Mutational Analysis Identifies Therapeutic Biomarkers in Inflammatory Bowel Disease–Associated Colorectal Cancers

Shahida Din1, Kim Wong2, Mike F. Mueller3, Anca Oniscu4, James Hewinson2, Catherine J. Black5, Martin L. Miller6, Alejandro Jimenez-Sanchez6, Roy Rabbie2, Mamun Rashid2, Jack Satsangi7, David J. Adams2, and Mark J. Arends3

Abstract

Purpose: Inflammatory bowel disease–associated colorectal cancers (IBD-CRC) are associated with a higher mortality than sporadic colorectal cancers. The poorly defined molecular pathogenesis of IBD-CRC limits development of effective prevention, detection, and treatment strategies. We aimed to identify biomarkers using whole-exome sequencing of IBD-CRCs to guide individualized management.

Experimental Design: Whole-exome sequencing was performed on 34 formalin-fixed paraffin-embedded primary IBD-CRCs and 31 matched normal lymph nodes. Computational methods were used to identify somatic point mutations, small insertions and deletions, mutational signatures, and somatic copy number alterations. Mismatch repair status was examined.

Results: Hypermutation was observed in 27% of IBD-CRCs. All hypermutated cancers were from the proximal colon; all but one of the cancers with hypermutation had defective mismatch repair or somatic mutations in the proofreading domain of DNA POLE. Hypermutated IBD-CRCs had increased numbers of predicted neo-epitopes, which could be exploited using immunotherapy. We identified six distinct mutation signatures in IBD-CRCs, three of which corresponded to known mechanisms of mutagenesis. Driver genes were also identified.

Conclusions: IBD-CRCs should be evaluated for hypermutation and defective mismatch repair to identify patients with a higher neo-epitope load who may benefit from immunotherapies. Prospective trials are required to determine whether IHC to detect loss of MLH1 expression in dysplastic colonic tissue could identify patients at increased risk of developing IBD-CRC. We identified mutations in genes in IBD-CRCs with hypermutation that might be targeted therapeutically. These approaches would complement and individualize surveillance and treatment programs. Clin Cancer Res; 1–10. ©2018 AACR.

Introduction

Inflammatory bowel disease–associated colorectal cancer (IBD-CRC) is an aggressive complication of chronic inflammation accounting for 10% to 15% of deaths from IBD. Patients with IBD-CRC are younger and have a higher mortality than those with sporadic colorectal cancer (1). IBD-CRC arises diffusely in chronically inflamed epithelium via low-grade dysplasia evolving to high-grade dysplasia and eventually adenocarcinoma, although it may also arise without these preceding changes. Technological advances in endoscopy have improved the ability to discriminate between normal and pre-cancerous mucosa (1); however, a third of cancers arise in the interval between scheduled colonoscopies, suggesting poor efficacy of surveillance programs (2). Colonoscopy-directed mucosal samples are analyzed for the presence of dysplasia, which remains the gold standard for predicting future cancer risk in IBD when reported by an expert gastrointestinal pathologist.

The poorly defined molecular and genetic pathogenesis of IBD-CRC impairs our ability to detect and treat colorectal cancer effectively. There is an urgent need to identify biomarkers to improve early detection of colorectal cancer and guide individualized therapy (2).

We elected to study primary IBD-CRC because molecular changes identified in the resected cancer specimen are most likely to represent biomarkers of cancer development. We characterized the hypermutator status of the cancers, defined the mutational signatures and underlying biological processes, and explored the potential clinical translation of this work with respect to immune therapy, and improving current surveillance programs.
**Translational Relevance**

There is an urgent need to identify predictive biomarkers in inflammatory bowel disease-associated colorectal cancers (IBD-CRC) to individualize the current prevention, surveillance, and treatment programs. This analysis demonstrates that proximal IBD-CRCs have high mutational rates associated with defects in MMR (MLH1 loss) and DNA POLE proofreading function. Hypermutation is associated with a predicted higher neo-epitope load, which could be exploited using immunotherapies in selected patients with hypermutated IBD-CRCs. Prospective studies are required to determine whether analysis for loss of MLH1 in surveillance colonic biopsies could distinguish those at increased risk of developing colorectal cancer, permitting a more targeted approach for cancer treatment and surveillance. The identification of driver genes in hypermutated IBD-CRCs could also be used to develop therapeutic agents targeting the corresponding molecular pathways. Our comprehensive analysis of the mutational landscape of IBD-CRCs has revealed several novel approaches that may complement and personalize surveillance and treatment programs.

**Materials and Methods**

**Ethical approval and case identification**

The study was approved by the Lothian NHS Research Scotland Human Annotated BioResource, which is an NHS Health Research Authority Ethics Committee-approved Research Tissue Bank (www.hra.nhs.uk; 15/ES/0094) for the use of human tissue surplus to diagnostic requirements. The study was conducted in accordance with the Declaration of Helsinki and the guidance from the Human Tissue Authority on the use of human tissue from diagnostic archives. The Lothian (Edinburgh, UK) pathology database was searched for IBD-CRCs using the terms IBD, Crohn's disease, ulcerative colitis, colon cancer, colorectal and rectal cancer between 1990 and 2011 by the Tissue Bank appointed pathologist and provided the anonymized tissue samples with linked clinical and 2011 by the Tissue Bank appointed pathologist and provided the anonymized tissue samples with linked clinical and 2011 by the Tissue Bank appointed pathologist and provided the anonymized tissue samples with linked clinical and 2011 by the Tissue Bank appointed pathologist and provided the anonymized tissue samples with linked clinical medical records, the histopathological evidence for IBD in previous colonic samples and in the IBD-CRC resection specimen.

Cases were only included from patients with a previous histopathologic diagnosis of IBD and where evidence of IBD in the resected colonic specimen was confirmed by both GI specialist pathologists (CJB and MJA) independently. Sporadic adenomas and colorectal cancers which arose in an area of the colon without prior evidence of IBD in that colonic area were excluded. For each case, we have clear histopathological evidence of preexisting IBD at each tumor site and have confirmed that we have selected IBD-associated colorectal cancers.

Formalin-fixed paraffin embedded (FFPE) cancer and uninvolved normal lymph node blocks, removed at the time of surgery, were sectioned, H&E stained, and used for further analysis. In total 31 cases [15 Crohn’s disease (CD); 16 ulcerative colitis (UC)] with 34 cancers were used in this study.

**Clinical phenotype data: statistical analysis**

The "survival" package within the R software environment (3) was used to generate the Kaplan–Meier survival curves and perform statistical analyses. Categorical clinical phenotype data were analyzed using the two-tailed Fisher exact test; a P-value of <0.05 was considered significant. No survival data were available for case 2L (hypermutator), which was excluded from the survival analysis. Cases 15G1 and 15G2 were considered as one case for the purpose of statistical comparison, as they appear to originate from the same precursor clone.

**Sequencing data generation, processing, and analyses**

Nucleic acids were extracted using Qiagen Allprep FFPE DNA extraction kits, and DNA/RNA was quantified using Agilent GenChip. For whole genome sequencing (WES), DNA was captured using the Agilent SureSelectXT Human All Exon V5 platform following the manufacturer's protocol. Sequencing was performed using the Illumina HiSeq 2000 platform at the Wellcome Trust Sanger Institute. Raw paired-end sequencing reads (75bp) were aligned and a modified version of the reference genome which includes the GRCh37 primary assembly and additional human contigs and viral sequences that reduce the number of reads erroneously mapped to the primary assembly. Further details on data processing, data quality assessment, and sequence alignment are available in the Supplementary Materials and Methods. Somatic point mutations and small insertions and/or deletions (InDels) were identified by comparing cancer and matched normal samples using MutTect (v1.1.7; ref. 4) and Strelka (v1.0.14; ref. 5), respectively. BCFtools (v1.3; ref. 6) was used to identify variants in each patient’s germline (normal lymph nodes) relative to the human reference genome described above. The Ensembl Variant Effect Predictor (7) software was used to predict the effect of each variant on protein sequences, and the clinical significance of the changes was predicted by comparison to variants in the ClinVar database (release date 2016-05-31). For genes with multiple transcripts, we report the effect of the variant on the protein derived from the canonical transcript as annotated in Ensembl release 81 for GRCh37 (Supplementary Table S7; Supplementary Materials and Methods). We also provide the predicted effect on all transcripts in Supplementary Table S8. The Sequenza software package (version 2.1.2; ref. 8) was used to identify somatic copy number alterations. For the non-hypermutator cases, MutSigCV (version 1.4; ref. 9), which identifies significantly mutated genes, and dNdScv (10), which identifies genes under positive selection in cancer, were used to identify driver genes. For the hypermutator cases, microsatellite InDels were excluded from analysis with MutSigCV. The microsatellite InDels were analyzed with a modified version of MSMSigS (11) obtained from the author, along with required input files. In dNdScv, which is able to analyze both the SNV and InDel mutations together in both hypermutator and non-hypermutator cohorts, two InDel models are available; one model considers the total number of InDel mutations per gene, and the other model considers unique InDel sites per gene (the "unique-sites model"). For the hypermutator cohort, dNdScv was run using both models, and the results were compared. Only the “unique-sites” model was used for the non-hypermutator cohort. Restricted hypothesis testing was performed on the results from dNdScv, using known cancer genes from the Cancer Gene Census version 81 (12). Further details are available in the Supplementary Materials and Methods. Mutation signatures were identified with the
Bioconductor package SomaticSignatures (version 2.6.0; ref. 13) using the non-negative matrix factorization (NMF) algorithm. Details of software parameters, databases used, variant annotation, variant filtering, validation, and the identification of driver genes are available in the Supplementary Materials and Methods.

Human leukocyte antigen typing and neo-epitope predictions

Human leukocyte antigen (HLA)-I 4-digit typing was performed using the OptiType 1.0 algorithm and neo-epitopes from missense mutations were predicted by mapping the corresponding protein sequences to the human proteome database (version GRCh37.74). Neo-epitopes with a relative percentile rank ≤1% for each HLA-I allele were considered binders (additional details are provided in Supplementary Materials and Methods).

Mismatch repair IHC

Immunohistochemistry (IHC) for MLH1, MSH2, MSH6, and PMS2 was undertaken on a cancer tissue microarray and staining was independently scored by two pathologists (MJA and AO). Discordant scores were resolved by staining whole tissue sections with additional MLH1, MSH2, MSH6, and/or PMS2 antibodies.

MLH1 promoter methylation analysis

MLH1 promoter methylation was analyzed using the EZ DNA Methylation Kit Gold (Zymo Research). Cases from patient 32N failed methylation analyses. Further experimental details are described in the Supplementary Materials and Methods.

Results

To identify biomarkers associated with IBD-CRC, WES was performed on 34 IBD-CRCs and matched normal lymph node pairs. Twenty-nine patients had one cancer, one patient had three primary cancers separated anatomically (32N) and one patient had two cancers in close proximity to each other (15G, Supplementary Table S1). WES yielded between 45- and 90-fold coverage of the cancer samples and 26- and 93-fold coverage of the normal lymph node samples (Supplementary Fig. S1). When considering protein coding and splice site mutations only, the cancers from 32N had only six mutations common to at least two of the three cancers, whereas the two cancers from 15G had over 2,345 common mutations (22% of the total mutant positions in these cancers).

Somatic mutation rates in IBD-CRC

Somatic point mutations and InDels were identified by comparing exome sequences from cancer tissue with those from uninvolved lymph nodes removed at the time of surgery (Supplementary Tables S7 and S8). The 34 cancers were divided into two groups based on distinct somatic mutation rates. There were 24 non-hypermutator cancers with 2.0 to 7.0 mutations/Mb, and 10 hypermutator cancers with 32.6 to 171.3 mutations/Mb (Fig. 1A; Table 1). Two of the 10 cases, 15G1 and 15G2, have a somatic mutation burden greater than 100 mutations/Mb and could thus be defined as ultra-hypermutators.

Figure 1. Somatic mutational rates and survival analysis of IBD-CRC. A, Mutational frequency in each of the 34 IBD-CRCs ordered by overall mutation rate. There is a clear separation between the 10 hypermutator cancers and the 24 non-hypermutator cancers. With the exception of 15G1, 15G2, and 6J, the hypermutator cancers showed elevated mutation rates of both SNVs and InDels. No InDels were found in the exome of 21M. B, Kaplan–Meier plot of overall survival stratified by cancer mutator phenotype. Patients with hypermutator cancers had increased survival compared with patients with non-hypermutator cancers (log-rank test, \( P = 0.04 \)).

Table 1. Mutational rates in IBD-CRC and sporadic colorectal cancer

<table>
<thead>
<tr>
<th>Tumor phenotype</th>
<th>TCGA (25)</th>
<th>Din</th>
<th>Robles (31)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypermutators</td>
<td>&gt;12 (728)</td>
<td>32.6–171.3 (56.2)</td>
<td>18.19–22.2 (N/A)</td>
</tr>
<tr>
<td>Non-hypermutators</td>
<td>&lt;8.24 (58)</td>
<td>2.1–7.0 (3.3)</td>
<td>0.5–5.1 (5.1)</td>
</tr>
</tbody>
</table>

NOTE: Mutational rates (mutations/Mb) from the whole-exome sequence analysis of TCGA sporadic colorectal cancer cohort (25) and IBD-colorectal cancers from this study and Robles and colleagues (31) in the hypermutator and non-hypermutator cancers. Numbers in parentheses are median mutation rates. Two hypermutators were present in the Robles cohort. The rates provided include total mutations with the exception of the median mutation rate of non-hypermutator cancers in the TCGA cohort.
15G2, and 6J, which have similar InDel mutation rates to the non-hypermutator cases (Fig. 1A). The median age at diagnosis was not significantly different between the hypermutators (median 66.2 years; IQR 57.4–78.1 years) and non-hypermutators (median 65.8 years; IQR 51.0–77.0 years).

Overall, there was a significant survival difference between the patients with hypermutator cancers and those with non-hypermutator cancers (log-rank test, \( P = 0.04 \)), with the estimated 10-year survival being 75% in the former group as compared with 36% in the latter (Fig. 1B). However, it is likely that additional factors other than mutator status including stage and age influenced patient survival although we were unable to measure the effects of these variables due to the small sample size.

Strikingly, the hypermutator cancers were all located in the proximal colon and none in the distal colon (two-tailed Fisher exact test, \( P = 0.004 \); Fig. 2).

Mismatch repair abnormalities

To characterize the difference in mutational rates, the cancers were analyzed for genetic aberrations associated with dMMR. Seven out of the 10 hypermutator cancers had a high frequency of InDels (Fig. 1A), which is indicative of MSI, and showed loss of expression of MLH1 and its heterodimeric binding partner PMS2 (Fig. 2; Supplementary Fig. S2). Loss of MLH1 protein expression results in dMMR, leading to increased somatic expression of MLH1 and its heterodimeric binding partner PMS2 (Fig. 2; Supplementary Fig. S2). Loss of MLH1 protein expression in this case. The normal expression of MLH1 would cause loss of function. We did not detect any point or InDel mutations in the MLH1 gene, resulting in a truncated protein with a predicted nonsense mutation within the ATPase domain (R100*) of the MLH1 gene, resulting in a truncated protein with a predicted loss of function. We did not find other coding mutations that would cause MLH1 to be biallelically inactivated in case 33W. We did not detect any point or InDel mutations in MLH1 in the germline or tumor in case 28E, nor epigenetic silencing by promoter hypermethylation that could explain the loss of MLH1 expression in this case. The normal expression of MLH1 and low level of somatic InDels in the remaining three hypermutator cancers, 15G1, 15G2, and 6J, point to alternative mechanisms leading to hypermutation. In contrast, loss of expression, promoter hypermethylation and nonsilent somatic mutations were not observed in MLH1 in the non-hypermutator cases (Fig. 2; Supplementary Table S2B).

Mutations in DNA polymerase proofreading domains

During DNA replication, DNA fidelity is maintained by the proofreading function of DNA polymerases. Germline mutations in the exonuclease proofreading domains of the DNA polymerases POLQ (codons 245–571) and POLE (codons 223–517) have been shown to predispose to the development of hypermutated microsatellite stable (MSS) sporadic colorectal cancer (16). Nonsilent somatic mutations in POLE and POLD1 were identified in four and seven of the hypermutator cancers, respectively (Fig. 2; Supplementary Table S2A). Two shared somatic POLE mutations (P286R and F348S) affected the exonuclease proofreading domain in cases 15G1 and 15G2, which were two adjacent cancers (adenocarcinoma and squamous cell carcinoma) in the same patient (Fig. 1A and 2).
No frameshift or nonsilent somatic POLE or POLQ point mutations were identified in the non-hypermutator cancers. The mechanism of hypermutation in case 6J remains undefined. Case 6J was an adenocarcinoma from a patient with a 20-year history of colonic Crohn's disease and autoimmune liver disease for which she received potent immunosuppression (prednisolone, tacrolimus, and azathioprine) throughout this period. This patient had four basal and two squamous cell carcinomas of skin removed but did not have any frameshift or nonsilent somatic or germline point mutations in PTTCH1, excluding Gorlin syndrome as a contributing factor to the underlying cancer predisposition (17). In addition, germline mutations that could explain the hypermutator phenotype were not found in other DNA repair genes including those involved in nucleotide and base excision. It is possible that the continuous, prolonged immunosuppression in this patient has altered the immune cancer response resulting in both colonic and multiple skin cancers.

Mutational signatures in IBD-CRC

It has been demonstrated that different mutational processes in cancers generate specific patterns of mutation, or “signatures,” with 30 distinct signatures identified thus far by Alexandrov and colleagues (18). The overall mutational spectrum our IBD-CRC cohort was very similar to the spectra derived from cohorts of sporadic colorectal cancers from Giannakis and colleagues (cosine similarity = 0.87; ref. 14) and from TCGA Data Portal (cosine similarity = 0.91; Fig. 3A; Supplementary Materials and Methods). Six distinct signatures, designated A–F, were extracted from the catalogue of IBD-CRC somatic mutations (Fig. 3B) and corresponded well to Alexandrov signatures 10, 1, 13/2, 17, 6, and 5, respectively, with cosine similarities ranging from 0.82 to 0.97 (Fig. 2, bottom; Supplementary Table S3). Some of the signatures have been associated with specific mutational mechanisms (18). Of particular note, are IBD-CRC Signature A (Alexandrov signature 10/POLE) that was predominant in the hypermutator cancers with DNA POLE mutations (15G1 and 15G2); and IBD-CRC Signature E (Alexandrov signature 6/dMMR and MSI) that was predominant in the seven hypermutator cancers which had loss of expression of the MLH1 protein (Fig. 2). IBD-CRC Signature C (Alexandrov signature 13 and 2/AID/APOBEC) is the major contributor in case 21M, however as our analysis is limited to the exome regions only, we have no evidence of AID/APOBEC activation in this case. Transcriptional upregulation of APOBEC3B is commonly found in bladder, cervical, lung, head/neck, and breast cancers with kataegis (18), however, we did not find evidence of kataegis in case 21M (Supplementary Fig. S4). It has been hypothesized that one cause of APOBEC3B upregulation may be infection by viruses, including human papillomavirus (HPV; ref. 19). We did not find HPV DNA associated with case 21M, however, because whole-exome sequencing was used in this study, any viral sequence present would be excluded unless it has integrated into the genome in the targeted sequences.

Although Alexandrov signatures 2 and 13 have not yet been identified in colorectal cancer, an APOBEC mutation pattern has been found, using different methods, in a variety of cancer types including whole-genome sequenced sporadic colorectal cancers (20). Signatures 1, 5, 6, and 10 have previously been found in sporadic colorectal cancers (18) and signature 17 has been identified in MSS colorectal cancer (21). Taken together, these results indicate that the mutational mechanisms in IBD-CRC and sporadic colorectal cancer are similar, although a much larger IBD-CRC cohort and whole-genome sequencing may reveal additional or novel signatures.

Recently, a signature attributed to the persistence of 8-oxoguanine G>T/C>A mismatches due to biallelic inactivation
of MUTYH has been identified in MUTYH-associated polyposis (MAP) colorectal cancers (22). Unsurprisingly, we did not find this signature in our IBD-CRCs as our cases did not have a history of MAP nor any germline nonsilent or splice site point or InDel mutations in MUTYH other than common SNPs. We did not discover a novel inflammation-associated mutational signature. Similarly, others have not found an inflammation-signature common to chronic inflammatory-associated gastrointestinal cancers (such as Barrett’s esophagus; ref. 23). Rather, epithelial regeneration and the subsequent cumulative effect of chronic inflammation-associated damage appears to be a major mechanism of promoting carcinogenesis in IBD (24).

Driver genes in IBD-CRC

To identify driver genes in non-hypermutator IBD-CRC, we used MutSigCV (9) to identify significantly mutated genes, and dNdScv to find genes under positive selection in cancer (Methods and Supplementary Materials and Methods; ref. 10). In our 24 cases, TP53, PIK3CA, APC, and KRAS were identified as driver genes (FDR-adjusted $P$-value, or $q < 0.10$; Fig. 4A), all of which are
as also driver genes in sporadic colorectal cancer (25). As observed in sporadic colorectal cancer (25), significantly more non-hypermutator cancers had nonsilent somatic mutations in TP53 than the hypermutator cancers (79% vs. 33%; two-tailed Fisher exact test, \( P = 0.03 \); Table 2).

In the hypermutator cases, no driver genes were identified using MutSigCV (9) or MSMutSig (11), due to small sample size. When using dNdScv (10), followed by restricted hypothesis testing with a list of known cancer genes from the Cancer Gene Census (version 81; ref. 12), KRAS (\( q = 0.03 \)) was identified as a driver gene in the hypermutator cohort, in which three of the four cases with KRAS mutations are p.G12D (Fig. 4A). Importantly, this analysis also identified hotspot InDel mutations in RPL22 (p.K15fs; \( q = 0.001 \)), ACVR2A (p.K437fs; \( q = 0.03 \)), and RNF43 (p.G659fs; \( q = 0.1 \); Fig. 4A; Supplementary Materials and Methods). The same ACVR2A and RNF43 mutations have previously been identified as driver events in a much larger cohort of sporadic colorectal cancer (11), and although they may represent driver events in IBD-CRC, they could also be genomic sites that are particularly prone to mutation. A much larger cohort of MSI IBC-CRCs would be required to confirm these as driver events. As reported for sporadic colorectal cancers (26), we find enrichment of RNF43 mutations in MSI cases, and mutual exclusivity of RNF43 truncating mutations (which includes frameshift InDels, nonsense mutations, and splice site mutations) with APC truncating mutations in our IBD-CRC cohort (Fig. 4A).

Several of the genes we identified are clinically actionable. For example, tumors with RNF43 mutations are sensitive to porcine inhibitors (27). In the same way, RPL22 mutant cancers may be susceptible to MDM2-p53 pathway compounds (28). Notably, the same RPL22 p.K15fs hotspot mutation we identified in our IBD-CRC cohort has been frequently found in endometrial, gastric, and colorectal cancers (28).

**Somatic IDH1 R132 mutations**

A potentially targetable IDH1 hotspot mutation at R132, which was mutated in 11% (5/47) of IBD-CRCs and only in CD-associated colorectal cancers, was previously reported by Yaeger and colleagues (29). In our cohort, one IDH1 R132C mutation was present in UC-associated colorectal cancer case 32N3. The observed frequency here is not significantly different than that of Yaeger and colleagues (\( \chi^2 \) test, \( P = 0.21 \); ref. 29). We note, however, that the median 17-fold coverage (range 9- to 36-fold in the tumor samples) at this site was lower than the whole-exome median of 70-fold, as this mutation is only 20bp from an exon boundary where coverage tends to be lower in whole-exome sequencing studies.

**Somatic copy number alterations in IBD-CRC**

As previously observed with sporadic colorectal cancer (30), somatic copy number alterations (SCNA) in IBD-CRCs were significantly more prevalent (one-tailed Student \( t \) test \( P = 0.001 \)) in the non-hypermutator cases compared with hypermutator cases (Supplementary Fig. S5; Supplementary Table S6).

Robles and colleagues (31) reported 15 focal copy number gains, 8 of which were observed in more than one tumor and Shivakumar and colleagues (32) reported 26 focal SCNAs in pooled IBD-associated dysplastic and carcinoma biopsy samples. Focal SCNAs common to our IBD-CRC cohort and these studies are gains in 12p13.33-12p13.31, 22q11.21, 4p16.3, 8q24.3, 10q26, 13q12, 20q11.23, and 20q13; however, none were common to all three.

We compared the frequency of predicted chromosome arm-level SCNAs to recurrent SCNAs found by Sheller and colleagues (33), and to SCNAs identified by TCGA (Fig. 4B; Supplementary Table S6; ref. 25). Our novel data demonstrate that the frequencies of specific arm-level SCNAs in IBD-CRC broadly concurs with those found in sporadic colorectal cancer.

**Germline variants in IBD patients**

Colorectal cancer affects up to 2.5% of the patients who suffer from IBD (1), and germline variations conferring cancer susceptibility have been poorly described in this population of patients. We examined the germlines of our cohort for variants in genes implicated in susceptibility to colorectal cancer (Supplementary Tables S4 and S5). The CHEK2 frameshift mutation at T367 (c.1100delC; rs555607708), in the germline of case 15N, is known to predispose to breast cancer, but association with colon cancer has not been resolved (34). The R242H mutation in SDHB (rs74315368) in case 3Q has been characterized in paragangliomas (35) but not, to our knowledge, in colon cancer. The remaining germline variants we have reported in Supplementary Table S5 are in genes known to be associated with increased risk of colorectal cancer, however, they have unknown or mixed reported clinical significance and currently we are unable to clarify as to whether these variants confer a predisposition to colorectal cancer in the setting of chronic inflammation.

**Neo-epitope and immune infiltrate analysis**

Nonsynonymous somatic mutations in cancer can generate novel antigens (neo-epitopes) that can be exploited in cancer immunotherapy. The number of predicted neo-epitopes is expected to be directly proportional to the mutational load of cancers (36). As expected, the IBD-CRC hypermutator cancers generated the largest number of HLA class I neo-epitopes, and were significantly higher than the number found in non-hypermutators (one-tailed Student \( t \) test, \( P = 0.004 \); Fig. 5). Similar to our study, the neo-epitope load was higher in sporadic colorectal cancer with dMMR/MSI-high status, than in pMMR/MSS tumors, and notably, they responded favorably to PD1 blockade (14). IBD is characterized by a dysregulated immune response and many therapeutic modalities targeting inflammation are directed at

Table 2. Frequency of TP53, KRAS, and APC mutations in IBD-CRC and sporadic colorectal cancer

<table>
<thead>
<tr>
<th>Gene</th>
<th>TCGA HM (25)</th>
<th>TCGA NHM (25)</th>
<th>Din HM</th>
<th>Din NHM</th>
<th>Robles NHM (31)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP53</td>
<td>20%*</td>
<td>60%*</td>
<td>33%</td>
<td>79%*</td>
<td>13%</td>
</tr>
<tr>
<td>KRAS</td>
<td>30%</td>
<td>43%</td>
<td>44%</td>
<td>25%</td>
<td>20%</td>
</tr>
<tr>
<td>APC</td>
<td>51%*</td>
<td>81%*</td>
<td>33%</td>
<td>29%</td>
<td>13%</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>13%</td>
<td>23%</td>
<td>56%</td>
<td>29%</td>
<td>10%</td>
</tr>
</tbody>
</table>

*Significant differences between HM and NHM colorectal cancers.

**NOTE:** Mutational frequencies from the whole-exome sequence analysis of TCGA sporadic colorectal cancer cohort (25) and IBD-CRCs from this study and Robles and colleagues (31) in hypermutator (HM) and non-hypermutator (NHM) cancers.

Published OnlineFirst June 27, 2018; DOI: 10.1158/1078-0432.CCR-17-3713

**Predictive Biomarkers in IBD-Associated Colorectal Cancers**
cytokines. Analysis of the cytokine gene immune expression profile demonstrated that overall the cytokine gene expression profile was similar in the hypermutator IBD-CRCs and the non-hypermutator IBD-CRCs (Supplementary Fig. S6), although this analysis was limited by the very small size of tumor epithelium (with very little immune cell-containing stroma) sampled.

**Discussion**

We have undertaken a comprehensive analysis of 34 IBD-CRCs from 31 patients using WES. Hypermutator cancers were observed in both our study and Robles and colleagues (31) and the mutational rates were within range of the larger TCGA sporadic colorectal cancer cohort (25).

In sporadic and hereditary colorectal cancers, the hypermutator phenotype is most frequent in cancers of the right (proximal) colon (25). In this study, we observed a strong association between CD-associated colorectal cancers occurring in the proximal colon (two-tailed Fisher exact test, \( P = 0.03 \)) and the hypermutator phenotype (two-tailed Fisher exact test, \( P = 0.05 \)). Long standing, extensive colonic CD has been shown to be associated with proximal cancers (37) and we have clear histopathologic evidence of preexisting IBD at each tumor site, more than half of which were in the proximal colon. Similar to previous studies of IBD-CRC (38), a higher proportion of UC-associated colorectal cancers occurred in the left colon (Fig. 2) when compared with CD-associated colorectal cancers (two-tailed Fisher exact test, \( P = 0.03 \)).

Previous studies have reported a variable frequency of \( MLH1 \) promoter hypermethylation and loss of expression of \( MLH1 \) in IBD-CRCs (39). The frequency of \( MLH1 \) promoter methylation in our series may be associated with the more advanced age at time of IBD-CRC diagnosis, although it remains lower than that of sporadic MSI-high colorectal cancer with CpG (\( MLH1 \)) hypermethylation (40). Lennerz and colleagues (39) have exclusively studied colorectal cancer complicating Crohn’s colitis and described the median age of cancer diagnosis to be 58 years (range 34–77), which is also slightly higher than that conventionally reported for IBD-CRCs. In addition to our strict inclusion criteria, and similar to previous studies (39), the mutation analysis identified a single \( BRAF \) V600E mutation in one hypermutator IBD-CRC (Fig. 2) confirming that these cancers are unlikely to be sporadic in nature.

\( BRAF \) V600E mutations occur in the majority (>85%) of hypermethylated sporadic MSI high colorectal cancers (25) and are therefore established in diagnostic algorithms to differentiate between sporadic and familial Lynch syndrome cases of colorectal cancer (41). Notably, we did not identify germline mutations in genes such as \( MLH1 \), \( MSH2 \), \( APC \), and \( MUTYH \) that predispose to the development of colorectal cancer in our IBD-CRC cohort. Although the proportion of hypermutator cancers in our cohort was not significantly different than that in sporadic colorectal cancer, these differential mutations distinguish our hypermutated IBD-CRC cohort from hypermutated sporadic and familial forms of colorectal cancer.

The effect of dMMR/MSI on survival outcomes in IBD-CRC has not been reported. A recent meta-analysis of 20 studies including 571,278 patients by Reynolds and colleagues (42) has reported that IBD-CRC does not affect the overall 5-year survival compared with sporadic colorectal cancer without any adjustment for molecular subtypes. In our series, the hypermutator IBD-CRCs (which included two MSS hypermutators 15G and 6J) had a significantly better survival compared with non-hypermutator IBD-CRCs. The increased survival of sporadic MSI cancers (43) is comparable to the data presented here for IBD-CRCs. The improved survival of early-stage sporadic dMMR/MSI cancers is postulated to be mediated by (anti-)tumor infiltrating lymphocytes (TIL) in response to the neo-epitopes generated by the high mutational rate. Immune blockade therapies targeting immune checkpoints and enhancing the anti-TIL response are currently being used in various cancers including colorectal cancer (44). In our series, the hypermutator IBD-CRCs had a higher predicted neo-epitope load but we were unable to identify any obvious differences in the immune-related gene expression profile. This is unsurprising as our samples had been enriched for cancer-containing cells and not the adjacent immune rich stromal compartment. Molecular phenotyping of all colorectal cancers for targeted therapy is a compelling reason for universal screening for dMMR. Of note, the FDA has approved Pembrolizumab (targeting the programmed cell death 1 receptor) for the treatment of all unresectable or metastatic MSI-H/dMMR tumors that have progressed after initial treatment (45). Immune checkpoint blockade has been associated with immune-mediated colitis and this is
particularly relevant for patients with coexisting IBD (46). Although our results require validation in a randomized prospective cohort, it is exciting to postulate that immune blockade therapies may be useful adjuncts in treating patients with dMMR (and/or hypermutator phenotype) associated IBD-CRC that have undergone a total colectomy as part of their treatment programme, negating the risk of immunotherapy mediated colitis. Importantly, immune therapies should be used with caution or avoided in IBD-CRC patients who have undergone a partial colectomy or those with stage IV disease with an intact colon as it can aggravate the underlying colitis (47).

To date, we have not been able to use genetic or molecular markers to improve the detection or treatment of IBD-CRC. In 2009, a specialist committee in the United States of America recommended universal screening for Lynch syndrome in all newly diagnosed colorectal cancers (48) and this has not been adopted for several reasons including financial, technical, ethical and health economical limitations. In the United Kingdom, the National Institute for Health and Care Excellence (NICE) is advocating use of mismatch repair IHC on all colorectal cancers (including IBD cases) to detect Lynch syndrome (41). A cost effective analysis limited to early onset colorectal cancers suggests that this would be economically viable assuming that all of the necessary subsequent health interventions are fully implemented to reduce cancer mortality and morbidity (49). Our data supports this recommendation to detect dMMR in IBD-CRC, but not necessarily to detect Lynch syndrome as we did not identify any known clinically significant pathogenic germline mutations in MMR genes in this IBD-CRC cohort. Although these important issues are resolved our data can be used to support the rational for universal testing in IBD-CRCs to profile tumors for the use of targeted therapies, which is not routinely undertaken at all institutions.

The IHC analysis of TP53 protein has not been widely accepted to aid in the discrimination between dysplastic and inflamed colonic epithelium. Prospective studies are required to determine whether analyzing for TP53 abnormalities and loss of expression of MLH1 together, in colonic biopsies with potentially dysplastic epithelium, could aid in the evaluation of those at highest risk of developing colorectal cancer, similar to the molecular profiling used in gastric cancer (50). Our retrospective study has inherent limitations and the small sample size may result in a reporting bias. Notwithstanding these issues, the power of next-generation sequencing technology has provided detailed information allowing clinically relevant statistical analyses to be undertaken. Future similar studies may discover many more similarities and differences between IBD-CRC and other types of colorectal cancer.

The categorization of disease guides individualized therapeutic strategies and can predict the response to therapy and prognosis. Our analysis demonstrates that proximal IBD-CRCs have high mutational rates associated with defects in MMR and DNA POLE proofreading function. This results in a predicted higher neo-epitope load, which could be exploited using immunotherapies. The hypermutator phenotype of this cohort is mostly associated with loss of MLH1 expression and this could be evaluated in colonic dysplastic lesions detected in IBD patients. The identification of driver genes in hypermutated IBD-CRCs can be used to develop therapeutic agents targeting the corresponding molecular pathways. In our series up to 90% of the hypermutated IBD-CRCs have actionable mutations. These approaches would complement and individualize current surveillance and treatment programs for IBD-CRC.

All exome sequencing data are available from the European Genome-phenome Archive under accession number EGAS00001001129.

Acknowledgments

CpG methylation primer sequences were generously provided by R.H. Masalmeh and D. Sproul (University of Edinburgh). This work was supported by the European Research Council Combat Cancer Program, Cancer Research UK, and the Wellcome Trust. S. Din acknowledges the financial support of NHS Research Scotland (NRS), through NHS Lothian. A. Jiménez-Sánchez was supported by a doctoral fellowship from the Cancer Research UK Cambridge Institute and the Mexican National Council of Science and Technology (CONACyT).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received December 13, 2017; revised May 3, 2018; accepted June 25, 2018, published first July 16, 2018.

References

27. Kallunki M, Dvorak M, Tavani C, Di对该文进行映射
Clinical Cancer Research

Mutational Analysis Identifies Therapeutic Biomarkers in Inflammatory Bowel Disease–Associated Colorectal Cancers

Shahida Din, Kim Wong, Mike F. Mueller, et al.

Clin Cancer Res  Published OnlineFirst June 27, 2018.

Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-17-3713

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2018/06/27/1078-0432.CCR-17-3713.DC1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link
http://clincancerres.aacrjournals.org/content/early/2018/08/18/1078-0432.CCR-17-3713.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.