A Specific Mutational Signature Associated with DNA 8-Oxoguanine Persistence in MUTYH-defective Colorectal Cancer

Alessandra Viela a,1, Alessandro Bruselles b,1, Ettore Meccia c,1, Mara Fornasarig d, Michele Quaia a, Vincenzo Canzonieri e, Eleonora Policicchio f, Emanuele Damiano Ursog M, Marco Agostini g,h, Maurizio Genuardi i, Emanuela Lucci-Cordisco i, Tiziana Venesio k, Aline Martayan l, Maria Grazia Diodoro l, Lupe Sanchez-Mete l, Vittoria Stigliano l, Filomena Mazzei c, Francesca Grasso c, Alessandro Giuliani c, Marta Baiocchi b, Roberta Maestro a, Giuseppe Giannini m, Marco Tartaglia f,2, Ludmil B. Alexandrov n,o,p, Margherita Bignami c,2,*

a Functional Onco-Genomics and Genetics, CRO Aviano National Cancer Institute, Aviano, Italy
b Department of Haematology, Oncology and Molecular Medicine, Istituto Superiore di Sanità, Rome, Italy
c Department of Environment and Primary Prevention, Istituto Superiore di Sanità, Rome, Italy
d Unit of Gastroenterology, CRO Aviano National Cancer Institute, Aviano, Italy
e Unit of Pathology, CRO Aviano National Cancer Institute, Aviano, Italy
f Genetics and Rare Diseases Research Division, Ospedale Pediatrico Bambino Gesù, Rome, Italy

First Surgical Clinic, Department of Surgical, Oncological and Gastroenterological Sciences, University of Padua, Padua, Italy

I Nano Imager Biomedicine Lab, Istituto di Ricerca Pediatrica (IRP) Padua, Italy
Institute of Genomic Medicine, Catholic University, Rome, Italy

Fondazione Polyclinico Universitario “A. Gemelli, Rome, Italy
Unit of Pathology, Candiolo Cancer Institute-FPO, IRCCS, Candiolo, Turin, Italy
Division of Gastroenterology and Digestive Endoscopy, Regina Elena National Cancer Institute, Rome, Italy

Department of Molecular Medicine, University La Sapienza, Rome, Italy

Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM, USA

Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM, USA

University of New Mexico Comprehensive Cancer Center, Albuquerque, NM 87102, USA

ARTICLE INFO

Article history:
Received 17 February 2017
Received in revised form 7 April 2017
Accepted 11 April 2017
Available online 13 April 2017

Keywords:
MUTYH-associated polyposis
Colorectal cancer
8-Oxoguanine
Base excision repair
Exome sequencing
Mutational signature

ABSTRACT

8-Oxoguanine, a common mutagenic DNA lesion, generates G:C→T:A transversions via mispairing with adenine during DNA replication. When operating normally, the MUTYH DNA glycosylase prevents 8-oxoguanine-related mutagenesis by excising the incorporated adenine. Biallelic MUTYH mutations impair this enzymatic function and are associated with colorectal cancer (CRC) in MUTYH-Associated Polyposis (MAP) syndrome. Here, we perform whole-exome sequencing that reveals a modest mutator phenotype in MAP CRCs compared to sporadic CRC stem cell lines or bulk tumours. The excess G:C→T:A transversion mutations in MAP CRCs exhibit a novel mutational signature, termed Signature 36, with a strong sequence dependence. The MUTYH mutational signature reflecting persistent 8-oxoG:A mismatches occurs frequently in the APC, KRAS, PIK3CA, FAT4, TP53, FAT1, AMER1, KDM6A, SMAD4 and SMAD2 genes that are associated with CRC. The occurrence of Signature 36 in other types of human cancer indicates that DNA 8-oxoguanine-related mutations might contribute to the development of cancer in other organs.

© 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

DNA repair provides a major protection against cancer, and germline DNA repair defects in inherited syndromes confer a significant cancer predisposition. DNA replication errors are a major source of mutations, and early-onset colorectal and other cancers are associated with germline mutations in DNA mismatch repair (MMR) genes (Lynch et al., 2015) or in the proofreading exonuclease domains of DNA
polymerases POLE and POLD1 (Palles et al., 2013). In a distinct mutational mechanism, endogenously or exogenously generated reactive oxygen species (ROS) induce pre-mutagenic DNA lesions. Most ROS-induced base damage is repaired by base excision repair (BER) initiated by DNA glycosylases. 8-Oxoguanine (8-oxoG), one of the most common oxidative DNA lesions can mispair with adenine during DNA replication to generate G:C–T:A transversion mutations. 8-oxoG-induced mutagenesis is prevented by the cooperative action of the DNA glycosylases encoded by the OGG1 and MUTYH genes. OGG1 removes 8-oxoG from 8-oxoG:C pairs and MUTYH scans the newly-synthesized daughter strand to locate and remove incorporated adenine mispaired with 8-oxoG (Mazzei et al., 2013; Markkanen et al., 2013; Tsuzuki et al., 2007). A hydrolase encoded by the MTH1 gene provides a third level of protection against oxidative mutagenesis by degrading 8-oxodGTP to prevent the incorporation of 8-oxodCMP into DNA (Mo et al., 1992). Inactivation of any of these genes increases steady-state DNA 8-oxoG levels and confers a mutator phenotype (Klungland et al., 1999; Tsuzuki et al., 2001; Hirano et al., 2003).

To date, no human disease has been associated with defective OGG1 or MTH1 activities. In contrast, germline biallelic MUTYH mutations underlie MUTYH-associated polyposis (MAP), a recessively heritable colorectal polyposis with a predisposition to colorectal cancer (CRC) (Al-Tassan et al., 2002; Sieber et al., 2003). CRCs in MAP patients bear distinctive somatic G:C–T:A transversions in the APC gene (Al-Tassan et al., 2002; Sieber et al., 2003; Jones et al., 2002). The relationship between defective repair of oxidized DNA and CRC susceptibility was strengthened by the recent report that germline mutations in NTHL1, encoding a DNA glycosylase involved in the BER of oxidized pyrimidines are associated with a polyposis clinically similar to MAP (Weren et al., 2015). Thus, two repair pathways counteract CRC susceptibility by acting at replication to process mismatches containing oxidized DNA bases.

Mutational signatures in cancer genomes provide indications of the mechanisms underlying neoplastic transformation (Alexandrov, 2015; Alexandrov et al., 2013a,b, 2015; Helleday et al., 2014). Thus, CRC from MAP patients offer the unique opportunity to identify a mutational fingerprint of persistent 8-oxoG:A mismatches. Here, we report that whole-exome DNA sequencing identifies a distinct mutational signature of G:C–T:A transversions in MAP CRC. The mutational signature is reflected in the specific pattern of oncocenes/tumour suppressors involved in colorectal carcinogenesis and associated with inactive MUTYH. Our findings also indicate the possible involvement of DNA oxidation-related mutations in other types of human cancer as the MUTYH mutational fingerprint is present in cancers of other organs.

2. Material and Methods

2.1. Subjects and Tissue Samples

All subjects included in this study gave informed consent for gene testing and related research (Table S1). The local scientific IRB approved the collection and usage of tumour samples for this research (CRO Aviano IRB-03-2016). Blood, normal intestinal mucosa, adenomas and CRCs were retrospectively obtained from MAP patients with confirmed constitutional biallelic mutations in the MUTYH gene. CRCs were collected at surgery, while adenomas and normal mucosa were harvested at colectomy or endoscopic polypectomy. In all cases histopathological evaluation was performed by a pathologist.

DNA from frozen tissues was extracted using the Qiagen DNeasy Tissue kit (Qiagen, Hilden, Germany) with a protocol including RNAse treatment. FFPE tissues were deparaffinized by two washes in Noyxil (15 min at 56 °C) and DNA was then obtained using the FFPE Qiagen Tissue kit (Qiagen). DNA quality and concentrations were assessed using Nanodrop spectrophotometer (ThermoFisher Scientific, MA, USA) and Qubit Fluorometer (Invitrogen, CA, USA).

2.2. Whole Exome Sequencing, Data Analysis and Validation

Whole Exome Sequencing was performed by the Beijing Genomics Institute (BGI, Hong Kong). Target enrichment was performed using in-solution technology (NimbleGen SeqCap EZ Library v.3.0, Roche) and the resulting target libraries were sequenced by Illumina sequencing technology (HiSeq2000) (Illumina, San Diego, CA, USA). Raw image files were processed by Illumina base calling software (CASAVA 1.7) using default parameters. Paired-end reads were aligned to the human genome (UCSC GRCh37/hg19) with the Burrows-Wheeler aligner (BWA v. 0.7.10). Presumed PCR duplicates were removed using Picard’s MarkDuplicates. The Genome Analysis Toolkit (GATK 3.2) was used for realignment of sequences encompassing indels and for base quality recalibration. Somatic single-nucleotide variants were detected using muTect software v.1.1.6, and small indels were identified through a comparison between indels called in individual CRCs and their matched non-tumoural samples by means of the GATK Haplotype Caller algorithm, applying the following quality filters: quality score > 100 and quality-by-depth score > 1.5; indels below these thresholds or resulting from four or more reads having ambiguous mapping (this number being >10% of all aligned reads) were discarded. The resulting SNVs and small indels were annotated by SnpEff v3.6 and dbNSFP2.8 in terms of functional impact of variants (missense or nonsense, coding or non-coding, location with respect to exon-intron junction, depth, reference/variant reads ratio, dbSNP ID, amino acid change and position, cancer association data and structural/functional impact). Variant validation and genotyping were performed by Sanger sequencing. Amplicons were directly sequenced using the ABI BigDye Terminator Sequencing Kit (Applied Biosystems, Foster City, CA, USA) and an automated capillary sequencer (ABI 3500, Applied Biosystems). Sequence electropherograms were analyzed using Sequencing Analysis Software v.5.4 (Applied Biosystems).

2.3. Microsatellite Instability Analysis

Five mononucleotide and two pentanucleotide markers were co-amplified using the MSI analysis system (Promega Madison, WI, USA) according to the manufacturer’s protocol. The fluorescent-labeled PCR products obtained from constitutional DNA and tumour DNA were separated by capillary electrophoresis using an ABI3130 xl sequencer and evaluated with the GeneMapper software (Applied Biosystems/Life technologies). The 5 quasi-monomorphic mononucleotide markers included in the kit have high sensitivity and specificity in detecting alterations in tumour samples with mismatch repair defects. Criteria for definition of MSS and MSI are according to the Bethesda guidelines (Umar et al., 2004).

2.4. Mutational Signature Analysis

Prior to performing mutational signatures analysis, the mutational catalogues of all examined samples were generated. These catalogues include the six types of substitutions (C–T, C–G, C–T, T–A, T–C, T–G; mutations are referred to by the pyrimidine of the mutated Watson–Crick base pair) as well as the base immediately 5′ and the base immediately 3′ of each somatic mutation. The immediate 5′ and 3′ sequence context for each somatic mutation was extracted using the ENSEMBL Core APIs for human genome build GRCh37.

Mutational signatures analysis was performed based on somatic substitutions and their immediate sequence context. Briefly, mutational signatures were deciphered for all cancer types together using our previously developed computational MATLAB framework (Alexandrov et al., 2013a).

The computational framework for deciphering mutational signatures is freely available and it can be downloaded from: http://www.mathworks.com/matlabcentral/fileexchange/38724. The algorithm deciphers the minimal set of mutational signatures that optimally explains
the proportion of each mutation type found in each catalogue and then estimates the contribution of each signature to each mutational catalogue. Overall, we identified multiple distinct mutational signatures, most of which were previously reported by others or us (Alexandrov et al., 2013a,b; Schulze et al., 2015). The names of mutational signatures used in this manuscript are consistent with the ones on the COSMIC website, http://cancer.sanger.ac.uk/cosmic/signatures.

2.5. Target Re-sequencing

A TruSeq Custom Amplicon (TSCA) workflow was applied for interrogating targeted regions of interest in a series of MAP tumours. The custom panel included all exons and ± 10 padding intronic regions of the following target genes: AMER1, APC, FAT1, FAT4, KDM6A, KRAS, PIK3CA, SMAD2, SMAD4, TPS3. A project of 632 amplicons having a median size of 175 bp was constructed and optimized using Design Studio tool (Illumina). These amplicons covered a cumulative target of 60,174 bp and, according to the design, there were only 13 gaps, with a predicted 97% in silico-coverage of the region of interest. Double stranded DNA (250 ng) from frozen tissues and DNA from FFPE tissues (250–500 ng) were used for library preparation according to the TSCA protocol (Illumina). After normalization, three pooled quantified libraries were sequenced on the MiSeq platform (Illumina) using the 2 × 175 configuration and run on a MiSeq Reagent kit V3 sequencing flow cell with 175-cycle paired-end run plus two eight-cycle index reads. Data were automatically aligned using the MiSeq Reporter and visualized using Illumina Amplicon Viewer (Illumina) and Integrative Genomics Viewer 2.3 (Broad Institute, Boston, U.S.A.). The vcf files were imported into Illumina VariantStudio v2.2 software and analyzed with the default setting. SNV and in/del variants were annotated using different filtering options. Briefly, a) variant call quality was based on the selection of the “PASS” filter; b) to narrow results to the somatic events, annotation of the vast majority of constitutional variants was avoided by filtering only variants with global Minor Allele Frequency <1 and discarding variants recurrent in more tumour tissues of the same individual; c) frequency of the alternative allele was set at a threshold of 5%; d) the “consequence” option was used to filter data by variants that alter the coding potential of the transcript, i.e. frameshift, stop gained, initiator codon, inframe insertion or deletion, splice and missense with deleterious (SIFT) or damaging (Polyphen) prediction; synonymous, intronic, 5′- and 3′- UTR and missense variants with tolerated (SIFT) and benign (Polyphen) prediction were filtered out, unless they were reported in the Catalogue of somatic mutations in cancer (Cosmic) database (http://cancer.sanger.ac.uk/cosmic).

2.6. MUTYH Glycosylase Assays

Two DNA duplexes, each containing a single 8-oxodG:A mismatch with different flanking sequences, were used as substrates for adenine DNA glycosylase assays. HPLC-purified 30-mer oligonucleotides (5′-CTT GCC TAC GCC ACA AGC TAC TAC and 5′-CTT GCC TAC GCC ACA AGC TAC TAC), 3′-end labeled with 6-carboxyfluorescein (6-FAM) were annealed to the complementary strands containing a single 8-oxoG at 16 nt and 17 nt from their 5′end. All oligonucleotides were from Thermo Fisher Scientific (Thermo Fisher Scientific GmbH, Ulm, Germany). To obtain human purified MUTYH, Escherichia coli strain BL21 CodonPlus (DE3) RILP (Strategene, La Jolla, CA, USA) competent cells were transformed with a pMAL-c4E vector containing the MUTYH-cDNA. Bacteria were grown at 37 °C in Luria Broth medium with 5 g/l of glucose until OD 600 nm = 0.6 and then induced with 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 3 h at 25 °C. Cells were harvested and resuspended in phosphate-buffered saline (PBS) (1/40th of culture volume). In all, 5 mM dithiothreitol (DTT), 1 mg/ml of lysozyme and protease inhibitors (Roche) were added to the cell suspension and incubated for 30 min on ice. Lysates were cleared by centrifugation and the supernatant loaded on an amylose column (New England BioLabs Inc., Ipswich, MA). The maltose binding protein (MBP)-MUTYH fusion protein was eluted in 10 mM Tris–HCl, pH 7.5, 10 mM maltose buffer. Protein containing fractions were analyzed by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS–PAGE) and stained with Comassie blue dye. Protein recovery was evaluated by gel analysis by ImageJ software, available free online at the NIH website (http://rsb.info.nih.gov/). MUTYH active fraction was evaluated as described before (Turco et al., 2013). In single-turnover conditions ([MUTYH][DNA]), DNA substrate (10 nM) was reacted with an active protein (20 nM). Aliquots were withdrawn at different times, ranging from 30s to 10 min, and reaction was stopped by addition of NaOH (80 mM) and heating at 90 °C for 4 min. Reaction products were separated by denaturing PAGE. Gel images were acquired by Typhoon 9600 scanner (GE Healthcare, Uppsala, Sweden) and analysis performed by Image J and Kaleidagraph software (Synergy software).

2.7. Statistical Analysis

The mutation profiles were analyzed by means of principal Component Analysis of the matrix having as statistical units the different types of mutations and as variables the samples.

3. Results

3.1. Determination of Somatic Mutation Frequencies by Exome Sequencing in MAP CRC

Whole-exome DNA sequencing was performed on seven matched CRC/normal mucosa pairs from six unrelated MAP patients. These individuals harbour either the p.Gly96Asp and p.Tyr750Cys variants that are common among Caucasians or the p.Arg245His, p.Gly264Trpfs*7 and p.Glu480del that occur relatively frequently in Italian MAP patients (Aretz et al., 2014). A summary of the constitutional MUTYH genotypes, clinical phenotypes and pathology reports is listed in Table S1. The age of cancer diagnosis in these MAP patients is quite ample with a range of 27–73 years, in accordance with a recent Italian survey (Ricci et al., 2017). All these variants compromise MUTYH activity and are associated with increased steady-state levels of DNA 8-oxoG in lymphoblastoid cell lines (Molatore et al., 2010; D’Agostino et al., 2010; Ruggieri et al., 2013; Grasso et al., 2014).

The number of somatic non-synonymous single nucleotide variants (SNVs) and small insertions/deletions (indels) was similar among MAP CRCs, with a median value of 340 mutations per sample (range: 247–423). The mutational landscape of MAP tumours was compared to a set of 23 well-characterized colorectal cancer stem cell lines (SCCs) previously shown to faithfully reproduce the genomic complexity of CRC (De Angelis et al., 2016). As expected, the number of mutations in these SCCs was dependent on MMR status, with median values of 210 and 2215 mutations/exome in MMR-proficient (MSS) and MMR–deficient (MSI) SCCs, respectively. These values correspond to 3.5 ± 0.3 and 34.6 ± 6.9 mutations per Mb for MSS and MSI SCCs, respectively (Fig. 1A and B) and are in good agreement with those reported by The Cancer Genome Atlas Network (2012). MAP tumours contained 5.3 ± 0.38 mutations per Mb, an increase of 1.5-fold over the frequency in MMR-proficient SCCs (Fig. 1A). Approximate values for mutation rates were derived by dividing the number of somatic mutations per tumour by the age of the patient at diagnosis (Tomasetti et al., 2015). Median ages of cancer diagnosis were 47 in MAP patients and 73 years for SCCs (range: 27–73 and 47–88 years, respectively). The mutation rate in MAP CRCs (6.8 ± 0.6) was 2-fold higher than that in MSS SCCs (3.3 ± 0.3) (p-value = 2.4 × 10⁻⁶ (Fig. 1C). In contrast MAP SCCs were characterized by an extreme value of 31.7 ± 6 with an almost 10-fold increase over MSS SCCs (data not shown).

The higher frequency of mutations in MAP tumours was driven predominantly by an increase in SNVs and the number of indels was relatively low compared to either MSS or MSI SCCs (Fig. 1D). MSI is
not a characteristic of MAP CRCs and there were no detectable alterations of microsatellite markers \((NR21, BAT-26, BAT-25, NR-24, MONO-27)\) (Fig. S1).

MAP CRCs exhibited a homogeneous mutational spectrum in the seven patients, which was characterized by the prevalence of two mutational classes, G:C→T:A transversions and G:C→A:T transitions (Fig. 2A).
By comparison with MSS and MSI CSCs, there was a significant overrepresentation of G:C→T:A in MAP CRC, with all the other mutational classes occurring at decreased frequencies (Fig. 2B). The transversions were equally distributed between coding and non-coding strands (1.01 ratio). Taken together, these findings are fully consistent with mutation induction by persistent DNA 8-oxoG:A mismatches resulting from the MUTYH defect.

3.2. Fingerprint of Base Substitutions in MAP CRCs

To identify the mutational signature associated with persistent 8-oxoG:A mispairs, we performed a refined classification of base substitutions to include the 3′ and 5′ flanking bases at the mutated site (Alexandrov et al., 2013a,b). This identified a strongly conserved MAP signature, mostly characterized by biases among the possible positions of C>A and C>T base substitutions (substitutions in mutational signatures are referred to by the pyrimidine of the mutated Watson–Crick base pair; Fig. 3). It is noteworthy that identical SNVs distributions occurred in tumours from the distal and proximal colon of MAP796 patient (T1 and T2 in the sigmoid and caecum, respectively) indicating that these tumours have a common molecular origin (Fig. 3).

To determine whether the mutational profiles of MAP CRCs differed from those observed in CSCs we applied principal component analysis (PCA). This revealed a perfect separation between the signatures of

![Fingerprints of transitions and transversions in MAP CRCs. Mutational signatures are defined by the substitution class (in different colours) and the sequence context immediately 5′ and 3′ to the mutated base. The mutation types are on the horizontal axes, whereas vertical axes show the recurrence of mutations occurring in the specific sequence context.](image)
MAP and CSCs (Fig. 4A and Table S2). The signatures endowed with the greater discriminant power for this separation were C->A transversions at 5’-TCT-3’, 5’-TCA-3’ and, to a lower extent, 5’-GCA-3’, 5’-CCA-3’, 5’-TCC-3’, 5’-GCT-3’ and 5’-CCT-3’ trinucleotides (Fig. 4B). All these signatures identify sequences where the damaged base (8-oxoG) is assumed to be the complementary guanine in the C-G pair.

Trinucleotide signatures from MAP CRCs and CSCs were compared to a catalogue of known mutational signatures identified from >12,000 samples derived from 40 types of human cancer (Alexandrov et al., 2013a,b). The frequently observed Signatures 1 and 5 were also present in MAP and CSCs (Fig. 4C and D) (Alexandrov et al., 2013a). In addition, fingerprints associated with MMR-defective CSCs corresponded to Signature 36, which is almost exclusively defined by the specific locations of G:C->T:A transversions (Fig. 4C and D). Signature 36 identifies the location at which mutations arise because of unrepaired 8-oxoG:A mispairs and is essentially identical (correlation 0.98) to that reported for germline mutations occurring in Mutyh/Ogg1/Mth1 triple knock-out mice (Ohno et al., 2014) (Fig. 4D). This suggests that the unique spectrum of mutations in MAP tumours mainly reflects MUTYH loss and the sequence specificity of DNA 8-oxoG formation.

### 3.3. Identification of Oncogenes/Tumour Suppressors Specifically Associated with MAP CRC

It has been proposed that carcinogenesis in MUTYH-associated polyposis follows a distinct genetic pathway (Lipton et al., 2003). Recurrently mutated genes in 7 MAP CRCs were identified using MuTect and manual curation (Fig. 5A). These included APC, KRAS, PIK3CA, FAT4, TP53, FAT1, AMER1 (or WTX), KDM6A, SMAD4 and SMAD2 (a more complete list of genes and detailed mutation analysis are reported in Tables S3 and S4). Targeted exome sequencing of 10 additional MAP CRCs (Table S5) confirmed the MAP-related mutagenic profiles of these 10 genes (Fig. 5B).

To examine whether these mutations were present at the adenoma stage, targeted sequencing was performed in 26 MAP adenomas (Fig. 5C and Table S5). AMER1 and APC were equally mutated in adenomas and CRCs consistent with early occurrence and confirming their role as driver genes in MAP (Rashid et al., 2016). In contrast SMAD2, SMAD4 and TP53 were rarely mutated in adenomas indicating that they represent late events in colorectal carcinogenesis. It was not possible to assign early or late occurrence to the other five genes.

Comparison of MAP mutations to 376 MMR-proficient and -deficient CRCs in COSMIC identified three different patterns of putative driver mutations (Fig. 5D). KRAS, SMAD4, SMAD2 and KDM6A

---

**Fig. 4.** Mutational signatures in MAP CRCs and MSS and MSI CSCs. a) Separation between MAP (red) and CSCs (blue) mutational fingerprints by PCA. Factor 1 and Factor 2 represent commonality and differences among profiles, respectively. b) G:C->T:A transversions at the indicated trinucleotides identify signatures with the greater discriminant power for MAP/SCS separation. c) Somatic mutations in MAP CRCs (red) and MSI/MSS CSCs (black) attributed to reported mutational signatures (Alexandrov et al., 2013a,b). d) Graphical representations of signatures identified in MSI/MSS CSCs and MAP CRCs. Signature 8-oxoguanine derives from mice defective in Mutyh/Ogg1/Mth1 (Ohno et al., 2014).
were preferentially mutated in MAP compared to MSS and MSI CRCs. A second group of genes (FAT4, FAT1 AMER1 and PIK3CA) were mutated at similar frequencies in MAP and MSI CRCs. APC and TP53 comprised a third group that was equally mutated in all CRCs.

Thus, alterations in MAP CRCs affect the well-established WNT, TGFβ, PI3K, RAS and p53 signalling pathways (Cancer Genome Atlas Network, 2012). The mutational profiles also reveal the specificity of MAP oncogenesis, with limited features in common with MMR-proficient and -deficient CRC.

3.4. Mutational Signature in Oncogenes/Tumour Suppressors

The identified trinucleotide mutational signature for G:C→T:A transversions was also present in recognized target genes for MAP (Al-Tassan et al., 2002; Jones et al., 2002, 2004). An excess of G:C→T:A transversions in APC leading to protein truncation has been reported (Al-Tassan et al., 2002; Sieber et al., 2003; Jones et al., 2002). Besides confirming this observation (G:C→T:A were 23/32, >70% of the identified APC variants), we noticed that these transversions occurred almost exclusively in the identified “hot-spots” for 8-oxoG (Table 1). Similarly the majority of SNV in the AMER1 driver gene were also G:C→T:A transversions (14/24, 58% of total mutations) in the first base of GAG and GAA codons that lead to the formation of TAG and TAA stop codons. Analysis of the 5’ and 3’ flanking bases at the mutated site confirmed that these mutations are also within the recognized trinucleotide signature (i.e. Signature 36) for 8-oxoG (Table 1).

Missense G:C→T:A transversions were also preferentially located in identified “hot-spots” of DNA oxidation in other key genes for MAP (PIK3CA, SMAD4 and SMAD2) (Table S6). This was true even if the frequency of the transversions was lower than expected (35–20% in TP53, FAT1, FAT4) (Table S7).

KRAS mutations in MAP CRCs exhibited a distinctive pattern of mutations. They were exclusively G:C→T:A transversions at the 1st G of codon 12 (GGT→TGT, nt 34) to generate a MAP-specific G12C variant (Jones et al., 2002, 2004; Venesio et al., 2013). This is in contrast to sporadic CRCs in which G:C→T:A transversions predominantly affect the 2nd G of the same codon (GGT→GT, nt 35) leading to the G12V variant (Fig. 6A) (Cox et al., 2014). Indeed G preceded by a T in 5′-TG/G-3′ is a
mutational “hot-spot” in MAP CRCs, whereas G preceded by G in $5'$-GG"T-3' is a “cold-spot” (Fig. 6B). To examine whether this mutational bias reflected the sequence dependence of DNA 8-oxoG formation or the particularly efficient processing of 8-oxoG at nt 34 by active MUTYH, we examined the activity of purified MUTYH on duplexes containing a single 8-oxoG:A basepair located at nt 34 or nt 35 of the KRAS sequence (Fig. 6A). We did not observe any difference in MUTYH processing of these two substrates (Fig. 6B,C) and we conclude that the overrepresentation of the G12C mutations in MAP CRC reflects the particular susceptibility of C34 to oxidation. Thus, the different probabilities of ROS-mediated guanine oxidation are likely to be a significant determinant in the mutational signature in MAP CRCs.

4. Discussion

Our analysis of cancer-related mutations associated with MUTYH inactivation in MAP reveals that the absence of MUTYH-mediated processing of 8-oxoG:A mispairs results in a relatively modest mutator phenotype. The increase in mutation rates in MAP CRC is similar to that in lymphoblastoid cell lines established from MAP patients (1.5- to 4.0-fold) (Ruggieri et al., 2013; Grasso et al., 2014). It therefore appears that even a small increase in mutation rate due to impaired repair of pre-mutagenic oxidative DNA lesions is an important risk factor for cancer. CRC is the main cancer in MAP. It occurs with a relatively early age of onset (47 years in our MAP cases) and a lifetime risk approaching 100% (Nielsen et al., 2010). It is noteworthy that MUTYH-defective cells are hypermutable by oxidants, indicating that the mild mutator phenotype can be significantly enhanced under oxidative stress conditions (Ruggieri et al., 2013; Grasso et al., 2014). Recently a correlation has been drawn between the lifetime risk of cancer and the total number of cell divisions in a given tissue (Tomasetti and Vogelstein, 2015). Based on the present observations we propose that a combination of endogenous and exogenous oxidation sources (environmental factors) and lack of DNA repair (inherited predisposition) acting on a highly self-renewing tissue are the main determinants of the increased MAP cancer susceptibility.

Detailed analysis of MAP CRC mutations revealed a substantial increase in a single class of base substitutions, G:C→T:A transversions. This is in accordance with the expected persistence of DNA 8-oxoG:A mismatches, the preferred MUTYH substrate. MUTYH also removes 2-hydroxyadenine from 2-hydroxyadenine: G mispairs (Ohnsubo et al., 2015). If unrepaired, these generate A:T→C:G transversions (Satou et al., 2007). We did not observe increase in this class of mutations in MAP CRCs and it appears that oxidation of DNA adenine is most likely a relatively infrequent occurrence. Mutations associated with the miscoding properties of DNA 8-oxoG are therefore the hallmark of MUTYH inactivation.

The MAP CRC G:C→T:A transversions were equally distributed in the transcribed and non-transcribed strands. This is at variance with the strong bias towards higher rates of G:C→T:A transversions located in the non-transcribed strand in cutaneous squamous cell carcinoma and UV-irradiated normal human skin (Martincorena et al., 2015). The lack of transcriptional strand bias is a strong indicator that 8-oxoG:A mismatches are not processed by transcription-coupled BER in the colon.

More importantly, analysis of the sequence context of the G:C→T:A transversions identified a specific mutational signature due to unrepaired DNA 8-oxoG (Signature 36). It is particularly significant that Signature 36 is almost identical to the signature of spontaneous mutations in Ogg1−/−, MutYh−/−, Mth1−/− mice in which all known repair systems for DNA 8-oxoG are disabled (Ohnsubo et al., 2014). These coincident signatures indicate that MUTYH plays the leading role in protecting the human genome against mutations induced by DNA 8-oxoG.

The types and location of mutations in oncogene/tumour suppressor genes also implicate DNA 8-oxoG in MAP CRC development. KRAS activation provides the best example of this. 100% of KRAS mutations are codon 12 G:C→T:A transversions that correspond to the acknowledged preferential oxidation of $5'$-G of GG and GGG sequences under oxidative stress conditions.

### Table 1

<table>
<thead>
<tr>
<th>Location of G:C→T:A transversions in the APC and AMER1 genes.a</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tumour</strong></td>
</tr>
<tr>
<td>MAP104T1</td>
</tr>
<tr>
<td>MAP356T2</td>
</tr>
<tr>
<td>MAP443T1</td>
</tr>
<tr>
<td>MAP256</td>
</tr>
<tr>
<td>MAP873T1</td>
</tr>
<tr>
<td>MAP450</td>
</tr>
<tr>
<td>MAP356T1</td>
</tr>
<tr>
<td>MAP797</td>
</tr>
<tr>
<td>MAP902T1</td>
</tr>
<tr>
<td>MAP901T1</td>
</tr>
<tr>
<td>MAP173</td>
</tr>
<tr>
<td>MAP976T1</td>
</tr>
<tr>
<td>MAP987T1</td>
</tr>
<tr>
<td>MAP798</td>
</tr>
<tr>
<td>MAP986T1</td>
</tr>
</tbody>
</table>

| **AMER1** | **G:C→T:A 14/24 (58.3%)** |
|---|
| **Tumour** | **Aa change** | **Sequence context b** | **Trinucleotide c** |
| MAP422T1 | E334* | AGAA>AT | AGA>TCT |
| MAP797 | E386* | AGAA>AT | AGA>TCT |
| MAP256 | E386* | AGAA>AT | AGA>TCT |
| MAP483T2 | E386* | AGAA>AT | AGA>TCT |
| MAP577T1 | E394* | AGAA>AT | AGA>TCT |
| MAP577T1 | E394* | AGAA>AT | AGA>TCT |
| MAP677T | E395* | TGGC>TAG | TGA>TCA |
| MAP181T1 | E407* | AGAA>AT | AGA>TCT |
| MAP181T6 | E462* | AGAA>AT | AGA>TCT |
| MAP901T1 | E469* | TGGT>TA | TGA>TCA |
| MAP577T1 | E480* | TGGT>TA | TGA>TCA |
| MAP483T4 | E485* | TGGT>TA | TGA>TCA |
| MAP798 | E587* | AGAA>AT | AGA>TCT |
| MAP356T1 | D753Y | AGAA>AT | AGA>TCT |

---

a The sequence context of G:C→T:A transversions in APC is shown only in CRCs and in both adenomas and CRCs (bolded) in AMER1.

b Codons are shown in red and the mutated base is underlined.

"Hot-spot" and "cold-spot" trinucleotides are respectively in red and black.
The probability of mutating specific oncogenes/tumour suppressors is a major determinant of the ionization potential

Fig. 6. KRAS sequence context and MUTYH adenine glycosylase activity. a) Duplex oligonucleotides containing a single 8-oxoG:A mispair at nt 34 or 35 of the KRAS sequence which are responsible for the G12C and G12V variants in MAP (top) and sporadic CRC (bottom). Codon 12 is in bold and trinucleotides containing 8-oxoG (G\textsuperscript{*}) are underlined. b) Base substitution frequency of C\textsuperscript{\textprime}–A transversions in MAP CRC. The black and red bars represent the GC\textsuperscript{\textprime} “cold-spot” and TG\textsuperscript{\textprime} “hot-spot” trinucleotides, respectively. c) Representative gels of single-turnover MUTYH assays using duplexes shown in panel A. DNA glycosylase activity of human MUTYH on single 8-oxoG:A mispairs was assayed using 10 nM duplex substrates and 20 nM active protein. Reaction products are indicated by an arrow. Lanes 1–9: 0, 0.5, 1, 2, 3, 4, 5, 10 and 20 min. d) Plots of MUTYH glycosylase assay data under single-turnover conditions. Data were derived from image analysis of gels by Image J and Kaleidagraph softwares. Values are the average from two independent experiments. G12C sequence (full black circle); G12 V (empty blue circle).

Mutations in other “driver” genes of colorectal carcinogenesis such as APC and AMER1 also showed the 8-oxoG-dependent sequence context specificity. In this case mutations occurred mainly at AGA and TGA sites. Thus similarly to the 5′-GG-3′ KRAS hot-spot, the G in GA dinucleotides is also easily oxidized. The preferential oxidation of G in GA or GG is consistent with a reduced ionization potential of these purine dinucleotides. Base type, orientation and stacking interaction of consecutive bases are the main determinants of the ionization potential (Sugiyama and Saito, 1996). The majority of mutations in both APC and AMER1 were G:C→T:A transversions generating stop codons. Interestingly the relatively rare AMER1 mutations in sporadic CRC (10%) show a wider spectrum of molecular changes including stop gain, missense, and synonymous variants (Sanz-Pamplona et al., 2015). Our data suggest that the accumulation of sequence specific 8-oxoG:A mispairs due to MUTYH inactivation is the cause of the increased mutability of this gene and the strong drift towards stop codon formation. The observation that a specific 8-oxoG mutational fingerprint increases the probability of mutating specific oncogenes/tumour suppressors is a rule confirmed in several examples (such as PIK3CA, SMAD4, SMAD2, TP53, FAT4 and P53). Although the pathways involved in MAP CRC are the WNT, TGFR, PTK, RAS and p53 signalling pathways (Cancer Genome Atlas Network, 2012), which are commonly mutated in sporadic CRC, the MUTYH defect directs oncogenesis towards specific genes that are mutated less frequently in sporadic CRCs. In this sense
MB and AV designed the study, AB, EM and MT analyzed and interpreted whole-exome sequencing data. AV and FM collected and prepared tumour samples and analyzed targeted sequencing data. VC, MA and RM performed genetic testing and somatic interpretation. EP and MBa provided a molecular characterization of CSCs. GG provided technical support and conceptual advice. MB and AV designed the study. AB, EM and MT analyzed and interpreted whole-exome sequencing data in mice defective in removal of oxidative base damage. Proc. Natl. Acad. Sci. U. S. A. 96, 13300–13305.


