP53 signalling pathway, is found mutated in 40% of the proximal CRC patients in this present study. For these patients with mutated ATM, they have the additional option of the ATM/ATR Kinase inhibitor as the alternative treatment.

The tumour progression and response towards treatment are believed to be dependent on both *ATM* and *TP53* status (Jiang *et al*., 2009; Song *et al*., 2007). For instance, ATM signalling is necessary for the survival of TP53-deficient cells after DNA damage (Reinhardt *et al*., 2007); whereas in cancers with wild type *TP53*, inactivation of ATM allows the survival of genomically unstable cells and induces chemoresistance (Jiang *et al*., 2009; Reinhardt *et al*., 2007). In TP53-deficient settings, inhibition of ATM dramatically sensitizes tumours to DNA-damaging chemotherapy, whereas, conversely, in the presence of functional TP53, inhibition of ATM actually promotes resistance effect towards chemotherapy (Jiang *et al*., 2009). Thus, the specific set of alterations induced during tumour development play an important role in determining both the tumour response towards chemotherapy and specific susceptibilities to targeted therapies in a given cancer type.

Other than the genes implicated in CRC tumorigenesis, we highlighted here six genes which were of particular interest due to recurrently mutated exclusively in both proximal and distal CRCs. Out of this six recurrently mutated genes, KRAS and GPR6 are known to be involved in cancer, as reported in COSMIC. The GPR6 gene was mutated uniquely in two patients with proximal CRC. GPR6 protein plays an important role in signal transmission and regulates many cellular functions. There are evidences implicating the GPR6 protein and its downstream signalling targets in cancer initiation and progression, where it can influence cell growth and survival through activation of AKT/mTOR and MAPK pathway (Morgan *et al*., 2014). Cancer cells may exploit this pathway resulting in promotion of tumour growth, angiogenesis and metastasis to distant sites (Gaorav and Joan 2006). Findings of recurrently altered GPR6 gene unique to the proximal CRC patients may explain the poor prognosis and low survival rate in these patients. By directly targeting GPR6 or its downstream signalling components, it may help to identify novel therapeutic opportunities for cancer prevention and treatment. However, further investigations will be warranted to examine the potential impact of this mutation.

Exome sequencing may lead to the discovery of novel targets, driver mutations as well as novel colorectal cancer predisposing mutations. This application is getting more common in clinical practice

and represents a cost-effective approach to characterize somatic mutations. This discovery study in our own local CRC patients provides a number of insights into the differences in genetic landscape of proximal and distal and identifies potential therapeutic targets in particular to the anatomic subsites of CRC, specifically in Malaysian patients. Nevertheless, further study in larger series of samples coupled with functional studies will be needed to confirm the identified variants and determine their role in the genesis of proximal and distal CRCs.

Acknowledgement

The authors thanked Ms. Chia Chiu Lim for her extensive help in the validation process of the identified variants.

References

Adrover E, Maestro ML, Sanz-Casla MT, del Barco V, Cerdan J, Fernandez C, Balibrea JL. 1999. Expression of high p53 levels in colorectal cancer: a favorable prognostic factor. *British Journal of Cancer* 81:122–126.DOI: 10.1038/sj.bjc.6690660.

Ashktorab H, Daremipouran M, Devaney P, Varma S, Rahi H, Lee E, Shokrani B, Schwartz R, Nickerson ML, Brim H. 2014. Identification of Novel Mutations by Exome Sequencing in African American Colorectal Cancer Patients. *Cancer* 121 (1):34-42. DOI: 10.1002/cncr.28922.

Benedix F, Kube R, Meyer F, Schmidt U, Gastinger I, Lippert H. 2010. Comparison of 17,641 patients with right and left-sided colon cancer: differences in epidemiology, preoperative course, histology, and survival. *Diseases of the Colon & Rectum* 53(1):57-64. DOI: 10.1007/DCR.0b013e3181c703a4.

Bogaert J & Prenen H. 2014. Molecular genetics of colorectal cancer. *Annals of Gastroenterology*. 27(1):9-14.

Bouwman P and Jonkers J. 2012. The effects of deregulated DNA damage signalling on cancer chemotherapy response and resistance. *Nature Review in Cancer* 12: 587–598. DOI: 10.1038/nrc3342

Bufill JA. Colorectal cancer: Evidence for distinct genetic categories based on proximal or distal location. 1990. *Annals of Internal Medicine* 113(10):779-88. DOI: 10.7326/0003-4819-113-10-779.

Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* 6(2):80-92. DOI: 10.4161/fly.19695.

Clevers H and Nusse R. 2012. Wnt/beta-catenin signalling and disease. *Cell* 149:1192-1205. DOI: 10.1016/j.

Daniel BL, Paul DH & Patrick GJ. 2003. 5-Fluorouracil: mechanisms of action and clinical strategies. *Nature Reviews Cancer* 3:330-338. DOI: 10.1038/nrc1074.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky M, Sivachenko AY, Cibulski K, Gabriel SB, Altshuler D, Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 43(5):491–498. DOI: 10.1038/ng.806.

Diaz LA Jr, Williams RT, Wu J, Kinde I, Hecht R, Berlin J, Allen B, Bozic I, Reiter JG, Nowak MA, Kinzler KW, Oliner KS, Vogelstein B. 2012. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* 486:537-540. DOI: 10.1038/nature11219.

Dukes CE, 1932. The classification of cancer of the rectum. *The Journal of Pathology*, 35: 323–332. DOI: 10.1002/path.1700350303.

Fang W, Radovich M, Zheng Y, Fu CY, Zhao P, Mao C, Zheng Y and Zheng S. 2014. Druggable alterations detected by Ion Torrent in metastatic colorectal cancer patients. *Oncology Letters* 7:1761-1766 DOI: 10.3892/ol.2014.2047.

Femia AP, Dolara P, Giannini A, Salvadori M, Biggeri A, Caderni G. 2007. Frequent mutation of Apc gene in rat colon tumours and mucindepleted foci, preneoplastic lesions in experimental colon carcinogenesis. *Cancer Research* 67:445-449. DOI: 10.1158/0008-5472.CAN-06-3861.

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. 2015. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. International Journal of Cancer 136: 359 – 386. DOI: 10.1002/ijc.29210.

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. 2015. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. International Journal of Cancer 136: 359 – 386. DOI: 10.1002/ijc.29210.

Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. 2013. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles using the cBioPortal. *Science Signaling.* 6(269): p11. DOI: 10.1126/scisignal.2004088.

Gaorav PG and Joan M. 2006. Cancer Metastasis: Building a Framework. *Cell Press.* 127(4): 679 – 695. DOI: 10.1016/j.cell.2006.11.001.

Guo Y, Ye F, Sheng Q, Clark T, Samuel DC. 2014. Three-stage quality control strategies for DNA re-sequencing data. *Briefings in Bioinformatics* 15(6):879–889 DOI:10.1093/bib/bbt069.

Goh KL, Quek KF, Yeo GT, Hilmi IN, Lee CK, Hasnida N, Aznan M, Kwan KL, Ong KT. 2005. Colorectal cancer in Asians: a demographic and anatomic survey in Malaysian patients undergoing colonoscopy. *Alimentary Pharmacology & Therapeautics* 22(9):859-64. DOI: 10.1111/j.1365-2036.2005.02646.x.

Gonzalez EC, Roetzheim RG, Ferrante JM, Campbell R. 2011. Predictors of proximal vs. distal colorectal cancers. *Diseases of the Colon & Rectum* 44(2):251-8. DOI: 10.1007/BF02234301.

Gregorieff A, Clevers H. 2005. Wnt signalling in the intestinal epithelium: from endoderm to cancer. *Genes Development* 19: 877–89 DOI: 10.1101/gad.1295405.

Iacopetta B. 2003. *TP53* Mutation in Colorectal Cancer. *Human Mutation* 21:271-276 DOI: 10.1002/humu.10175.

Irby K, Anderson WF, Henson DE, Devesa SS. 2006. Emerging and widening colorectal carcinoma disparities between Blacks and Whites in the United States (1975-2002). *Cancer Epidemiology & Biomarkers Prevention*. 15(4):792-7. DOI: 10.1158/1055-9965.EPI-05-0879.

Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA, Leiserson MDM, Miller CA, Welch JS, Walter MJ, Wendl MC, Ley TJ, Wilson RK, Raphael BJ, Ding L. 2013. Mutational landscape and significance across 12 major cancer types. *Nature*. 502:333-339. DOI: 10.1038/nature12634.

Li FY, Lai MD. 2009. Colorectal cancer, one entity or three. *Journal of Zhejiang University SCIENCE B* 10(3):219-29. DOI: 10.1631/jzus.B0820273.

Lim GCC, Rampal S, Halimah Y (Eds). 2008. Cancer Incidence in Peninsular Malaysia 2003-2005. National Cancer Registry, Ministry of Health Malaysia.

Mahdavinia M, Bishehsari F, Verginelli F, Cumashi A, Lattanzio R, Satoudeh M, Ansari R, Semeraro D, Hormadzi M, Fakheri H, Rakhshani N, De Lellis L, Curia MC, Cama A, Piantelli M, Malakzadeh R, Iacobelli S, Mariani-Constantini R. 2008. P53 mutations in colorectal cancer from northern Iran: Relationships with site of tumour origin, microsatellite instability and K-ras mutations. *Journal of Cellular Physiology* 216(2): 543-550. DOI: 10.1002/jcp.21428.

Malekzadeh R, Bishehsari F, Mahdavinia M, Ansari R. 2009. Epidemiology and molecular genetics of colorectal cancer in Iran: a review. *Archives of Iranian Medicine* 12 (2): 161–169.

Manne U, Weiss HL, Myers RB, Danner OK, Moron C, Srivastava S, Grizzle WE. 1998. Nuclear accumulation of p53 in colorectal adenocarcinoma: prognostic importance differs with race and location of the tumour. *Cancer* 83:2456–2467. DOI: 10.1002/(sici)1097-0142(19981215)83:12<2456::aid-cncr8>3.0.co;2-5.

Matanoski G, Tao X, Almon L, Adade A A, Davies-Cole JO. 2006. Demographics and tumour characteristics of colorectal cancers in the United States. *Cancer* 107: 1112-1120. DOI: 10.1002/cncr.22008.

Maus MKH, Hanna DL, Stephen CL, Astrow SH, Yang D, Grimminger PP, Loupakis F, Hsiang JH, Zeger G, Wakatsuki T, Barzi A, Lenz HJ. 2015. Distinct gene expression profiles of proximal and distal colorectal cancer: implications for cytotoxic and targeted therapy. *The Pharmacogenomics Journal* 15: 354-362 DOI: 10.1038/tpj.2014.73

Meldrum C, Doyle MA & Tothill RW. 2011. Next generation sequencing for cancer diagnostics: a practical perspective. *The Clinical Biochemist Reviews* 32:177-195. DOI: PMCID: PMC3219767.

Minoo P, Zlobec I, Peterson M, Terracciano L, Lugli A. 2010. Characterization of rectal, proximal and distal colon cancers based on clinicopathological, molecular and protein profiles. *International Journal of Oncology* 37(3):707-18. DOI: 10.3892/ijo_00000720.

Missiaglia E, Jacobs B, D'Ario1 G, Di Narzo1 AF, Soneson C, Budinska E, Popovici V, Vecchione L, Gerster S, Yan P, Roth AD, Klingbiel1 D, Bosman FT, Delorenzi1 M, Tejpar S. 2014. Distal and proximal colon cancers differ in terms of molecular, pathological, and clinical features. *Annals of Oncology* 00: 1–7, 2014. DOI:10.1093/annonc/mdu275.

Moreno-Bueno G, Hardisson D, Sanchez C, Sarrio D, Cassia R, Garcia-Rostan G, Prat J, Guo M, Herman JG, Matias-Guiu X, Esteller M, Palacios J. 2002. Abnormalities of the APC/beta-catenin pathway in endometrial cancer. *Oncogene* 21:7981-7990. DOI: 10.1038/sj.onc.1205924.

Morgan OH, Maria SD, and Silvio JG. 2014. Novel insights into G protein and G protein-coupled receptor signalling in cancer. *Current Opinion in Cell Biology* 27: 126–135. DOI:10.1016/j.ceb.2014.01.005.

Myer PA, Mannalithara A, Singh G, Ladabaum U. 2012. Proximal and distal colorectal cancer resection rates in the United States since widespread screening by colonoscopy. *Gastroenterology* 5:1227-36. DOI: 10.1053/j.gastro.2012.07.107.

Nagel R, le Sage C, Diosdado B, van der Waal M, Joachim AF, Vrielink O, Bolijn A, Meijer GA, Agami R. 2008. Regulation of the adenomatous polyposis coli gene by the miR-135 family in colorectal cancer. *Cancer Research* 68:5795-5802. DOI: 10.1158/0008-5472.CAN-08-0951.

Nawa T, Kato J, Kawamoto H, Okada H, Yamamoto H, Kohno H, Endo H, Shiratori Y. 2008. Differences between right- and left-sided colon cancer in patient characteristics, cancer morphology and histology. *Journal of Gastroenterology & Hepatology*23: 418-423. DOI: 10.1111/j.1440-1746.2007.04923.x

Nikhil W, Michael FB, Matthew JD, Brendan B, Matthew DF, Panisa P, Matthew D, Paul VH, Laura EM, William CH, Matthew M, Stacey BG, Levia AG. 2012. High-throughput detection of actionable genomic alterations in clinical tumour samples by targeted, massively parallel sequencing. *Cancer Discovery* 2(1):82-93 DOI: 10.1158/2159-8290.CD-11-0184.

Pamplona RS, Doriga AL, Brunet LP, Lazaro K, Bellido F, Alonso MH, Ausso S, Guino E, Beltran S, Giner FC, Gut M, Sanjuan X, Closa A, Cordero D, Moron-Duran FD, Soriano A, Salazar R, Valle

L, Moreno V. 2015. Exome Sequencing Reveals AMER1 as a Frequently Mutated Gene in Colorectal Cancer. *Clinical Cancer Research* 21(20): 4709-4718. DOI: 10.1158/1078-0432.CCR-15-0159.

Perreault N, Katz JP, Sackett SD, Kaestner KH. 2001. Foxl1 controls the Wnt/beta-catenin pathway by modulating the expression of proteoglycans in the gut. *The Journal of Biological Chemistry* 276:43328-43333. DOI: 10.1074/jbc.M104366200.

Reinhardt HC, Aslanian AS, Lees JA, Yaffe MB. 2007. p53-deficient cells rely on ATM and ATR-mediated checkpoint signalling through the p38MAPK/MK2 pathway for survival after DNA damage. *Cancer Cell* 11: 175–189.

Reya T & Clevers H. 2005. Wnt signalling in stem cells and cancer. *Nature*. 434:843-850. DOI: 10.1038/nature03319.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. (2011). Integrative Genomics Viewer. *Nature Biotechnology*, *29*(1), 24–26. http://doi.org/10.1038/nbt.1754.

Roychowdhury S, Iyer MK, Robinson DR, Lonigro RJ, Wu YM, Cao X, Kalyana-Sundaram S, Sam L, Balbin OA, Quist MJ, Barrette T, Everett J, Siddiqui J, Kunju LP, Navone N, Araujo JC, Troncoso P, Logothetis CJ, Innis JW, Smith DC, Lao CD, Kim SY, Roberts JS, Gruber SB, Pienta KJ, Talpaz M, Chinnaiyan AM. 2011. Personalized oncology through integrative high-throughput sequencing: a pilot study. *Science Translational Medicine* 3: 111ra121. DOI: 10.1126/scitranslmed.3003161.

Segditsas S, Rowan AJ, Howarth K, Jones A, Leedham S, Wright NA, Gorman P, Chambers W, Domingo E, Roylance RR, Sawyer EJ, Sieber OM, Tomlison IP. 2009. APC and the three-hit hypothesis. *Oncogene* 28:146-155. DOI: 10.1038/onc.2008.361.

Samowitz WS, Curtin K, Ma KN, Edwards S, Schaffer D, Leppert MF, Slattery ML. 2002. Prognostic significance of p53 mutations in colon cancer at the population level. *International Journal of Cancer* 99:597–602. DOI: 10.1002/ijc.10405

Seshagiri S, Stawiski EW, Durinck S, Modrusan Z, Storm EE, Conboy. 2012. Recurrent R-spondin fusions in colon cancer. *Nature* 488: 660–664. DOI: 10.1038/nature11282.

Scholer-Dahirel A, Schlabach MR, Loo A, Bagdasarian L, Meyer Ronald, Guo R, Woolfenden S, Yu KK, Markovits J, Killary K, Sonkin D, Yao YM, Warmuth M, Sellers WR, Schlegel R, Stegmeier F, Mosher RE, McLaughlin ME. 2011. Maintenance of adenomatous polyposis coli (APC)-mutant colorectal cancer is dependent on Wnt/β-catenin signalling. *Proceeding of The National Academy of Sciences of the United States of America* 108 (41) 17135-17140. DOI: 10.1073/pnas.1104182108.

Song H, Hollstein M, Xu Y. 2007. p53 gain-of-function cancer mutants induce genetic instability by inactivating ATM. *Nature Cell Biology* 9: 573–580. DOI: 10.1038/ncb1571.

Soong R, Powell B, Elsaleh H, Gnanasampanthan G, Smith DR, Goh HS, Joseph D, Iacopetta B. 2000. Prognostic significance of TP53 gene mutation in 995 cases of colorectal carcinoma. Influence of tumour site, stage, adjuvant chemotherapy and type of mutation. *European Journal of Cancer* 36:2053–2060. DOI: 10.1016/s0959-8049(00)00285-9.

Soong R, Grieu F, Robbins P, Dix B, Chen D, Parsons R, House A, Iacopetta B. 1997. p53 alterations are associated with improved prognosis in distal colonic carcinomas. *Clinical Cancer Research* 3:1405–1411.

Sparks AB, Morin PJ, Vogelstein B, Kinzler KW. 1998. Mutational analysis of the APC/β-catenin/Tcf pathway in colorectal cancer. *Cancer Research* 58*:* 1130-1134.

The Cancer Genome Atlas Network. 2012. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487:330-337 DOI: 10.1038/nature11252.

Thorvaldsdóttir H, Robinson James T, Mesirov JP**.** 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* 14, 178-192. DOI: 10.1093/bib/bbs017.Iacopetta B. 2002. Are there two sides to colorectal cancer? *International Journal of Cancer*. 101: 403-408 DOI: 10.1002/ijc.10635.

499 Torrent Mapping Alignment Program. Available at *https://github.com/iontorrent/TMAP* (accessed on 22
500 March 2014).
501 Van Es JH, Jay P, Gregorieff A, van Gijn ME, Jonkheer S, Hatzis P, Thiele A, van den Born M, Begthel
502 H, Brabletz T, Taketo MM, Clevers H. 2005. Wnt signalling induces maturation of Paneth cells in
503 intestinal crypts. *Nature Cell Biology* 7:381386. DOI: 10.1038/ncb1240.
504 Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. 2013. Cancer genome
505 landscapes. *Science* 339:1546-1558. DOI: 10.1126/science.1235122.
506 Wang K, Li M, Hakonarson H. 2010. ANNOVAR: functional annotation of genetic variants from high-
507 throughput sequencing data. *Nucleic Acids Research* 38(16):e164. DOI: 10.1093/nar/gkq603.
508 Wang L, Tsutsumi S, Kawaguchi T, Nagasaki K, Tatsuno K, Yamamoto S, Sang F, Sonoda K, Sugawara
509 M, Saiura A, Hirono S, Yamaue H, Miki Y, Isomura M, Totoki Y, Nagae G, Isagawa T, Ueda H,
510 Murayama-Hosokawa S, Shibata T, Sakamoto H, Kanai Y, Kaneda A , Noda T, and Aburatani H.
511 2012. Whole-exome sequencing of human pancreatic cancers and characterization of genomic
512 instability caused by MLH1 haploinsufficiency and complete deficiency. *Genome Research* 22(2)
513 208–219 DOI: 10.1101/gr.123109.111
514 Wu X, Chen VW, Martin J, Roffers S, Groves FD, Correa CN, Hamilton-Byrd E, Jemal A. 2004.
515 Comparative Analysis of Incidence Rates Subcommittee, Data Evaluation and Publication
516 Committee, North American Association of Central Cancer Registries. Subsite-specific colorectal
517 cancer incidence rates and stage distributions among Asians and Pacific Islanders in the United
518 States, 1995 to 1999. *Cancer Epidemiology Biomarkers & Prevention* 13(7):1215-22.
519 Zainal Ariffin O, Nor Saleha IT. National Cancer Registry Report: Malaysia Cancer Statistics- Data and
520 Figure 2007. 2011. Ministry of Health, Malaysia. 2011.

Table 1(on next page)

Clinicopathological Characteristics of CRC Patients

Table 1. Clinicopathological Characteristics of CRC Patients

Characteristics	Number of patients, n (%)
Age	
> 50	1 (10)
> 60	9 (90)
Gender	
Male	7 (70)
Female	3 (30)
Race/Ethnicity	
Malay	4 (40)
Chinese	6 (60)
Site	
Distal	5 (50)
Proximal	5 (50)
Stage	
Dukes' A	2 (20)
Dukes' B	4 (40)
Dukes' C	4 (40)
Differentiation	
Moderately Differentiated	3 (30)
Well Differentiated	3 (30)
Poorly Differentiated	4 (40)

Table 2(on next page)

Exome Sequencing Coverage for Normal and Tumour Samples

Table 2. Exome Sequencing Coverage for Normal and Tumour Samples

Sample	Total Aligned Output (G)	# of mapped reads	% Reads On Target	Average Coverage	Uniformity (%)	Target Base Coverage at 1x (%)	Target Base Coverage at 20x (%)	Target Base Coverage at 100x (%)	Target Base Coverage at 500x (%)	Variants
L1N	13.7	38,327,122	80.75	71.39	93.48	98.03	90.72	20.66	0.11	34,706
L1T	13.7	43,581,666	84.79	84.68	93.78	98.17	92.64	29.79	0.21	35,675
L2N	13.6	37,248,504	79.6	67.21	94.07	98.5	90.28	16.84	0.14	37,472
L2T	13.6	50,527,425	76.29	88.39	94.40	98.71	93.55	30.66	0.35	37,487
L3N	14.1	45,586,546	83.14	86.35	93.33	98.08	92.29	31.86	0.21	36,077
L3T	14.1	41,217,844	82.66	77.71	93.21	98.01	91.12	25.17	0.19	35,358
L4N	10.7	38,273,507	77.96	62.24	92.97	97.85	88.15	14.38	0.1	34,197
L4T	10.7	33,604,520	78.61	55.4	92.77	97.79	85.53	10.65	0.07	33,519
L5N	12.3	41,524,901	78.49	71.81	93.48	98.04	90.67	20.48	0.16	35,113
L5T	12.3	37,647,763	79.99	66.22	93.04	97.98	88.90	17.12	0.13	34,563
R1N	12.5	40,760,456	81.92	79.26	93.65	97.95	92.03	26.01	0.17	35,213
R1T	12.5	32,407,582	82.14	62.67	92.85	97.76	87.69	15.10	0.12	32,804
R2N	12.7	36,590,333	86.11	73.47	93.83	98.12	91.43	21.70	0.15	34,752
R2T	12.7	40,262,039	86.99	80.74	93.87	98.22	92.34	26.17	0.21	34,862
R3N	11.9	38,279,157	73.01	65.69	93.76	98.06	90.18	16.24	0.09	34,870
R3T	11.9	31,589,214	79.78	58.71	93.51	97.91	88.09	11.93	0.07	35,274
R4N	14.6	40,855,511	77.97	72.79	93.01	98.02	90.25	22.26	0.12	34,516
R4T	14.6	47,748,666	78.07	85.35	93.35	98.11	92.27	31.51	0.18	39,337
R5N	11.6	37,243,895	75.27	60.64	93.54	98.04	88.53	13.48	0.06	34,497
R5T	11.6	37,455,362	75.79	62.11	92.59	97.82	87.48	15.11	0.07	33,175
Mean				**71.64**	**93.42**	**98.06**	**90.21**	**20.86**	**0.15**	**35,173**

L= Left (Distal), R= Right (Proximal), N= Normal, T= Tumour

Table 3(on next page)

Novel and Known Mutations in CRC-Related Genes in Proximal and Distal CRCs

Table 3. Novel and Known Mutations in CRC-Related Genes in Proximal and Distal CRCs

Sample ID	Gene	Start	End	Ref	Alt	Protein Change	Exonic Function	dbSNP ID	COSMIC ID	Mutation Type
R1T	APC	112175255	112175255	G	T	E1304X,E1322X	Stopgain	NA	COSM18702	Known
	TP53	7578550	7578550	G	A	S88F,S127F	Nonsynonymous	NA	COSM216414, COSM3378368, COSM44226, COSM216412, COSM216413, COSM1637542	Known
	TP53	7577022	7577022	G	A	R174X,R267X, R306X	Stopgain	rs121913344	COSM3388168, COSM10663, COSM1640820, COSM99947	Known
	KRAS	25398284	25398284	C	T	G12D	Nonsynonymous	rs121913529	COSM521, COSM1135366	Known
R2T	KRAS	25398284	25398284	C	T	G12D	Nonsynonymous	rs121913529	COSM521, COSM1135366	Known
	ACVR2A	148683686	148683686	A	-	K327fs,K435fs	Frameshift deletion	NA	COSM252949	Known
	APC	112175639	112175639	C	T	R1432X,R1450X	Stopgain	rs121913332	COSM13127	Known
R3T	ATM	108141828	108141828	A	G	Y959C	Nonsynonymous	NA	NA	Novel
	CTNNB1	41266136	41266136	T	C	S45P	Nonsynonymous	rs121913407	COSM5663	Known
	CTNNB1	41277302	41277302	A	G	R591G	Nonsynonymous	NA	NA	Novel
	IGF2-AS	2167618	2167618	A	C	T150P	Nonsynonymous	NA	NA	Novel
	KRAS	25380283	25380283	C	T	A59T	Nonsynonymous	rs121913528	COSM546, COSM1562187	Known
R4T	ACVR1B	52387841	52387841	C	T	R489C,R437C,R530C	Nonsynonymous	NA	NA	Novel
	ACVR1B	52379132	52379132	G	A	R379Q,R327Q,R420Q	Nonsynonymous	NA	NA	Novel
	ACVR2A	148683718	148683718	G	A	W337X,W445X	Stopgain	NA	NA	Novel
	ERBB2	37879658	37879658	G	A	R678Q,R648Q,R663Q	Nonsynonymous	NA	COSM436498	Known
	ERBB2	37864598	37864598	G	A	V84M,V54M,V69M	Nonsynonymous	rs376524324	NA	Known
	ERBB3	56495023	56495023	G	A	R1127H	Nonsynonymous	rs2271188	COSM1363018, COSM1363017	Known
	ERBB3	56478854	56478854	G	A	V104M	Nonsynonymous	NA	COSM172423, COSM20710, COSM1152549	Known
	FBXW7	153249456	153249456	C	T	R323Q,R361Q, R441Q	Nonsynonymous	NA	COSM1052092, COSM1052093, COSM1052091, COSM1052094	Known
	LRP5	68201247	68201247	G	A	R733H,R1314H	Nonsynonymous	NA	NA	Novel
	PTEN	89720812	89720812	A	-	T321fs	Frameshift deletion	NA	COSM5823	Known
	TCF7L2	114925436	114925436	G	A	R482Q,R505Q,R499Q	Nonsynonymous	NA	NA	Known
	TP53	7577138	7577138	C	T	R135Q,R108Q,R228Q, R267Q	Nonsynonymous	NA	COSM43923, COSM3691863, COSM1290766, COSM3691864	Known
	TP53	7578458	7578458	G	A	R26C,R119C,R158C	Nonsynonymous	NA	COSM1750371, COSM984954, COSM984957, COSM3932746, COSM984958, COSM984956, COSM43848	Known
R5T	APC	112176559	112176559	TT	GC	5214_5215GC, 5268_5269GC, 5268_5269GC	Nonframeshift Substitution	NA	NA	Novel
	ATM	108106477	108106477	G	T	G138X	Stopgain	NA	NA	Novel
	SMAD4	48573529	48573536	GAGCA ATT	-	c.113_120del:p.38_40del	Frameshift deletion	NA	NA	Novel
L1T	APC	112162896	112162896	T	G	Y482X,Y500X	Stopgain	NA	NA	Novel
L2T	APC	112175212	112175216	AAAAG	-	1289_1291del,	Frameshift deletion	NA	COSM18764	Known

	ERBB3	56478854	56478854	G	C	1307_1309del V104L	Nonsynonymous	NA	COSM160824	Known
L3T	TP53	7577121	7577121	G	A	R141C,R114C, R234C,R273C	Nonsynonymous	rs121913343	COSM3355991, COSM10659, COSM1645518, COSM99933	Known
L4T	APC	112175322	112175322	C	G	S1326X,S1344X	Stopgain	NA	NA	Novel
	FBXW7	153249385	153249385	G	A	R347C,R385C,R465C	Nonsynonymous	NA	COSM1154293, COSM170727, COSM170726, COSM170725, COSM22932	Known
L5T	CTNNB1	41266071	41266100	GTCACT GGCAG CAACA GTCTTA CCTGGA CT	-	23_33del	Nonframeshift Deletion	NA	NA	Novel
	TP53	7578263	7578263	G	A	R64X,R37X, R157X,R196X	Stopgain	NA	COSM1640847, COSM99667, COSM99666, COSM99668, COSM3378446, COSM99665, COSM10705	Known

NA = Not Available

Table 4(on next page)

Novel and Known Mutations in Most Frequently Mutated Genes in Proximal and Distal

1 Table 4. Novel and Known Mutations in Most Frequently Mutated Genes in Proximal and Distal

Gene	Sample ID	Start	End	Ref	Alt	Protein Change	Exonic Function	dbSNP ID	COSMIC ID	Mutation Type
APC	R1T	112175255	112175255	G	T	R1432X R1450X	Stopgain	NA	COSM18702	Known
	R2T	112175639	112175639	C	T	5214_5215GC 5268_5269GC	Stopgain	rs121913332	COSM13127	Known
	R5T	112176559	112176560	TT	GC	E1304X E1322X	Nonframeshift substitution	NA	NA	Novel
	L1T	112162896	112162896	T	G	Y482X Y500X	Stopgain	NA	NA	Novel
	L2T	112175212	112175216	AAAAG	-	1289_1291del 1307_1309del	Frameshift Deletion	NA	COSM18764	Known
	L4T	112175322	112175322	C	G	S1326X S1344X	Stopgain	NA	NA	Novel
TP53	R1T	7578550	7578550	G	A	S127F S88F	Nonsynonymous	NA	COSM216414, COSM3378368, COSM44226, COSM216412, COSM216413, COSM1637542	Known
	R1T	7577022	7577022	G	A	R174X R147X R306X R267X	Stopgain	rs121913344	COSM3388168,COSM10663,COSM1640820, COSM99947	Known
	R4T	7577138	7577138	C	T	R135Q G323A R108Q R267Q R228Q	Nonsynonymous	NA	COSM43923, COSM3691863, COSM1290766, COSM3691864	Known
	R4T	7578458	7578458	G	A	R26C R158C R119C	Nonsynonymous	NA	COSM1750371, COSM984954, COSM984957, COSM3932746, COSM984958, COSM984956, COSM43848	Known
	L3T	7577121	7577121	G	A	R141C R273C R234C	Nonsynonymous	rs121913343	COSM3355991, COSM10659, COSM1645518, COSM99933	Known
	L5T	7578263	7578263	G	A	R64X R37X R196X R157X	Stopgain	NA	COSM1640847, COSM99667, COSM99666, COSM99668, COSM3378446, COSM99665, COSM99668, COSM3378446, COSM99665	Known
KRAS	R1T	25398284	25398284	C	T	G12D	Non synonymous	rs121913529	COSM521, COSM1135366	Known
	R2T	25398284	25398284	C	T	G12D	Non synonymous	rs121913529	COSM521, COSM1135366	Known
	R3T	25380283	25380283	C	T	A59T	Non synonymous	rs121913528	COSM546, COSM1562187	Known
ATM	R3T	108141828	108141828	A	G	Y959C	Nonsynonymous	NA	NA	Novel
	R5T	108106477	108106477	G	T	G138X	Stopgain	NA	NA	Novel

Table 5(on next page)

Recurrent variants and Mutated Genes in Proximal and Distal CRC

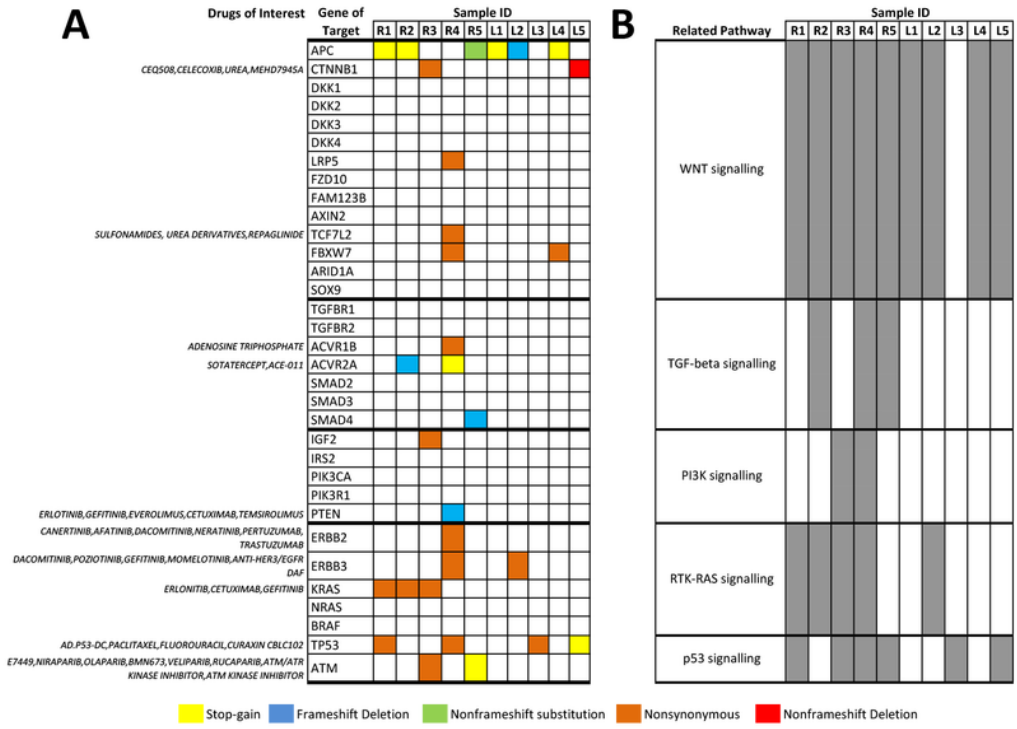
Table 5. Recurrent variants and Mutated Genes in Proximal and Distal CRC

Sample ID	Gene	Start	End	Ref	Alt	Protein Change	Exonic Function	dbSNP ID	COSMIC ID	Mutation Type
R2T, R5T	C9orf50	132377900	132377900	C	-	R248fs	Frameshift Deletion	NA	NA	Novel
R1T, R4T	GPR6	110301081	110301081	G	A	A256T, A271T	Nonsynonymous	NA	COSM3429854 COSM3429853	Known
R1T, R2T	KRAS	25398284	25398284	C	T	G12D	Nonsynonymous	rs121913529	COSM521 COSM1135366	Known
R3T, R4T	ZNF337	25657029	25657029	G	A	R299X	Stopgain	NA	NA	Novel
R3T, R4T	ZNF783	148963763	148963763	G	A	R121H	Nonsynonymous	NA	NA	Novel
L3T, L5T	CFAP74	1887018	1887019	TA	TG	2287_2288CA	Nonframeshift Substitution	NA	NA	Novel

1

Altered genes and pathway implicated in ten CRC patients (five patients of each proximal and distal CRCs).

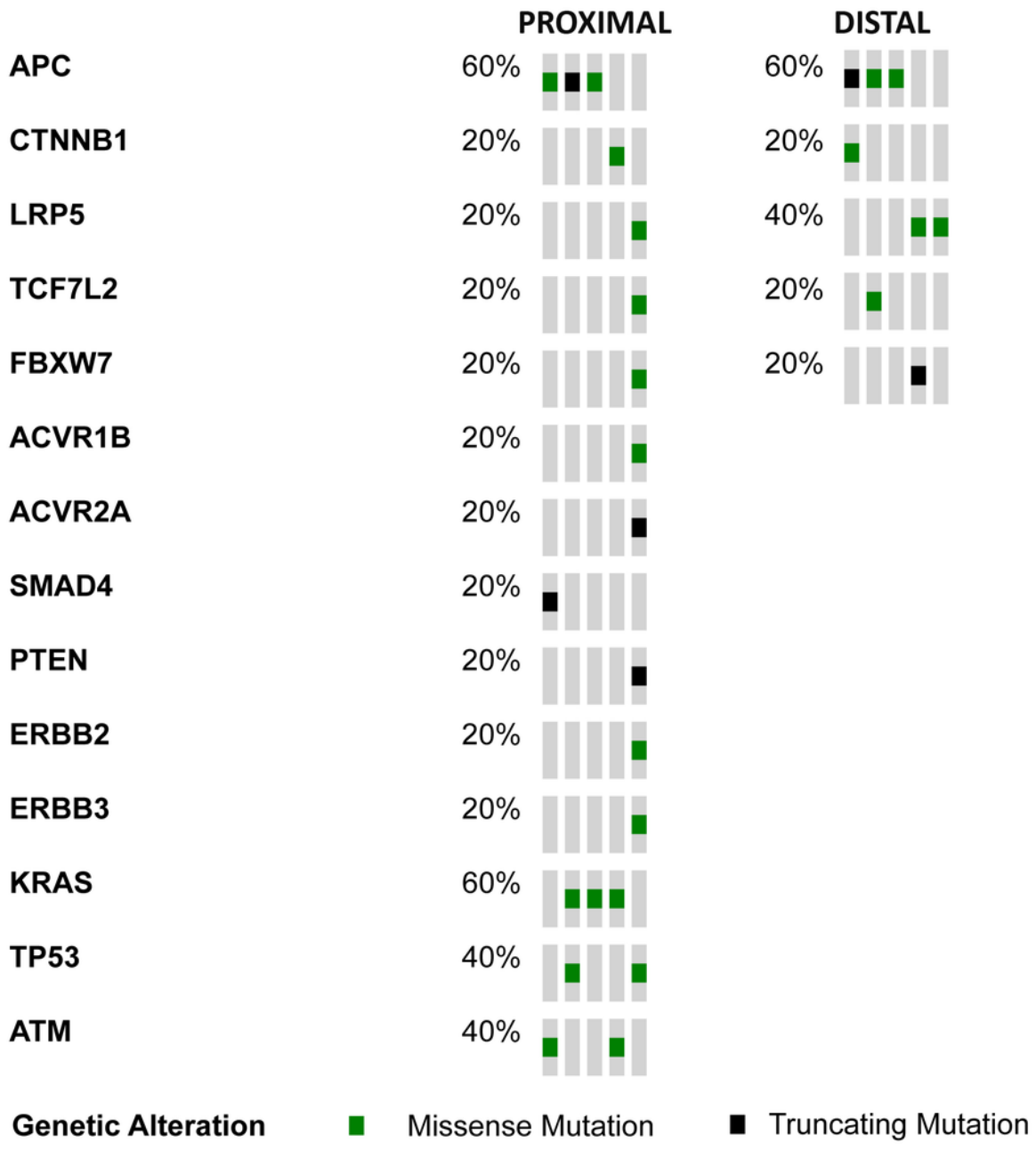
The genes were sorted according to the related pathway, separated by dark horizontal lines. The genes were associated with clinical treatment option or currently being investigated in clinical trials of novel targeted therapies.



2

Altered genes and pathway implicated in ten CRC patients (five patients of each proximal and distal CRCs). .

The genes were sorted according to the related pathway, separated by dark horizontal lines. The genes were associated with clinical treatment option or currently being investigated in clinical trials of novel targeted therapies



3

The most commonly mutated canonical pathway.

In present study, the key mutated genes detected in proximal CRC are highlighted in pink while mutated genes detected in distal CRCs are highlighted in blue.