FULL PAPER

BJC

British Journal of Cancer (2017) 117, 136–143 | doi: 10.1038/bjc.2017.148

Keywords: ulcerative colitis; targeted sequencing; ion torrent; somatic mutations

Targeted sequencing-based analyses of candidate gene variants in ulcerative colitis-associated colorectal neoplasia

Sanjiban Chakrabarty[1], Vinay Koshy Varghese[1], Pranoy Sahu[1], Pradyumna Jayaram[1], Bhadravathi M Shivakumar[1,2], Cannanore Ganesh Pai[2] and Kapaettu Satyamoorthy*[,1]

[1]*Department of Cell and Molecular Biology, School of Life Sciences, Manipal University, Manipal, Karnataka 576104, India and* [2]*Department of Gastroenterology and Hepatology, Kasturba Medical College, Manipal University, Manipal, Karnataka 576104, India*

Background: Long-standing ulcerative colitis (UC) leading to colorectal cancer (CRC) is one of the most serious and life-threatening consequences acknowledged globally. Ulcerative colitis-associated colorectal carcinogenesis showed distinct molecular alterations when compared with sporadic colorectal carcinoma.

Methods: Targeted sequencing of 409 genes in tissue samples of 18 long-standing UC subjects at high risk of colorectal carcinoma (UCHR) was performed to identify somatic driver mutations, which may be involved in the molecular changes during the transformation of non-dysplastic mucosa to high-grade dysplasia. Findings from the study are also compared with previously published genome wide and exome sequencing data in inflammatory bowel disease-associated and sporadic colorectal carcinoma.

Results: Next-generation sequencing analysis identified 1107 mutations in 275 genes in UCHR subjects. In addition to *TP53* (17%) and *KRAS* (22%) mutations, recurrent mutations in *APC* (33%), *ACVR2A* (61%), *ARID1A* (44%), *RAF1* (39%) and *MTOR* (61%) were observed in UCHR subjects. In addition, *APC*, *FGFR3*, *FGFR2* and *PIK3CA* driver mutations were identified in UCHR subjects. Recurrent mutations in *ARID1A* (44%), *SMARCA4* (17%), *MLL2* (44%), *MLL3* (67%), *SETD2* (17%) and *TET2* (50%) genes involved in histone modification and chromatin remodelling were identified in UCHR subjects.

Conclusions: Our study identifies new oncogenic driver mutations which may be involved in the transition of non-dysplastic cells to dysplastic phenotype in the subjects with long-standing UC with high risk of progression into colorectal neoplasia.

Colorectal cancer is the cancer of lower digestive system which has been implicated in patients with a history of chronic ulcerative colitis (UC), an inflammatory bowel disease (IBD) (Eaden *et al*, 2001; Xie and Itzkowitz, 2008; Navaneethan *et al*, 2013). It is the cancer of western world but has recently been reported with increasing frequencies in several South Asian countries, including India (Sung *et al*, 2005; Ray, 2011). Molecular alterations associated with sporadic form of colorectal cancer (S-CRC) are well established (Leary *et al*, 2012; Yashiro, 2015). Based on the previous studies, it is known that spectrum of mutations of colorectal tumours in patients with long-standing UC has both similar and distinct features when compared with sporadic tumours. The S-CRCs exhibits different clinicopathological features compared with that of UC-associated CRCs (UC-CRC), thus indicating unique genetic variations or mutations responsible for dysplasia (Ali *et al*, 2011). Loss of *p53* expression occurs much earlier in IBD-associated CRC compared with the sporadic form (Goretsky *et al*, 2012). *KRAS*, a proto-oncogene, has been reported

*Correspondence: Dr K Satyamoorthy; E-mail: ksatyamoorthy@manipal.edu

Received 14 December 2016; revised 25 April 2017; accepted 26 April 2017; published online 18 May 2017

with decreased mutations (8–24%) rates in UC-CRC compared with S-CRC where the mutations are reported with higher frequency (40–50%) (Tomlinson *et al*, 1998). Similarly, gene expression profiles and miRNA analysis showed distinct patterns in UC-associated and in sporadic form of colorectal carcinoma (Colliver *et al*, 2006; Olaru *et al*, 2011; Kanaan *et al*, 2012; Wu *et al*, 2014). Analysis of gene-specific methylation in UC with and without neoplasia (Konishi *et al*, 2007; Dhir *et al*, 2008; Garrity-Park *et al*, 2010) showed similar and distinct signatures when compared with S-CRC. Taking a clue from these observations, it is clear that the molecular pathways that drives the malignant progression in two scenarios have some similarities with majority of them are being unique signatures. Chronic inflammation of colonic mucosa in long-standing UC has an important role in the generation of reactive oxygen species, which in turn can aid in genomic instability. A detailed analysis of mutational spectrum using a comprehensive gene panel could help to understand the promutagenic role of inflammation and other molecular alteration favouring carcinogenesis in subjects with long-standing UC. At present, little is known about the molecular alterations that govern the colitis-derived neoplasia in several population. UC-associated CRC studies have shown sequential mutations in the *KRAS, BRAF* and *TP53* genes in an Indian population (Shivakumar *et al*, 2012; Laskar *et al*, 2015). Previously, our group has reported few molecular signatures using conventional Sanger sequencing, PCR-RFLP and multiplex PCR (Shivakumar *et al*, 2012). Subsequently, we have reported genome-wide analysis of copy number variation in UCLR, ulcerative colitis high risk (UCHR) and S-CRN, and showed significant copy number change and its association with the progression of cancer from its early stage (Shivakumar *et al*, 2015, 2016). Recently, two independent studies have demonstrated the utility of targeted sequencing of cancer gene panel to identify clinically relevant mutations from CRC patient tissues using ion torrent next-generation sequencing (NGS) platform (Singh *et al*, 2014; Malapelle *et al*, 2015). In the present study, we have performed targeted sequencing of 409 genes for the identification of driver gene mutation in long-standing UC patients with and without dysplasia, that is, progressor and non-progressor phenotype, who are at high risk of developing colorectal carcinoma.

MATERIALS AND METHODS

Study subjects. This study was approved by Ethics Committee of Kasturba Medical Hospital, Manipal University, Manipal. All the patients provided informed consent before participating in the present study. Ulcerative colitis patients ($n=26$) at high risk of associated colorectal neoplasia (UC-HR, $\geqslant 7$ years of extensive colitis or $\geqslant 10$ years of left-sided colitis), who were in remission underwent magnification chromo-colonoscopy as a part of surveillance colonoscopy from September 2008 to January 2011. These subjects were recruited and divided in two groups. Dysplasia was identified histologically according to the Riddell grading as indefinite, low grade (LGD) or high grade (HGD) on colonoscopic biopsies obtained from areas of abnormal pit pattern identified by at the time of magnifying chromo colonoscopy performed using an $\times 80$ magnification colonoscope (EC-3870 LZK, Pentax Corporations, Tokyo, Japan) (Hurlstone *et al*, 2005; Shivakumar *et al*, 2012). Ulcerative colitis high-risk subjects were subdivided as non-progressors or progressors depending on the absence or presence of neoplastic lesions on histopathology. The groups were as follows: (a) UCHR non-progressors (UCHR-NP), which consisted of 13 UC patients with high risk but without any dysplasia and (b) UCHR progressors (UCHR-P) wherein there were 13 patients with dysplasia or cancer. Tissue biopsies were collected from study subjects with UC at high risk of associated colorectal neoplasia (UC-HR, $\geqslant 7$ years of extensive colitis or $\geqslant 10$ years of left-sided colitis) for molecular analyses.

Targeted sequencing of 409 genes using ion torrent. DNA was extracted from fresh tissue biopsies using NucleoSpin tissue extraction kit (MACHEREY-NAGEL Inc., Bethlehem, PA, USA). Purity of the DNA was determined using NanoDrop ND1000 Spectrophotometer (Thermo Scientific, Waltham, MA, USA). Purified DNA was quantified using Qubit dsDNA HS (High-Sensitivity) Assay Kit (ThermoFisher Scientific, Waltham, MA, USA) and subsequently used for library preparation. Next-generation sequencing for the detection of clinically relevant mutations in the exonic region of Ion AmpliSeq comprehensive cancer panel (CCP 409 gene panel) genes in the UCHR-P and -NP patient samples was carried out using Ion Proton NGS platform.

Library preparation for ion comprehensive cancer panel genes. In the present study, we have performed targeted NGS of long-standing UC subjects ($n=18$) who are at high risk of progressing into CRC (UCHR-P = 9 and UCHR-NP = 9). Library preparation for 409 genes was performed using Ion AmpliSeq DNA library kit (ThermoFisher Scientific) along with primer pools from Ion Comprehensive Cancer Panel (4000 primer pairs in 4 pools) as per the manufacturer's instructions (ThermoFisher Scientific). Comprehensive cancer panel facilitates PCR based amplification capture and sequencing the coding regions of the 409 genes involved in various cellular pathway frequently altered during the process of carcinogenesis. The panel requires 60 ng DNA per tissue samples in four PCR reactions with each reaction utilising 15 ng DNA involving a total of 16 000 primer pairs (4000 primer pairs in each PCR reaction) for targeted sequencing of 409 genes. Each sample was barcoded using Ion Express Barcode Adapter 1–16 kit (ThermoFisher Scientific). Each library prepared were subsequently quantified using Agilent Bioanalyzer 2100 using Agilent HS DNA kit (Agilent Technologies, Santa Clara, CA, USA).

Emulsion PCR and targeted sequencing of cancer panel genes using ion proton. Automated emulsion PCR was performed to clonally amplify the CCP gene DNA library using Ion PI Template OT2 kit v2 (ThermoFisher Scientific) following the manufacturer's instruction. The enriched template positive ion sphere particles were then loaded onto a Proton PI v2 chip and sequenced using Ion proton system (ThermoFisher Scientific).

Sequencing analysis and variant calling. Sequencing reads obtained from Ion Proton were aligned to human reference sequence (human genome build-19) and variant calling was performed using Ion Torrent Variant Caller v5.0 in Torrent Suite software v5.0 (ThermoFisher Scientific). After the reads were aligned to the reference sequence, noisy and low quality reads were removed and sequence variants were detected using Ion Torrent Variant Caller Plugin software v5.0 with optimised parameters (AmpliSeq Designer, ThermoFisher Scientific) for low-frequency variant detection with minimal false-positive calls. Further, variant calls were filtered based on the technical characteristics such as (1) variant quality score, (2) variant coverage and (3) variant allele frequency. Amplicon coverage was determined via Coverage Analysis Plugin software v5.0. ThermoFisher Scientific Alignment reads and variants called, with respect to the reference human genome sequence, were viewed using Integrative Genomic Viewer software (Thorvaldsdóttir *et al*, 2013) and to check for strand biases, homopolymer length and sequencing errors. The variant caller plug-in generates a variant caller file or VCF file, which was subsequently imported to Ion Reporter software v5.0 (ThermoFisher Scientific) for variant annotation and filtering.

Functional significance of missense variants in UCHR subjects. Ion Reporter employs public databases such as, COSMIC v67, Ensembl74 and dbSNP for annotation. Variants with base call quality below 20 were filtered out. Non-coding and synonymous mutations were excluded for downstream analysis in all the UCHR

barcoded samples. CRAVAT (Cancer-Related Analysis of Variants Toolkit), a tool specifically tailored to analyse cancer-specific variants, was used for identification and prioritisation of genes with possible role in cancer tumorigenesis in each UCHR subjects (Douville *et al*, 2013). Identification and annotation of cancer-specific driver missense mutation was performed using CHASM (Cancer-specific High-throughput Annotation of Somatic Mutations) for UCHR subjects (Carter *et al*, 2009, 2010). To identify and prioritise pathogenic missense mutation, we have used Variant Effect Scoring Tool, a supervised machine learning-based classifier for each UCHR subject (Carter *et al*, 2013). cBio cancer genomics portal (http://www.cbioportal.org/) was used for the representation of the mutations in UCHR subjects (Cerami *et al*, 2012). We have taken additional care to filter germline variants as UCHR-P and -NP in high-risk UC patients were analysed against unmatched normal control. We have excluded silent and known germline variants using dbSNP, ESP (http://evs.gs.washington.edu/EVS/) and ExAC (http://exac.broadinstitute.org/) (variants with $>1\%$ minor allele frequency), and remaining missense variants were analysed with known somatic mutations using COSMIC v67, The Cancer Genome Atlas Research Network (TCGA) and whole-exome sequencing data sets for IBD-associated colorectal carcinoma. Analysis of mutation pattern was performed by collating the base substitutions into six different categories representing six different base change pattern. This was further subdivided into 16 possible combinations by extracting the trinucleotide context of the mutated base yielding 96 categories. Mutation signature, analysis of transition and transversion mutation, and mutation spectrum in UCHR subjects were performed using Maftools (Mayakonda and Koeffler, 2016).

Copy number analysis in UCHR subjects. Copy number analysis was performed by comparing normalised coverage of individual genes in UCHR subjects and in normal control using Ion Reporter software v5.0 (ThermoFisher Scientific). The patient who underwent colonoscopy for clinical symptoms and found to be normal during colonoscopy and histopathology were considered as control group. DNA was extracted from 10 male subjects with no organic colonic disease confirmed by colonoscopy and histopathological analysis, pooled at equimolar concentration and employed for copy number analysis. The quality and quantity of pooled control DNA was analysed using NanoDrop ND1000 Spectrophotometer and Qubit dsDNA HS Assay Kit.

RESULTS

Technical performance of the ion comprehensive cancer panel. Our study aimed to identify potential driver mutations in colorectal carcinogenesis process in subjects with long-standing UC at risk of CRC. To achieve this, targeted sequencing of 409 genes, involved in various molecular pathways altered during the process of carcinogenesis, were performed in 18 UCHR subjects (Supplementary Table 1). An average of 16 million mapped reads were generated in each subject with $>97\%$ reads were on target in all UCHR subjects. In our data, the average mean depth per sample was $>1000\times$ with uniformity of base coverage $>94\%$ in all UCHR subjects (Supplementary Table 2).

Mutational spectrum in UCHR subjects. Variant calling was performed using Ion Torrent Variant Caller v5.0 (TSV) in Torrent Suite software v5.0. Identified variants in each UCHR subjects were annotated and filtered with custom filters using Ion Reporter software (Supplementary Figure 1). After filtering, our bioinformatics analysis narrowed down to 1107 potentially deleterious somatic mutation including 1074 missense and 33 nonsense mutations in 275 genes in 18 UCHR subjects (Supplementary Table 3, and Figure 1A and B). We observed more number of transitions than transversion mutations in all UCHR subjects (Figure 1C). Although comparing the base change pattern for all the mutations, we observed majority of them were C to T transition as previously reported in other cancers (Kandoth *et al*, 2013) including colorectal adenocarcinoma and may occur due to spontaneous deamination of methylated cytosine (Figure 1C and D, Supplementary Figure 2). Mutational signature analysis using somatic single base substitutions identified in UCHR subjects was performed as previously discussed (Alexandrov *et al*, 2013). When compared with validated mutational signatures, it was found to be most similar to mutational signature observed in CRC with predominant $C>T$ transitions (Figure 1E). Mutations identified in UCHR subjects were plotted using CIRCOS visualisation tool (Zhang *et al*, 2013) (Figure 2).

Recurrent mutation in UCHR subjects. Overall, we identified 1107 mutations in 275 genes in all UCHR subjects. Among them, 62 genes were recurrently mutated in two or more UCHR subjects (Supplementary Table 4). Among the 62 recurrently mutated genes, oncogenes *KRAS*, *PIK3CA*, *FGFR2*, *FGFR3*, *PDGFRA* and *RAF1*, and tumour suppressor genes *APC*, *TP53*, *TET2* and *ATM* showed mutation rate of >10 mutations/Mb (Figure 3). We found *ARID1A*, *EP300* and *SMARCA4*, involved in chromatin remodelling process, were recurrently mutated in UCHR subjects (Supplementary Figure 4). *ARID1A*, a subunit of the SWI/SNF chromatin remodelling complex, regulates gene expression by controlling the accessibility of DNA to transcriptional machinery. Earlier studies have established *ARID1A* gene as novel tumour suppressor in several cancers and it has been found frequently mutated in microsatellite unstable colorectal carcinoma (Cajuso *et al*, 2014). In addition, we have identified recurrent mutation is *MLL*, *MLL2*, *MLL3*, *SETD2*, *KDM5C* and *TET2* genes involved in histone modification and DNA demethylation processes. Among them, somatic mutations in histone-lysine *N*-methyltransferase 2 (KMT2) family proteins (*MLL*, *MLL2* and *MLL3*) have been previously reported (Guo *et al*, 2013). Among the recurrently mutated genes in UCHR subjects, *ATM*, *FGFR3*, *AKAP9*, *MTOR* and *ACVR2A* are known cancer drivers identified by IntOGen-mutations analysis pipeline (Gonzalez-Perez *et al*, 2013) (Figure 3B). We compared recurrent mutations identified in UCHR subjects with recurrent mutated genes identified in IBD-associated colorectal carcinoma subjects (Robles *et al*, 2016). In addition to *TP53* and *KRAS* mutation, we have identified multiple mutations in *APC*, *MTOR*, *TRRAP* and *EP300* genes in more than one UCHR subjects (Supplementary Table 5).

Cancer driver mutations in UCHR subjects. We have applied CHASM to all the missense mutations identified in the UCHR subjects to identify driver mutation in CRC progression. Missense mutations were identified and annotated as drivers (false discovery rate <0.15) with the CHASM algorithm in the UCHR subjects. In addition, we considered all nonsense mutations and splice-site changes as drivers, as these changes are involved in structural and functional alteration of the protein products. In addition to *TP53* and *KRAS* mutations, we have identified oncogenic driver mutations in *APC*, *FGFR3*, *FGFR2*, *PDGFRA* and *PIK3CA* in UCHR subjects. Among the others *AKT1*, *ATM* and *TET2* were found to be mutated in both UCHR-P and -NP subjects (Supplementary Table 6 and Supplementary Figure 3). We compared the mutations identified in UCHR subjects with the sporadic colorectal neoplasia data reported by TCGA. When compared with TCGA-CRC data Cancer Genome Atlas Network (2012), we identified *SYNE1*, *LRP1B*, *ARID1A*, *ACVR2A*, *ABL2*, *FGFR3*, *PDGFRA*, *AKAP9* and *TAF1L* gene with high mutation frequency (40% or more) in UCHR subjects (Supplementary Table 7). Driver genes identified in UCHR subjects were stratified based on their known role in cancers using the Broad Institute's GSEA analysis. We identified 15 oncogenes (*AKT1*, *CCND1*, *CREBBP*, *EGFR*, *ERBB2*, *FGFR2*, *FGFR3*, *JAK3*, *KRAS*, *MET*, *MPL*,

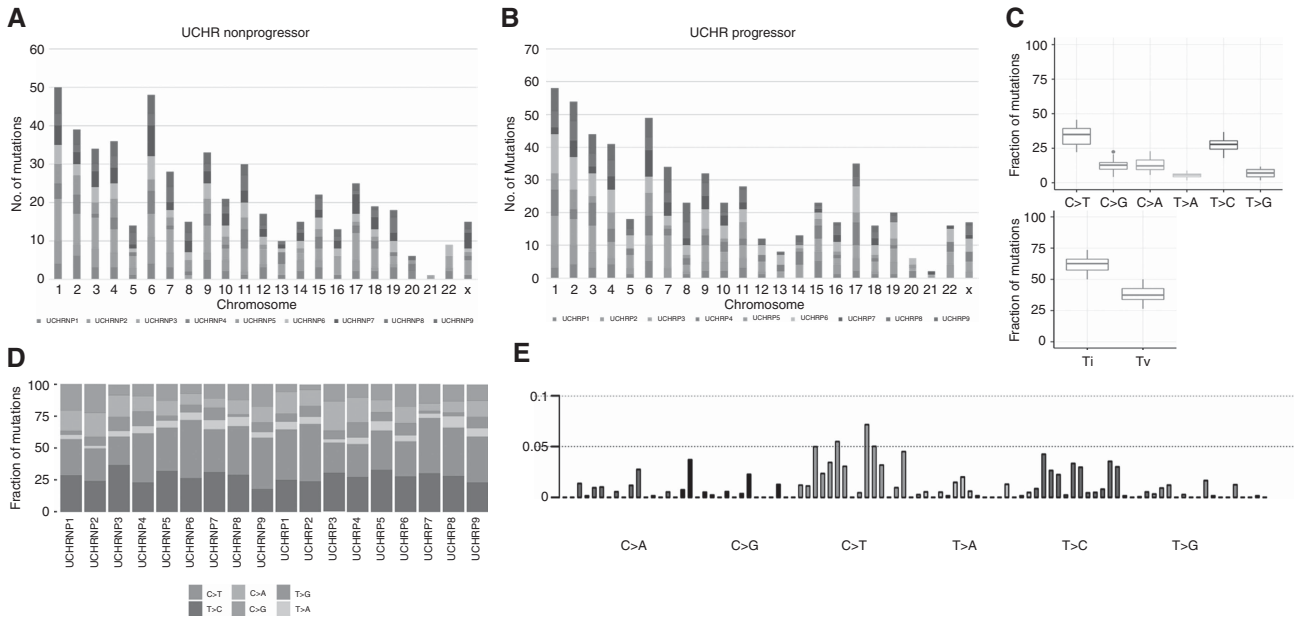

Figure 1. Mutation spectrum identified in ulcerative colitis high-risk subjects by targeted sequencing. (**A**) Chromosomal distribution of mutations identified in nine UCHR-NP subjects. (**B**) Chromosomal distribution of mutations identified in nine UCHR-P subjects. (**C**) Base substitution pattern and distribution of transition and transversion observed in 18 UCHR subjects. (**D**) Distribution of mutation frequency in UCHR-NP and -P subjects. (**E**) Mutational signature analysis using somatic single base substitutions identified in UCHR subjects. A full color version of this figure is available at the *British Journal of Cancer journal* online.

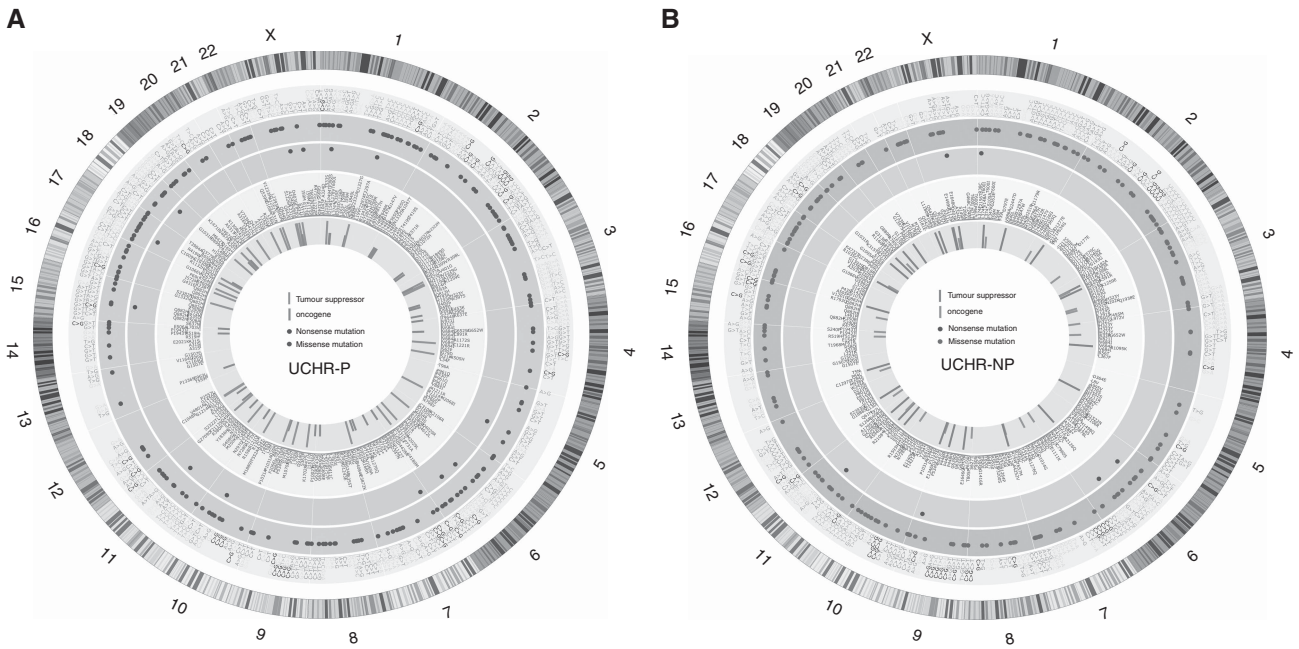

Figure 2. Genome-wide representation of somatic mutations identified in UCHR subjects. (**A**) Somatic mutations identified in UCHR-Ps. (**B**) Somatic mutations identified in UCHR-NPs. Mutations identified in UCHR subjects were plotted using CIRCOS visualisation tool. A full color version of this figure is available at the *British Journal of Cancer journal* online.

PDGFRA, *PIK3CA*, *RUNX1* and *TSHR*) and 10 tumour suppressor genes (*APC*, *ASXL1*, *ATM*, *KDM5C*, *MSH6*, *PTCH1*, *SMARCA4*, *STK11*, *TET2* and *TP53*) with potential driver mutations in UCHR subjects (Supplementary Table 8).

Gene Ontology and pathway altered in UCHR subjects. Gene Ontology analysis identified enrichment in ontologies associated with DNA recombination, DNA repair, cell differentiation, immune response and cell cycle (Supplementary Table 9).

Mutated genes were significantly enriched for most of the major cancer pathways including CRC, prostate cancer, pancreatic cancer, endometrial cancer, melanoma and for glioma pathways (Supplementary Table 10). We found that WNT signalling pathway was altered, which includes inactivating mutation in *APC* in 6 out of 18 UCHR subjects, *CTNNB1* mutation in 2 out of 18 and *EP300* mutations in 3 out of 18 UCHR subjects respectively. Similar to the observations made in previous CRC studies, we noticed recurrent mutation in genes belong to PI3K-Akt signalling

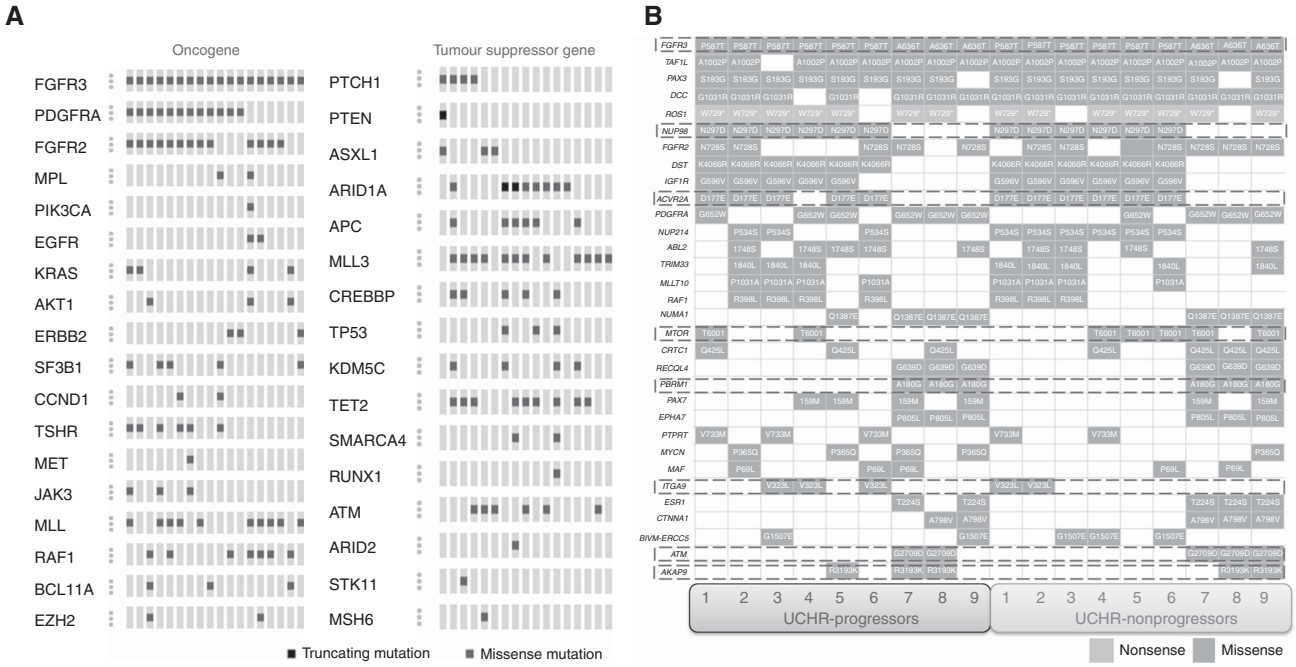

Figure 3. Recurrently mutated oncogene and tumour suppressor genes identified in UCHR subjects. (**A**) Recurrent mutation identified in oncogene and tumour suppressor genes in UCHR subjects. (**B**) Recurrent mutations identified in genes in UCHR-P and -NP subjects. Genes highlighted in red boxes are reported as cancer drivers by IntOGen database. A full color version of this figure is available at the *British Journal of Cancer* journal online.

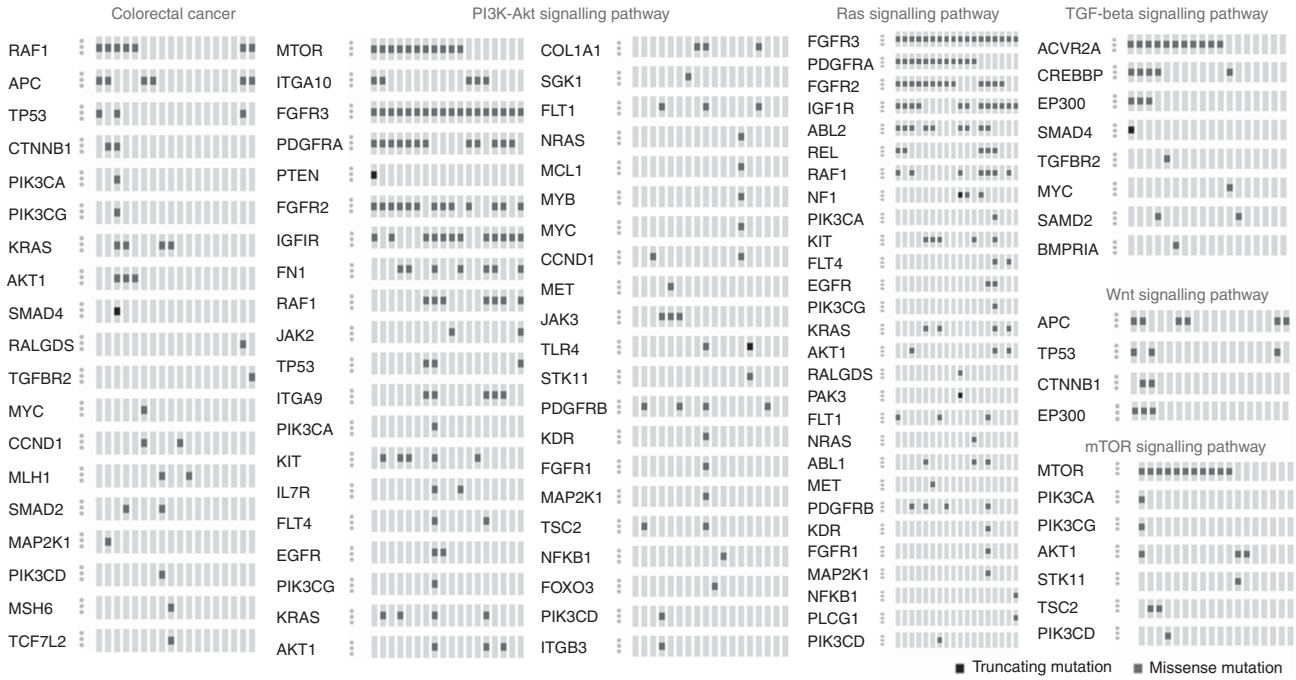

Figure 4. Distributions of somatic mutations identified in UCHR subjects in various molecular pathways. The diagram was prepared in OncoPrinter tool using cBio cancer genomics portal (http://www.cbioportal.org/). A full color version of this figure is available at the *British Journal of Cancer journal* online.

pathway (*AKT1*, *FGFR2*, *FGFR3*, *IGF1R*, *PDGFRA*, *TP53* and *RAF1*), TGF-β receptor pathway (*ACVR2A*, *SMAD2*/4 and *TGFBR2*), Ras signalling (*ABL2*, *KRAS*, *FGFR2*, *FGFR3*, *PDGFRA*, *IGF1R* and *RAF1*), mTOR signalling pathway (*MTOR*, *AKT1*, *PIK3CA*, *PIK3CG*, *PIK3CD*, *STK11* and *TSC2*) (Figure 4). We observed *TRRAP*, *EP300*, *MLL*, *MLL2* and *MLL3* genes mutated in UCHR subjects involved in chromatin remodelling. *MLL3* gene frameshift mutation with loss of expression has been previously reported in gastric and CRC (Je *et al*, 2013). *EP300* and *TRRAP* are involved in chromatin remodelling via histone acetylation facilitating p53-mediated transcription.

Recurrent mutation in chromatin remodelling complex proteins and histone-modifying proteins in UCHR subjects. Chromatin remodelling is a dynamic process that regulates transcription in normal cell. Previous studies have shown genetic alteration and aberrant regulation of ATP-dependent chromatin remodelling complex proteins and nucleosome histone-modifying complexes

have been identified as one of the key process facilitating carcinogenesis (Nakazato *et al*, 2016). Epigenetic alteration, which includes histone methylation, has an important role in carcinogenesis. Altered expression of several histone methyltransferase enzymes have been implicated in several cancers including CRC. Human KMT2 family, also known as mixed-lineage leukaemia (MLL) family, initially discovered in this disease. Recently, several genome-wide and targeted sequencing studies have revealed widespread mutations in KMT2 gene family in many types of human cancers. Ongoing studies identified the link between KMT2 gene family deregulation and its roles in tumorigenesis (Rao and Dou, 2015). In our study, we identified recurrent mutations in *MLL* (*KMT2A*), *MLL2* (*KMT2D*) and *MLL3* (*KMT2C*) in UCHR subjects. In addition, we identified mutation in *ARID1A*, *ATRX*, *SETD2*, *PRDM1*, *EP300* and *NSD1* gene involved in regulation of gene expression by histone modification and chromatin remodelling (Supplementary Table 3 and Supplementary Figure 4A and B). In addition, *SMARCA4* gene, a member of the SWI/SNF family of proteins, was mutated in three out of nine UCHR-P subjects (Supplementary Figure 4A and B). Mutation in *SMARCA4* gene has been reported in lung cancer and ovarian cancer (Rodriguez-Nieto *et al*, 2011). We have further identified a heterozygous insertion mutation in the *MLL3* (*KMT2C*) gene on chromosome 7 (151945071 bp, insT; Human Genome assembly GRCh37/hg19) in both UCHR-P and -NP subjects. *MLL3* gene insertion, which starts at codon 817 in exon 14, results a frameshift mutation causing premature stop codon at codon 827. In addition, we identified recurrent *MLL3* somatic mutations in all the UCHR subjects (Supplementary Figure 4A).

Copy number variants identified in UCHR subjects. Analysis of relative copy number in UCHR subjects were also assessed from the targeted sequencing data. We have identified 10 copy number-amplified regions with more than 3 copy number in UCHR subjects (Supplementary Table 11). Copy number-amplified regions identified in our study include *MYC* oncogene (8q24.21), *ERBB2* (17q12), *MYCN* (2p24.3) and *IRS2* (13q34) reported previously in IBD-associated colorectal carcinoma (Robles *et al*, 2016). *MYC* gene amplification was previously identified in UCHR subjects by our group (Shivakumar *et al*, 2016). *MET* proto-oncogene and *ABL1* oncogene amplification, previously reported in colon cancer, have been observed in UCHR subjects. Although *ABL1* oncogene amplification is observed in both UCHR-P and -NP, *MET* proto-oncogene amplification is observed only in UCHR-Ps ($n = 3$). We observed copy number gain in 20q12 harbouring topoisomerase 1 (*TOP1*), receptor-type tyrosine-protein phosphatase T (*PTPRT*) and phospholipase C, gamma 1 (*PLCG1*) in UCHR subjects. Chromosome 20q12 gain is commonly observed in sporadic colorectal carcinoma and may play a possible role in transformation of adenoma to carcinoma (Carvalho *et al*, 2009). *PTPRT* functions as an oncogene and has been reported to be recurrently amplified in colorectal carcinoma (Laczmanska *et al*, 2014).

DISCUSSION

Long-standing UC with LGD experiences a higher incidence of colorectal carcinoma than the rest of the population (Herrinton *et al*, 2012; Choi *et al*, 2015). Our study population belongs to region with low incidence of both UC and UC-associated CRC. Moreover, little is known about the molecular alterations aiding the malignant transformation in UC patients in Indian population. Recently, few reports have highlighted close association of molecular signature patterns identified in CRC in India and that of the western population (Gupta *et al*, 2010; Laskar *et al*, 2015). In colorectal carcinoma, the malignant transformation occurs by acquiring series of driver mutations in sequential manner. Recent NGS studies in sporadic CRCs have extensively classified driver mutations accumulated during progression of CRC. Molecular alterations commonly observed include (a) defective DNA mismatch repair with microsatellite instability, (b) somatic copy number alterations (8q, 13q, 17q, 20q), (c) activating mutation in *KRAS* and *PIK3CA* oncogene and (d) inactivating mutation and LOH in *APC* and *TP53* tumour suppressor genes. Although, some of the cellular pathways involved in carcinogenesis are same in sporadic and UC-associated colorectal carcinoma, recent studies highlighted that *APC* mutations appear early, whereas *TP53* mutation appears late in sporadic colorectal carcinoma. Similarly, *TP53* mutations is an early event and *APC* inactivating mutation is seen in later stages of UC-associated colorectal carcinoma (Ullman and Itzkowitz, 2011; Carethers and Jung, 2015; Robles *et al*, 2016).

The objective of our study was to identify (a) common molecular alteration in the genes in UCHR subjects previously associated in UC-associated and sporadic colorectal carcinoma, and (b) unique molecular alteration in the genes that are not frequently altered in sporadic colorectal carcinoma however may have a role in UC-associated carcinogenesis. Long-standing UC subjects were classified by using magnification chromo-colonoscopy into UCHR-P with neoplastic lesions and UCHR-NP were without neoplastic lesions. Targeted sequencing of 409 genes performed in 18 UCHR subjects with and without neoplastic lesion identified 1107 somatic mutations in 275 genes implicated in various molecular pathways, which are altered in UC-associated and S-CRC. Somatic mutation in *APC* was observed in 6 out of 18 UCHR subjects. *APC* mutations with altered WNT signalling pathway is frequently observed in sporadic colorectal carcinoma subjects. We have identified a nonsense mutation with premature stop codon in *APC* (R1114*) in one UCHR subject with HGD. Among the other genes in WNT signalling pathway, *CTNNB1* mutation was found in 2 out of 18 UCHR subjects. *KRAS* mutation (G13D, G12D and Q61H) was observed in 4 out of 18 UCHR subjects. Although *TP53* mutation was found in 4 UCHR subjects with HGD (UCHR-P), which is commonly seen as an early molecular event that occurs during UC-CRC tumorigenic process. However, we did not observe *TP53* mutation in UCHR subjects without dysplastic phenotype (UCHR-NP), which could possibly be because of the small sample size. In our study, we did not find *BRAF* mutation in any of the UCHR subjects. Based on the previous reports, frequency of *BRAF* mutations is less in sporadic CRCs (6–11%) and in UC-associated colorectal neoplasia (9%) (Aust *et al*, 2005). It is reported that *BRAF* mutations frequency is higher in microsatellite instable CRCs developed through the mutator or serrated pathway (Aust *et al*, 2005). Frequent somatic mutations in *TGFBR2* and *SMAD4*, members of canonical TGF-β signalling pathway, were reported in CRCs. In our study, *TGFBR2* missense mutation and a nonsense mutation with premature stop codon in *SMAD4* (Q169*) was found in two UCHR subjects with high grade dysplasia.

Deregulation of histone lysine methyltransferase activity leading to aberrant chromatin remodelling has been reported in various cancers (Rao and Dou, 2015). It is known that altered chromatin modification contribute to aberrant cell proliferation *via* altered expression pattern of oncogenes and tumour suppressor genes or by inducing genome instability (Benard *et al*, 2014). Mutation in *MLL2*, encoding a histone methyltransferase, has been implicated in substantial transcription stress and genomic instability (Kantidakis *et al*, 2016). *MLL3* gene germline insertion mutation was previously reported in a pedigree of CRC (Li *et al*, 2013). In our study, we observed recurrent mutations in *MLL* (*KMT2A*, H3K4me1/2/3), *MLL2* (*KMT2D*, H3K4me1/2/3) and *MLL3* (*KMT2C*, H3K4me1/2/3) in UCHR subjects. Mutation in *NSD1* (H3K36me1/2/3), *SETD2* (H3K36me1/2/3) and *PRDM1* (H3K9me1/2/3) was also observed in UCHR subjects. Massively parallel sequencing of cancer genome projects involving large number of subjects have identified recurrent mutation in mammalian SWI/

SNF complex and its associated proteins. SWI/SNF nucleosome remodelling complex is a known tumour suppressor in many human malignancies. We have identified frequent somatic missense mutation in *ARID1A* and *SMARCA4* in UCHR subjects. *ARID1A* has been previously reported to be frequently mutated in colon and rectal carcinoma and is a known epigenetic tumour suppressor gene. Recent study has shown *ARID1A* deficiency promotes colon cancer in animal model (Mathur *et al*, 2017). Together, these finding suggests an in-depth interrogation is required to understand the widespread mutation in epigenetic regulators, chromatin remodelling complex genes and their contribution in epigenetic regulation of histone, transcriptional stress, and genome instability during colorectal carcinogenesis.

Several mutations and copy number alteration, which are already established as oncogenic driver genes in the colorectal carcinogenesis have been identified in our study. These include *AKT1* and *RAF1* (EGF receptor signalling pathway) *APC*, *CTNNB1* and *TP53* (WNT signalling pathway), *FGFR2*, *FGFR3*, *AKT1*, *RAF1* (FGFR signalling pathway), *PIK3CA* mutation along with *IRS2*, *ERBB2*, *EGFR*, *MYC*, *MET* and *ABL1* gene amplifications in UCHR subjects. In addition, we have observed *PTPRT* (20q12) gene amplification in 10 out of 18 UCHR subjects. *PTPRT* belongs to tyrosine phosphatase gene superfamily and has been previously reported to have dual roles as oncogene and tumour suppressor gene in a context-dependent manner in human cancers. Previously, *PTPRT* gene was reported to be frequently amplified in sporadic colorectal carcinoma without *BRAF* mutation (Laczmanska *et al*, 2014).

From a low resource country, this is one of the few attempts to study by targeted sequencing-based comprehensive analysis of mutation spectrum in cancer-associated genes for the identification of driver mutation in long-standing UC subjects. Data generated from this study highlight a possible mechanism of early molecular changes that occur during transformation of colitis to low grade dysplasia into carcinoma. Although the study population represents a low incidence rate of UC-associated colorectal carcinoma, findings from the study could be tested in a larger cohort to assess the role of somatic driver mutations and its potential for an early molecular signature. To complement our findings, an in-depth analysis of gene expression and epigenetic regulation of driver genes could facilitate deeper perspective of the pathophysiology of CRC and may also identify potential therapeutic targets.

ACKNOWLEDGEMENTS

This work was supported by Department of Biotechnology, Government of India, (BT/01/COE/06/02/07), DST-FIST and TIFAC-CORE. We thank Manipal University for providing the infrastructure to conduct the study. We gratefully thank the patients for participating in this study.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Børresen-Dale AL, Boyault S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjörd JE, Foekens JA, Greaves M, Hosoda F, Hutter B, Ilicic T, Imbeaud S, Imielinski M, Jäger N, Jones DT, Jones D, Knappskog S, Kool M, Lakhani SR, López-Otín C, Martin S, Munshi NC, Nakamura H, Northcott PA, Pajic M, Papaemmanuil E, Paradiso A, Pearson JV, Puente XS, Raine K, Ramakrishna M, Richardson AL, Richter J, Rosenstiel P, Schlesner M, Schumacher TN, Span PN, Teague JW, Totoki Y, Tutt AN, Valdés-Mas R, van Buuren MM, van 't Veer L, Vincent-Salomon A, Waddell N, Yates LR, Australian Pancreatic Cancer Genome Initiative, ICGC Breast Cancer Consortium, ICGC MMML-Seq Consortium, ICGC PedBrain, Zucman-Rossi J, Futreal PA, McDermott U, Lichter P, Meyerson M, Grimmond SM, Siebert R, Campo E, Shibata T, Pfister SM, Campbell PJ, Stratton MR (2013) Signatures of mutational processes in human cancer. *Nature* **500**: 415–421.

Ali RA, Dooley C, Comber H, Newell J, Egan LJ (2011) Clinical features, treatment, and survival of patients with colorectal cancer with or without inflammatory bowel disease. *Clin Gastroenterol Hepatol* **9**: 584–589. e1-e2.

Aust DE, Haase M, Dobryden L, Markwarth A, Löhrs U, Wittekind C, Baretton GB, Tannapfel A (2005) Mutations of the BRAF gene in ulcerative colitis-related colorectal carcinoma. *Int J Cancer* **115**: 673–677.

Benard A, Goossens-Beumer IJ, van Hoesel AQ, de Graaf W, Horati H, Putter H, Zeestraten EC, van de Velde CJ, Kuppen PJ (2014) Histone trimethylation at H3K4, H3K9 and H4K20 correlates with patient survival and tumor recurrence in early-stage colon cancer. *BMC Cancer* **14**: 531.

Cajuso T, Hänninen UA, Kondelin J, Gylfe AE, Tanskanen T, Katainen R, Pitkänen E, Ristolainen H, Kaasinen E, Taipale M, Taipale J, Böhm J, Renkonen-Sinisalo L, Mecklin JP, Järvinen H, Tuupanen S, Kilpivaara O, Vahteristo P (2014) Exome sequencing reveals frequent inactivating mutations in ARID1A, ARID1B, ARID2 and ARID4A in microsatellite unstable colorectal cancer. *Int J Cancer.* **135**: 611–623.

Cancer Genome Atlas Network (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**: 330–337.

Carethers JM, Jung BH (2015) Genetics and genetic biomarkers in sporadic colorectal cancer. *Gastroenterology* **149**: 1177–1190.e3.

Carter H, Chen S, Isik L, Tyekucheva S, Velculescu VE, Kinzler KW, Vogelstein B, Karchin R (2009) Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. *Cancer Res* **69**: 6660–6667.

Carter H, Douville C, Stenson P, Cooper D, Karchin R (2013) Identifying Mendelian disease genes with the Variant Effect Scoring Tool. *BMC Genomics* **14**: S3.

Carter H, Samayoa J, Hruban RH, Karchin R (2010) Prioritization of driver mutations in pancreatic cancer using cancer-specific high-throughput annotation of somatic mutations (CHASM). *Cancer Biol Ther* **10**: 582–587.

Carvalho B, Postma C, Mongera S, Hopmans E, Diskin S, van de Wiel MA, van Criekinge W, Thas O, Matthäi A, Cuesta MA, Terhaar Sive Droste JS, Craanen M, Schröck E, Ylstra B, Meijer GA (2009) Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression. *Gut* **58**: 79–89.

Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**: 401–404.

Choi CH, Ignjatovic-Wilson A, Askari A, Lee GH, Warusavitarne J, Moorghen M, Thomas-Gibson S, Saunders BP, Rutter MD, Graham TA, Hart AL (2015) Low-grade dysplasia in ulcerative colitis: risk factors for developing high-grade dysplasia or colorectal cancer. *Am J Gastroenterol* **110**: 1461–1471.

Colliver DW, Crawford NP, Eichenberger MR, Zacharius W, Petras RE, Stromberg AJ, Galandiuk S (2006) Molecular profiling of ulcerative colitis-associated neoplastic progression. *Exp Mol Pathol* **80**: 1–10.

Dhir M, Montgomery EA, Glöckner SC, Schuebel KE, Hooker CM, Herman JG, Baylin SB, Gearhart SL, Ahuja N (2008) Epigenetic regulation of WNT signaling pathway genes in inflammatory bowel disease (IBD) associated neoplasia. *J Gastrointest Surg* **12**: 1745–1753.

Douville C, Carter H, Kim R, Niknafs N, Diekhans M, Stenson PD, Cooper DN, Ryan M, Karchin R (2013) CRAVAT: Cancer-Related Analysis of VAriants Toolkit. *Bioinformatics* **29**: 647–648.

Eaden J, Abrams K, Mayberry J (2001) The risk of colorectal cancer in ulcerative colitis: a meta-analysis. *Gut* **48**: 526–535.

Garrity-Park MM, Loftus Jr EV, Sandborn WJ, Bryant SC, Smyrk TC (2010) Methylation status of genes in non-neoplastic mucosa from patients with ulcerative colitis-associated colorectal cancer. *Am J Gastroenterol* **105**: 1610–1619.

Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Tamborero D, Schroeder MP, Jene-Sanz A, Santos A, Lopez-Bigas N (2013) IntOGen-mutations identifies cancer drivers across tumor types. *Nat Methods* **10**: 1081–1082.

Goretsky T, Dirisina R, Sinh P, Mittal N, Managlia E, Williams DB, Posca D, Ryu H, Katzman RB, Barrett TA (2012) p53 mediates TNF-induced epithelial cell apoptosis in IBD. *Am J Pathol* **181**: 1306–1315.

Guo C, Chen LH, Huang Y, Chang CC, Wang P, Pirozzi CJ, Qin X, Bao X, Greer PK, McLendon RE, Yan H, Keir ST, Bigner DD, He Y (2013) KMT2D maintains neoplastic cell proliferation and global histone H3 lysine 4 monomethylation. *Oncotarget* **4**: 2144–2153.

Gupta S, Bhattacharya D, Acharya A, Majumdar S, Ranjan P, Das S (2010) Colorectal carcinoma in young adults: a retrospective study on Indian patients: 2000–2008. *Colorectal Dis* **12**: e182–e189.

Herrinton LJ, Liu L, Levin TR, Allison JE, Lewis JD, Velayos F (2012) Incidence and mortality of colorectal adenocarcinoma in persons with inflammatory bowel disease from 1998 to 2010. *Gastroenterology* **143**: 382–389.

Hurlstone DP, Sanders DS, Lobo AJ, McAlindon ME, Cross SS (2005) Indigo carmine-assisted high -magnification chromoscopic colonoscopy for the detection and characterisation of intraepithelial neoplasia in ulcerative colitis: a prospective evaluation. *Endoscopy* **37**: 1186–1192.

Je EM, Lee SH, Yoo NJ, Lee SH (2013) Mutational and expressional analysis of MLL genes in gastric and colorectal cancers with microsatellite instability. *Neoplasma* **60**: 188–195.

Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA, Leiserson MD, 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.

Kanaan Z, Rai SN, Eichenberger MR, Barnes C, Dworkin AM, Weller C, Cohen E, Roberts H, Keskey B, Petras RE, Crawford NP, Galandiuk S (2012) Differential microRNA expression tracks neoplastic progression in inflammatory bowel disease-associated colorectal cancer. *Hum Mutat* **33**: 551–560.

Kantidakis T, Saponaro M, Mitter R, Horswell S, Kranz A, Boeing S, Aygün O, Kelly GP, Matthews N, Stewart A, Stewart AF, Svejstrup JQ (2016) Mutation of cancer driver MLL2 results in transcription stress and genome instability. *Genes Dev* **30**: 408–420.

Konishi K, Shen L, Wang S, Meltzer SJ, Harpaz N, Issa JP (2007) Rare CpG island methylator phenotype in ulcerative colitis-associated neoplasias. *Gastroenterology* **132**: 1254–1260.

Laczmanska I, Karpinski P, Kozlowska J, Bebenek M, Ramsey D, Sedziak T, Ziolkowski P, Sasiadek MM (2014) Copy number alterations of chromosomal regions enclosing protein tyrosine phosphatase receptor-like genes in colorectal cancer. *Pathol Res Pract* **210**: 893–896.

Laskar RS, Ghosh SK, Talukdar FR (2015) Rectal cancer profiling identifies distinct subtypes in India based on age at onset, genetic, epigenetic and clinicopathological characteristics. *Mol Carcinog* **54**: 1786–1795.

Leary RJ, Sausen M, Kinde I, Papadopoulos N, Carpten JD, Craig D, O'Shaughnessy J, Kinzler KW, Parmigiani G, Vogelstein B (2012) Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. *Sci Transl Med* **4**: 162ra154.

Li WD, Li QR, Xu SN, Wei FJ, Ye ZJ, Cheng JK, Chen JP (2013) Exome sequencing identifies an MLL3 gene germ line mutation in a pedigree of colorectal cancer and acute myeloid leukemia. *Blood* **121**: 1478–1479.

Malapelle U, Vigliar E, Sgariglia R, Bellevicine C, Colarossi L, Vitale D, Pallante P, Troncone G (2015) Ion Torrent next-generation sequencing for routine identification of clinically relevant mutations in colorectal cancer patients. *J Clin Pathol* **68**: 64–68.

Mathur R, Alver BH, San Roman AK, Wilson BG, Wang X, Agoston AT, Park PJ, Shivdasani RA, Roberts CW (2017) ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice. *Nat Genet* **49**: 296–302.

Mayakonda A, Koeffler HP (2016) Maftools: efficient analysis, visualization and summarization of MAF files from large-scale cohort based cancer studies. *bioRxiv*.

Nakazato H, Takeshima H, Kishino T, Kubo E, Hattori N, Nakajima T, Yamashita S, Igaki H, Tachimori Y, Kuniyoshi Y, Ushijima T (2016) Early-stage induction of SWI/SNF mutations during esophageal squamous cell carcinogenesis. *PLoS One* **11**: e0147372.

Navaneethan U, Jegadeesan R, Gutierrez NG, Venkatesh PG, Hammel JP, Shen B, Kiran RP (2013) Progression of low-grade dysplasia to advanced neoplasia based on the location and morphology of dysplasia in ulcerative colitis patients with extensive colitis under colonoscopic surveillance. *J Crohns Colitis* **7**: e684–e691.

Olaru AV, Selaru FM, Mori Y, Vazquez C, David S, Paun B, Cheng Y, Jin Z, Yang J, Agarwal R, Abraham JM, Dassopoulos T, Harris M, Bayless TM, Kwon J, Harpaz N, Livak F, Meltzer SJ (2011) Dynamic changes in the expression of MicroRNA-31 during inflammatory bowel disease-associated neoplastic transformation. *Inflamm Bowel Dis* **17**: 221–231.

Rao RC, Dou Y (2015) Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. *Nat Rev Cancer* **15**: 334–346.

Ray G (2011) Inflammatory bowel disease in India—changing paradigms. *Int J Colorectal Dis* **26**: 635–644.

Robles AI, Traverso G, Zhang M, Roberts NJ, Khan MA, Joseph C, Lauwers GY, Selaru FM, Popoli M, Pittman ME, Ke X, Hruban RH, Meltzer SJ, Kinzler KW, Vogelstein B, Harris CC, Papadopoulos N (2016) Whole-exome sequencing analyses of inflammatory bowel disease-associated colorectal cancers. *Gastroenterology* **150**: 931–943.

Rodriguez-Nieto S, Cañada A, Pros E, Pinto AI, Torres-Lanzas J, Lopez-Rios F, Sanchez-Verde L, Pisano DG, Sanchez-Cespedes M (2011) Massive parallel DNA pyrosequencing analysis of the tumor suppressor BRG1/SMARCA4 in lung primary tumors. *Hum Mutat* **32**: E1999–E2017.

Shivakumar BM, Kumar BL, Bhat G, Suvarna D, Rao L, Pai CG, Satyamoorthy K (2012) Molecular alterations in colitis-associated colorectal neoplasia: study from a low prevalence area using magnifying chromo colonoscopy. *J Crohns Colitis* **6**: 647–654.

Shivakumar BM, Rotti H, Vasudevan TG, Balakrishnan A, Chakrabarty S, Bhat G, Rao L, Pai CG, Satyamoorthy K (2015) Copy number variations are progressively associated with the pathogenesis of colorectal cancer in ulcerative colitis. *World J Gastroenterol* **21**: 616–622.

Shivakumar BM, Chakrabarty S, Rotti H, Seenappa V, Rao L, Geetha V, Tantry BV, Kini H, Dharamsi R, Pai CG, Satyamoorthy K (2016) Comparative analysis of copy number variations in ulcerative colitis associated and sporadic colorectal neoplasia. *BMC Cancer* **16**: 271.

Singh R, Patel K, Routbort M, Aldape K, Lu X, Manekia J, Abraham R, Reddy N, Barkoh B, Veliyathu J (2014) Clinical massively parallel next-generation sequencing analysis of 409 cancer-related genes for mutations and copy number variations in solid tumours. *Br J Cancer* **111**: 2014–2023.

Sung JJ, Lau JY, Goh KL, Leung WK. Asia Pacific Working Group on Colorectal Cancer (2005) Increasing incidence of colorectal cancer in Asia: implications for screening. *Lancet Oncol* **6**: 871–876.

Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**: 178–192.

Tomlinson I, Ilyas M, Johnson V, Davies A, Clark G, Talbot I, Bodmer W (1998) A comparison of the genetic pathways involved in the pathogenesis of three types of colorectal cancer. *J Pathol* **184**: 148–152.

Ullman TA, Itzkowitz SH (2011) Intestinal inflammation and cancer. *Gastroenterology* **140**: 1807–1816.

Wu CW, Ng SC, Dong Y, Tian L, Ng SS, Leung WW, Law WT, Yau TO, Chan FK, Sung JJ, Yu J (2014) Identification of microRNA-135b in stool as a potential noninvasive biomarker for colorectal cancer and adenoma. *Clin Cancer Res* **20**: 2994–3002.

Xie J, Itzkowitz SH (2008) Cancer in inflammatory bowel disease. *World J Gastroenterol* **14**: 378–389.

Yashiro M (2015) Molecular alterations of colorectal cancer with inflammatory bowel disease. *Dig Dis Sci* **60**: 2251–2263.

Zhang H, Meltzer P, Davis S (2013) RCircos: an R package for Circos 2D track plots. *BMC Bioinformatics* **14**: 244.

Supplementary Information accompanies this paper on British Journal of Cancer website (http://www.nature.com/bjc)