Catalytically impaired hMYH and NEIL1 mutant proteins identified in patients with primary sclerosing cholangitis and cholangiocarcinoma

Monika Forsbring[1], Erik S.Vik[1], Bjørn Dalhus[1,2], Tom H.Karlsen[3], Annika Bergquist[4], Erik Schrumpf[3], Magnar Bjørås[1,2], Kirsten M.Boberg[3] and Ingrun Alseth[1,*]

[1]Institute of Medical Microbiology, Rikshospitalet University Hospital HF and Centre for Molecular Biology and Neuroscience, [2]Institute of Clinical Biochemistry, University of Oslo, N-0027 Oslo, Norway, [3]Department of Medicine, Rikshospitalet University Hospital HF, N-0027 Oslo, Norway and [4]Department of Gastroenterology and Hepatology, Karolinska University Hospital, Huddinge, S-14186 Stockholm, Sweden

*To whom correspondence should be addressed. Tel: +47 23074069;
Fax: +47 23074061;
Email: ingrun.alseth@rr-research.no

The human *hMYH* and *NEIL1* genes encode DNA glycosylases involved in repair of oxidative base damage and mutations in these genes are associated with certain cancers. Primary sclerosing cholangitis (PSC), a chronic cholestatic liver disease characterized by inflammatory destruction of the biliary tree, is often complicated by the development of cholangiocarcinoma (CCA). Here, we aimed to investigate the influence of genetic variations in the *hMYH* and *NEIL1* genes on risk of CCA in PSC patients. The *hMYH* and *NEIL1* gene loci in addition to the DNA repair genes *hOGG1*, *NTHL1* and *NUDT1* were analyzed in 66 PSC patients (37 with CCA and 29 without cancer) by complete genomic sequencing of exons and adjacent intronic regions. Several single-nucleotide polymorphisms and mutations were identified and severe impairment of protein function was observed for three non-synonymous variants. The NEIL1 G83D mutant was dysfunctional for the major oxidation products 7,8-dihydro-8-oxoguanine (8oxoG), thymine glycol and dihydrothymine in duplex DNA, and the ability to perform δ-elimination at abasic sites was significantly reduced. The hMYH R260Q mutant had severe defect in adenine DNA glycosylase activity, whereas hMYH H434D could excise adenines from A:8oxoG pairs but not from A:G mispairs. We found no overall associations between the 18 identified variants and susceptibility to CCA in PSC patients; however, the impaired variants may be of significance for carcinogenesis in general. Our findings demonstrate the importance of complete resequencing of selected candidate genes in order to identify rare genetic variants and their possible contribution to individual susceptibility to cancer development.

Introduction

Oxidative DNA damage induced by reactive oxygen species is a major source of mutation load in living organisms and is believed to play a causative role in aging and degenerative diseases such as cancer (1,2). Reactive oxygen species are formed as by-products of the electron transport machinery, during conditions of chronic inflammation and by exogenous stimuli such as ionizing radiation and various chemical oxidants (3). 7,8-dihydro-8-oxoguanine (8oxoG) is one of the most stable products of oxidative DNA damage and has deleterious effects because it can mispair with adenine during DNA replication (4,5). Many tumors have been shown to have G:C to T:A transversion mutations associated with 8oxoG and also to have increased levels of 8oxoG in the DNA (6–8).

To counteract the potential harmful effects of oxidative DNA damage, cells primarily use the base excision repair (BER) pathway. BER is initiated by lesion-specific DNA glycosylases that identify and excise the damaged base, leaving an abasic (apurinic/apyrimidinic) site in the DNA that is further processed by the other enzymes of the BER pathway (reviewed in ref. 9). Mammalian cells are equipped with several different DNA glycosylases and 8oxoG DNA glycosylase (hOGG1) is the major enzyme for removal of 8oxoG from an 8oxoG:C pair but not when 8oxoG is already mispaired with adenine. In this case, A is removed by adenine DNA glycosylase (hMutY/hMYH) (reviewed in ref. 10). Finally, the NUDT1 (MutT in bacteria) enzyme hydrolyzes 8oxoG triphosphate to 8oxoG monophosphate, thus avoiding the incorporation of oxidized nucleotides into genomic DNA. The DNA glycosylases endonuclease III (NTHL1) and endonuclease VIII-like 1 (NEIL1) are the principal enzymes for removal of oxidized pyrimidines such as thymine glycol (Tg), 5-hydroxy cytosine (5ohC) and 5-hydroxy uracil (5ohU) (reviewed in ref. 11).

Recent studies suggest that polymorphisms in DNA glycosylase genes are associated with several cancers. For example, polymorphisms in *NEIL1* have been linked to gastric cancer (12), whereas those in *hOGG1* have been related to higher risk for prostate (13) and lung cancer (reviewed in ref. 14). For the *hMYH* gene, inactivating mutations and polymorphisms are associated with development of colorectal polyposis (MutY associated-polyposis) (7,15). Also, approaches based on functional analyses have shown that reduced 8oxoG DNA glycosylase activity is a risk factor in lung, head and neck cancer (16,17).

Primary sclerosing cholangitis (PSC) is a chronic inflammatory disease of the bile ducts leading to bile duct strictures and eventually liver cirrhosis and liver transplantation (18). Importantly, PSC is a considerable risk factor for the development of cholangiocarcinoma (CCA) that results from a malignant transformation of bile duct epithelial cells. The prevalence of CCA in PSC ranges from 6–13% in most reports (19–21). The duration and severity of PSC does not seem to correlate with development of CCA, suggesting that individual susceptibility factors are important (22,23). As recently demonstrated (24), genetic variants may influence risk of CCA in PSC. Further, an emerging concept in association studies is that rare sequence variants with strong phenotypic effects might contribute substantially to the aggregated risk of complex traits, including cancer. The exact mechanism for CCA development in PSC is unknown. In experimental models, bile has been shown to induce oxidative stress and to increase oxidative DNA damage in cholangiocytes (25). Consequently, individual differences in processing of DNA damage could be of importance for transformation of the inflammatory changes of the biliary tree into cancer. In this study, we have evaluated the role of germ line variation in genes important for repair of oxidative DNA damage and their potential role for development of CCA in PSC patients.

Materials and methods

Patients and controls

The study included 66 Scandinavian PSC patients recruited on admission to the Medical Department, Rikshospitalet University Hospital, Oslo, Norway,

Abbreviations: BER, base excision repair; CCA, cholangiocarcinoma; DHT, dihydrothymine; EGFP, enhanced green fluorescent protein; EMSA, electrophoretic mobility shift assay; 5ohC, 5-hydroxy cytosine; 5ohU, 5-hydroxy uracil; 8oxoG, 7,8-dihydro-8-oxoguanine; PSC, primary sclerosing cholangitis; SNP, single-nucleotide polymorphism; ssDNA, single-stranded DNA; Tg, thymine glycol; WT, wild-type.

(*n* = 48) and Division of Gastroenterology and Hepatology, Karolinska University Hospital, Huddinge, Stockholm, Sweden (*n* = 18). PSC was diagnosed according to accepted criteria, including characteristic cholangiographic findings (18). Inflammatory bowel disease was diagnosed by colonoscopy with biopsies and classified according to standard criteria. The cohort comprised 37 (28 males) PSC patients with CCA (one patient with a carcinoma of the gallbladder was included in this group) and 29 (20 males) without cancer of the biliary tract. CCA was confirmed by histology. Concomitant inflammatory bowel disease had been diagnosed in 78% of the patients with CCA and in 86% of those without. An ethnically matched group of healthy controls (*n* = 30) was randomly selected from the Norwegian Bone Marrow Donor Registry. All participants gave their written informed consent. The study was approved by The Regional Committee for Research Ethics in Southern Norway and The Ethics Committee of Karolinska Institutet.

DNA sequencing

Genomic DNA was extracted from peripheral blood leukocytes from PSC patients and controls and whole genome amplification was performed using the GenomiPhi® kit (GE Healthcare Systems, Chalfont St Giles, UK), yielding high molecular amplified DNA thoroughly validated for genotyping (26).

The exons including flanking intronic regions of the genes *hMYH*, *hOGG1*, *NEIL1*, *NUDT1* and *NTHL1* were amplified using primers listed in supplementary Table 1 (available at *Carcinogenesis* Online). Detailed polymerase chain reaction conditions are freely available from the authors upon request. Subsequent sequencing (primers available upon request) was performed with Big-Dye® Terminator v3.1 chemistry (Applied Biosystems, Foster City, CA) and an ABI3730 capillary sequencer (Applied Biosystems) according to manufacturer's protocols. Subsequent mutation detection was performed using Seq-Scape v2.5 (Applied Biosystems).

Construction and purification of hMYH, hOGG1 and NEIL1mutants

Primers used for site-specific mutagenesis are listed in supplementary Table 2 (available at *Carcinogenesis* Online). pGEV1-*hMYH* (27) was used to generate the three *hMYH* mutants (R260Q, H434D and S501F) with QuickChange polymerase chain reaction. The *hOGG1* gene was subcloned from pUC18-*hOGG1* (28) into the EcoRI and BglII sites of pETDuet-1 (Novagen, Madison, WI) and this construct was used as the template to generate pETDuet-1-*hOGG1*-S31P. *NEIL1*cDNA subcloned into the NdeI and XhoI sites of pET22b was used as template to generate the two mutated constructs pET22b-*NEIL1*-G83D and pET22b-*NEIL1*-E181K. All constructs were verified by DNA sequencing of the entire open reading frames. Proteins were expressed in *Escherichia coli* and purified as described in supplementary Data (available at *Carcinogenesis* Online). Purified proteins were analyzed by sodium dodecyl sulfate–polyacrylamide gel and detected by Coomassie Blue staining (supplementary Figure 1 is available at *Carcinogenesis* Online).

Assays for DNA glycosylase activity

Substrates for enzyme cleavage analysis were: 5′-ctcggtacccggggatctggccatg[A]ccgtgcaggcatgc (8oxoG or G opposite the marked A gave the A:8oxoG and A:G substrates, respectively); 5′-ggcggcatgaccc[8oxoG]gaggcccatc (C opposite 8oxoG gave the 8oxoG:C substrate); 5′-ctcgtcagcatca[Tg]catcatacagtcagtg (to yield the stereoisomer *5R* Tg:A substrate) and 5′-gcatgcctgcacgg[x]catggccagatccccgggtaccgag [x is dihydrothymine (DHT), 5ohC, 5ohU or U with A, G, G or G, respectively, opposite the damage in the complementary strand gave the DHT:A, 5ohC:G, 5ohU:G or U:G substrates). All oligonucleotides were 5′ end labeled with T4 polynucleotide kinase (New England Biolabs, Ipswich, MA) and [γ-³²P]adenosine triphosphate (Amersham Biosciences, Pittsburgh, PA). In order to generate double-stranded substrates, the labeled oligonucleotides were annealed to their respective complementary strands by heating the solution to 90°C for 2 min followed by slow cooling to room temperature. The U:G substrate was treated with uracil DNA glycosylase (New England Biolabs) to make the apurinic/apyrimidinic substrate.

Enzyme reactions were carried out in 70 mM 3-(*N*-morpholino)propanesulfonic acid, pH 7.5, 1 mM dithiothreitol, 1 mM ethylenediaminetetraacetic acid and 5% glycerol for 30 min at 37°C. Reaction mixtures contained 10 fmol substrate and enzymes as indicated in the figures in a total volume of 10 µl. To obtain strand cleavage after hMYH and hOGG1 base removal, NaOH (100 mM) was added to the reactions and incubation continued at 70°C for 15 min before neutralization with HCl (100 mM). All reactions were stopped by adding formamide DNA loading buffer. The reaction products were analyzed by 20% denaturing polyacrylamide gel electrophoresis and bands visualized by Phosphorimaging. Quantifications were done with the ImageQuant Software.

FaPy DNA glycosylase assays were done in the same reaction buffer as above and [*N*-³H]methyl-*N*′-nitrosourea (18 Ci/mmol) was used to prepare poly (dG·dC) DNA containing faPy residues (5000 d.p.m./µg DNA) as described (29). Briefly, DNA glycosylase activity was measured in a total volume of 50 µl containing 0.4 µg [³H-]faPy-poly(dG-dC) with different amounts of

enzyme as indicated. After incubation at 37°C for 30 min, DNA was precipitated with ethanol. The radioactivity of the supernatant was determined by scintillation counting.

Electrophoretic mobility shift assays

DNA binding of hMYH variants was determined by electrophoretic mobility shift assays (EMSAs) using ³²P-5′-end labeled duplex DNA substrates containing an A:8oxoG or an A:G basepair. DNA (10 fmol) and enzymes as indicated in the figure were incubated on ice for 30 min in 20 mM Tris–HCl (pH7.5), 80 mM NaCl, 1 mM ethylenediaminetetraacetic acid, 1 mM dithiothreitol and 3% glycerol in a total volume of 10 µl. Two-hundred femtomole competitor DNA (same duplex without damage) was included in all reactions. Reactions were stopped by adding 2 µl 6× DNA loading buffer (Fermentas, Ontario, Canada). Samples were analyzed on 10% native polyacrylamide gel electrophoresis in 1× Taurin at 4°C and visualized by Phosphorimaging. EMSAs for hOGG1 and NEIL1 enzymes were performed in the same buffer as the DNA glycosylase assays with substrates as indicated. hOGG1 samples were analyzed as for hMYH, whereas the NEIL1 samples were run in 0.5× Tris/Borate/EDTA.

Intracellular localization of NEIL1 variants in HeLa cells

HeLa S3cells were cultivated in Dulbecco's Modified Eagle's Medium with 4.5 g/l glucose, 10% fetal calf serum, 0.3 mg/ml glutamine, 100 U/ml penicillin and 0.1 mg/ml streptomycin at 37°C in a 5% CO₂ air atmosphere. Transfections of HeLa cells were done using FuGENE 6 Transfection Reagent (Roche Applied Science, Mannheim, Germany) according to the manufacturer's recommendations. Plasmids for transfections were pEGFP-*NEIL1* (30), pEGFP-*NEIL1*-G83D, pEGFP-*NEIL1*-E181K and pEGFP-N1 (Clontech, Montain View, CA) as control. The site-specific mutants were generated by Quick-Change polymerase chain reaction with the same primers as for the expression constructs.

One or two days after transfection, the cells were stained with 5 µg/ml Hoechst 33342 (H-3570, Molecular Probes, Carlsbad, CA) at 37°C for 20 min and imaged directly with a Zeiss Axiovert 200M fluorescence microscope (Carl Zeiss, Jena, Germany).

Results

DNA from patients with PSC (*n* = 29), PSC with CCA (*n* = 37) and healthy controls (*n* = 30) were screened for sequence variations in the *hMYH, hOGG1, NEIL1, NUDT1* and *NTHL1*genes. In total, 18 sequence variants were identified in the coding regions of the five DNA repair genes examined (Table I). Association analysis with Bonferoni correction (18 tests) revealed no significant differences in genotype frequencies between PSC patients with and without CCA (Table I). Of the 18 variants, six were validated single-nucleotide polymorphisms (SNPs), whereas 12 were novel mutations.

In a functional perspective, two splice site mutations were found in two healthy controls in the *NEIL1* and *NUDT1* gene, respectively. Three of 12 novel mutations were synonymous, whereas nine were non-synonymous. All these 12 mutations were monoallelic and were found in the *hMYH* (4), *hOGG1* (5) and *NEIL1* (3) genes. For *NUDT1*, two frequent SNPs (V83M and D119) were found but no new mutations and, interestingly, no sequence variation at all was identified in the *NTHL1* gene. All novel non-synonymous mutations (except one non-synonymous but conservative mutation) identified in the PSC or PSC-CCA patients were subjected to functional analysis.

Analysis of hMYH

Of the four non-synonymous *hMYH* mutations identified, one was in the healthy control group (E155Q) and three in the PSC group (R260Q, H434D and S501F) (Table I, Figure 1A). The S501F mutation was also found in one PSC patient with CCA. From the 3D structure of MutY from *Bacillus stearothermophilus* (31), it can be inferred that amino acid Arg260 in hMYH is located on the surface of hMYH, on the opposite side of the DNA-binding groove (Figure 1A). Further, His434 resides within a loop involved in 8oxoG recognition (Figure 1A) and interacts with the phosphate backbone of the DNA strand containing 8oxoG. Finally, Ser501 is in a region predicted to be disordered that is also outside of the structural model of *B.stearothermophilus* MutY.

The three mutations found in the PSC patients were introduced into *hMYH* expression plasmids and the mutant proteins were expressed in

Table I. Sequence changes and allele frequencies in four DNA repair genes[a] in healthy, PSC and PSC-CCA populations

Gene	Reference/variant allele	Major/minor allele	Homozygotic or heterozygotic SNP/mutation	dbSNP ID	Amino acid change	Amino acid sequence accession number	Minor allele frequency		
							Healthy	PSC	PSC-CCA
hMYH	agcc(G/A)tggg	G/A	Het	rs3219484	Val22Met	AAC50618	0.21	0.07	0.11
	agcc(G/A)tggg	—/A	Homo	rs3219484	Val22Met	AAC50618	0	0.03	0
	acag(G/C)aggt[b]	G/C	Het	—	Glu155Gln	AAC50618	0.03	0	0
	gccc(G/A)gcca	G/A	Het	—	Arg260Gln[c]	AAC50618	0	0.03	0
	gaca(G/C)tgcc	G/C	Het	rs3219489	Gln324His	AAC50618	0.34	0.41	0.35
	gaca(G/C)tgcc	—/C	Homo	rs3219489	Gln324His	AAC50618	0.03	0	0.08
	ctct(C/G)acat	C/G	Het	—	His434Asp[c]	AAC50618	0	0.03	0
	tcct(C/T)tccg	C/T	Het	—	Ser501Phe[c]	AAC50618	0	0.03	0.03
hOGG1	tcgc(T/C)ctga	T/C	Het	—	Ser31Pro[c]	BAA19103.1	0	0.03	0
	gtgt(A/T)ctag	A/T	Het	—	Val58	BAA19103.1	0	0.03	0
	tggg(G/A)acct	G/A	Het	—	Gly308Glu	BAA19103.1	0.03	0	0
	caat(C/G)ccgc	C/G	Het	rs1052133	Ser326Cys	BAA19103.1	0.41	0.48	0.30
	caat(C/G)ccgc	—/G	Homo	rs1052133	Ser326Cys	BAA19103.1	0.10	0.07	0.05
	aagg(C/A)cccc	C/A	Het	—	Gly342	O15527-4[d]	0	0.07	0
	atga(C/G)ccag	C/G	Het	—	Thr398Ser	O15527-4[d]	0	0.03	0
NEIL1	tccg(G/A)ctct	G/A	Het	rs5745906	Gly83Asp[c]	Q96FI4	0	0	0.05
	ggca(G/A)agat	G/A	Het	—	Glu181Lys[c]	Q96FI4	0	0.03	0
	ggca(T/C)ggcc	T/C	Het	—	His275	Q96FI4	0	0	0.03
	ccag(C/T)ggcc	C/T	Het	—	Arg339Trp	Q96FI4	0.03	0	0
NUDT1	ggac(G/A)tgca	G/A	Het	rs4866	Val83Met	CAG28583.1	0.03	0.10	0
	ccga(C/T)gaca	C/T	Het	rs1799832	Asp119	CAG28583.1	0.45	0.31	0.32
	ccga(C/T)gaca	—/T	Homo	rs1799832	Asp119	CAG28583.1	0.07	0	0.03

[a]No genetic alterations were identified in the fifth gene sequenced; *NTHL1*.
[b]Part of this sequence is intronic.
[c]Amino acid changes subjected for functional studies.
[d]Name of an alternative transcript.

a *mutY* deficient *E.coli* strain to avoid contamination of endogenous MutY. Despite high expression levels, most hMYH was insoluble (data not shown) and we were able to obtain only partially purified hMYH proteins after nickel–agarose purification (supplementary Figure 1 is available at *Carcinogenesis* Online). Using oligonucleotide substrates with the A:8oxoG at a defined position, hMYH DNA glycosylase activity can be monitored by running denaturing polyacrylamide gels, where the cleaved product (C) after NaOH treatment will migrate faster than the intact substrate (I; Figure 2A). DNA glycosylase activity assays revealed that the mutant enzymes hMYH H434D and hMYH S501F removed adenines from duplex oligonucleotides containing A:8oxoG base pairs at rates comparable with wild-type (WT) hMYH (Figure 2A). Contrary, for a substrate containing an A:G mispair, hMYH H434D activity was severely reduced (4% cleavage as compared with 36% for WT hMYH with 18 ng enzyme), whereas the S501F mutant showed the same A:G DNA glycosylase activity as intact hMYH (Figure 2A). The mutant form hMYH R260Q displayed reduced ability to excise adenines from an A:8oxoG duplex, cleaving only 15% of the substrate as compared with complete removal by WT hMYH with 18 ng enzyme (Figure 2A). By increasing the amount of hMYH R260Q up to 240 ng, 40% cleavage of the A:8oxoG duplex was obtained (Figure 2B). For A:G-containing substrates, R260Q was almost inactive (1% compared with 36% cleavage by WT hMYH with 18 ng enzyme) (Figure 2A). Despite its reduced DNA glycosylase activity, EMSAs revealed that R260Q has retained the ability to bind A:8xoG-containing duplexes and for the A:G substrate R260Q binds even stronger than the WT hMYH (Figure 2C). It appears that the hMYH R260Q substitution results in loss of the ability to discriminate between A:8oxoG and A:G when binding to DNA. The mutant hMYH S501F binds the two substrates with similar affinities as the WT enzyme, whereas no shift was observed with the H434D mutant neither for A:8oxoG nor for A:G substrates (Figure 2C). With higher protein concentration, H434D could bind weakly to the A:G substrate but not to A:8oxoG (data not shown). Under conditions for EMSA, no cleavage of the DNA substrates was observed as assayed by running a small sample of the reaction mixture on a denaturing/

urea gel (data not shown). In summary, two of the *hMYH* mutations identified (R260Q and H434D) lead to proteins with defect DNA glycosylase activity.

Analysis of hOGG1

Sequencing of *hOGG1* revealed five novel mutations: one in the control group (G308E) and four in PSC patients (S31P, V58 synonymous, G342 synonymous and T398S; Table I). As two of the mutations found in the PSC group are silent mutations and one is conservative (T398S), we proceeded with functional studies of the hOGG1 S31P mutation only (Figure 1B). hOGG1 S31P was expressed in *E.coli*, purified (supplementary Figure 1 is available at *Carcinogenesis* Online) and subjected to 8oxoG DNA glycosylase analysis. hOGG1 S31P showed the same capacity as WT hOGG1 to remove 8oxoG from an 8oxoG:C DNA duplex (Figure 3A). EMSA reveals that WT and S31P bind with the same affinity to an 8oxoG:C DNA oligonucleotide (Figure 3B). The alteration S31P does not seem to affect the activity nor DNA binding capacity of hOGG1.

Analysis of NEIL1

In *NEIL1*, three novel mutations were found; one in the PSC group (E181K), one in the PSC-CCA group (H275 synonymous) and one in the healthy control group (R339W). E181K and the SNP G83D were subjected for functional studies. Based on a comparison with the *E.coli* endonuclease VIII structure in complex with DNA (34), glycine 83 is found in a loop that forms part of the groove of NEIL1 in which the DNA is expected to bind, whereas glutamate 181 is a conserved residue within the helix-2-turn-helix motif that interacts with the lesion-containing strand (Figure 1C). The NEIL1 E181K protein was produced in *E.coli*, but the protein formed inclusion bodies and we were not able to obtain a pure preparation. NEIL1 G83D was purified to apparent homogeneity (supplementary Figure 1 is available at *Carcinogenesis* Online), tested for co-purifying contaminating activities (supplementary Figure 2 is available at *Carcinogenesis* Online) and assayed toward several oxidative DNA lesions known to be removed by NEIL1. Because NEIL1 has been shown to
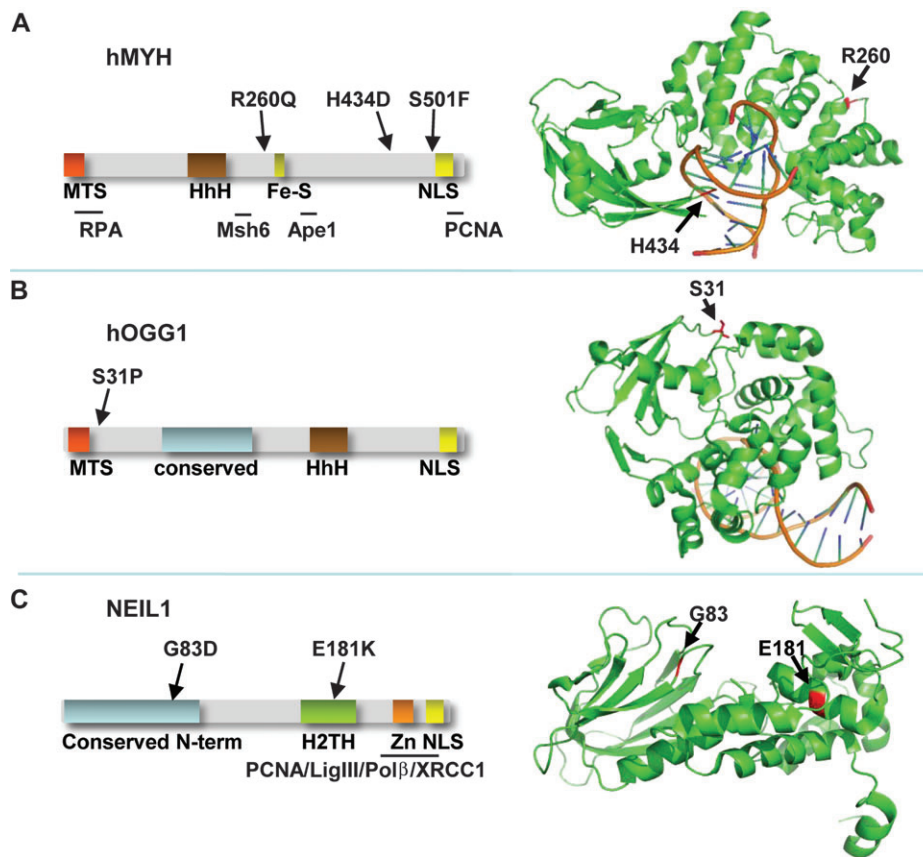
Fig. 1. Schematic diagram and 3D structures of hMYH, hOGG1 and NEIL1. (**A**) Schematic diagram of the hMYH showing functional domains (left) and 3D structure (right) of MutY from *Bacillus stearothermophilus* in complex with DNA [Protein Data Bank identifier 1RRQ] (31). The positions of the mutations studied in this work, R260 and H434 are shown and were identified by amino acid sequence alignment of hMYH with BsMutY. Serine501 is outside of the ordered structure. (**B**) Schematic diagram showing functional domains (left) and 3D structure (right) of hOGG1 in complex with 8oxoG-containing DNA (Protein Data Bank identifier 1EBM) (32). The location of S31P mutation is shown. (**C**) Schematic diagram of NEIL1 showing functional domains (left) and 3D structure (right) (Protein Data Bank identifier 1TDH) (33) with position of G83 and E181 indicated. For hMYH domains interacting with RPA, MSH6, APE1 and PCNA are indicated. hOGG1 and NEIL1 have a conserved domain (blue) and the NEIL1 C-terminus is necessary for interactions with PCNA, DNA ligase III, DNA polymerase β and XRCC1. The figures are not drawn to scale. Fe-S, Iron-sulfur cluster; H2TH, helix-2-turn-helix motif; HhH, helix-hairpin-helix motif; MTS, mitochondrial targeting signal; NLS, nuclear localization signal; Zn, zinc finger domain.

process β,δ-elimination activity (30,35), its activity assay does not need NaOH treatment to detect strand breaks as for hMYH and hOGG1. In contrast to the WT enzyme, NEIL1 G83D showed no activity on 8oxoG:C duplex DNA and had strongly impaired cleavage activity on DHT duplex DNA (Figure 4A). The residual mutant activity on DHT:A was <10% of WT activity. WT NEIL1 is shown previously not to possess 8oxoG (36) or DHT DNA glycosylase activity on single-stranded DNA (ssDNA; our unpublished observation). As expected, the G83D mutant protein showed no activity on ssDNA containing a single 8oxoG or DHT (data not shown). Another oxidative pyrimidine lesion cleaved by NEIL1 is Tg. WT Neil1 has higher activity for Tg in double-stranded DNA than in single stranded (Figure 4A). The mutant G83D had weak activity for Tg in double-stranded DNA and almost no cleavage was observed for the single-stranded Tg substrate. Moreover, removal of formamidopyrimidine by G83D NEIL1 was strongly impaired as compared with the WT enzyme (Figure 4B). These data demonstrate an important role of Gly83 in substrate recognition.

Next, we tested NEIL1 G83D activity toward other oxidative pyrimidine lesions such as 5ohC and 5ohU. In contrast to the data above, NEIL1 G83D cleaved 5ohC:G (Figure 4A) and 5ohU:G (data not shown) duplex DNA at the same efficiency as the WT enzyme (Figure 4A). Also for 5ohC and 5ohU on ssDNA, G83D showed almost the same cleavage activity as the WT enzyme (Figure 4A). However, NEIL1 G83D showed significantly reduced δ-elimination activity

(50%) as compared with the WT enzyme for both single-stranded substrates. With high concentration of G83D and 5ohC and 5ohU ssDNA substrates, there appears to be some 5′-phosphtase or ssDNA nuclease contamination in the reactions. As this is not observed for the Tg single-stranded substrate, the possible contaminating activity might be sequence dependent. To further characterize NEIL1 G83D, EMSAs were performed. The results reveal that G83D binds to 5ohC:G DNA substrate with slightly reduced affinity as compared with WT NEIL1 (Figure 4C). In the lane with the highest concentration of WT NEIL1 (100 ng), there is a 25% reduction in total signal intensity that might be ascribed to copurifying contaminating 5′-phosphatase/exonuclease activity.

To further investigate the role of the G83D mutation on the β,δ-reaction, we examined cleavage of a duplex oligonucleotide containing a single abasic site. Surprisingly, we find that the mutant enzyme incise the abasic site more efficiently (two times) than the WT (Figure 4A). However, only 50% of the cleaved abasic site has been fully processed to the δ-elimination end product by NEIL1 G83D, whereas 50% of the cleaved product remains as the β-elimination intermediate. With WT NEIL1 almost 100% of the cleavage product has been fully processed to the δ-elimination product. It thus appears that the NEIL1 G83D mutation modulates the catalytic capacity of NEIL1 as well as substrate specificity.

Intracellular localization studies have revealed that NEIL1 is located to the nucleus of human cells with an accumulation in the
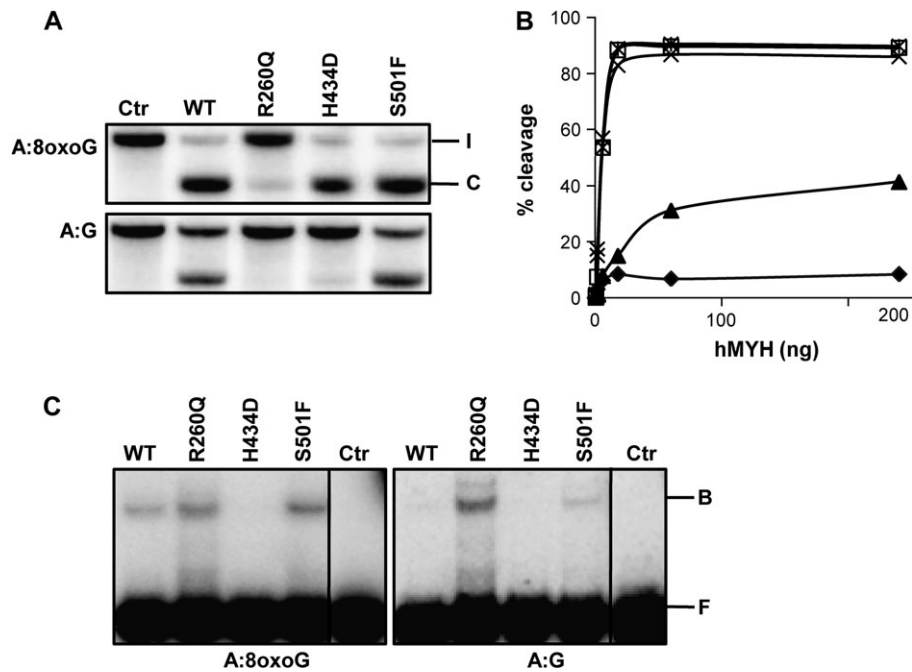
Fig. 2. Analysis of hMYH variants. (**A**) Adenine DNA glycosylase activities of hMYH WT, R260Q, H434D and S501F variants were measured by incubating the respective proteins (18 ng) with a duplex oligodeoxyribonucleotide containing a single A:8oxoG or A:G basepair at 37°C for 30 min. Strand cleavage after NaOH treatment was analyzed by 20% polyacrylamide gel electrophoresis and phosphorimaging (I = intact strand and C = cleavage product). (**B**) Different amounts (0.6–240 ng) of hMYH WT (□), R260Q (▲), H434D (X) and S501F (∗) were assayed for A:8oxoG DNA glycosylase activities and percentage strand cleavage quantified with ImageQuant. Extract from *Escherichia coli* cells expressing empty vector and purified similarly as hMYH was used to measure the background level (◆). (**C**) DNA binding properties of hMYH WT, R260Q, H434D and S501F (24 ng) to substrates containing A:8oxoG (left panel) or A:G (right panel). After incubation on ice, DNA–protein complexes (B = bound substrate) were separated from free DNA (F) by 10% native polyacrylamide gel electrophoresis. Control lanes were without addition of protein.
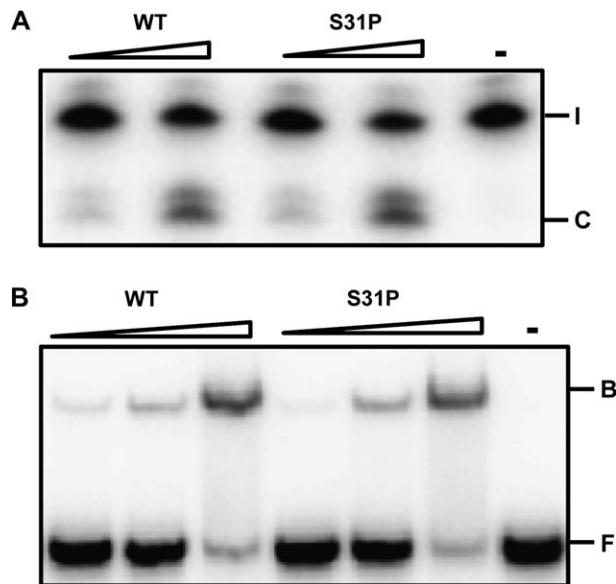

Fig. 3. Analysis of hOGG1 variants. (**A**) 8oxoG DNA glycosylase activity of S31P compared with WT hOGG1. A total of 3 and 10 ng enzymes were incubated with an 8oxoG:C oligonucleotide at 37°C for 30 min before cleavage of the phosphodiester backbone by NaOH. The reaction products were separated by 20% polyacrylamide gel electrophoresis and visualized by phosphorimaging. (I = intact strand and C = cleavage product). (**B**) DNA binding properties of hOGG1 WT and S31P. WT and S31P hOGG1 (10, 30 and 100 ng) were incubated with 8oxoG:C DNA on ice and DNA–protein complexes (B = bound substrate) were separated from free DNA (F) by 10% native polyacrylamide gel electrophoresis. Control lanes were without addition of protein.

nucleoli and a putative nuclear localization signal is found in the C-terminus of NEIL1 (Figure 1C) (30). Constructs of NEIL1 G83D and E181K fused to enhanced green fluorescent protein were generated and used for transfection of HeLa S3 cells. Fluorescence microscopy showed that the distribution of the NEIL1 mutants was confined to the nucleus as seen for WT NEIL1 (Figure 4D). Although we were not able to purify recombinant NEIL1 E181K expressed in *E.coli*, the corresponding fusion protein localized as WT NEIL1 in HeLa cells indicating that the mutant is folded and distributed correctly in mammalian cells.

Discussion

PSC is a chronic inflammatory disorder of the bile ducts in which the patients have a high risk for developing CCA. In order to better understand individual susceptibility to CCA, we evaluated the role of SNPs and germ line mutations in genes involved in the repair of oxidative DNA damage. Eighteen genetic variants were identified, six were validated SNPs, whereas 12 were new mutations. Three rare variants identified in PSC and PSC-CCA patients were found to critically affect protein function (hMYH R260Q, hMYH H434D and NEIL1 G83D).

Importantly, there is increasing awareness of the contribution from rare variants to the genetic susceptibility of complex traits (37). The detection of such variants is inert to the association study design and requires resequencing of entire study populations with subsequent assessment of functional consequences of identified variants (reviewed in ref. 38). Noteworthy, the three variants with severe consequence for the BER pathway were found among patients with PSC with or without CCA. The lack of overall association between any of the investigated variants and PSC or CCA susceptibility may clearly result from lack of statistical power. For instance, it is well established that the oxidative DNA lesion 8-oxoguanine is mutagenic and that hOGG1 represents the major 8oxoG removing activity in human cells. hOGG1 S326C is the most studied SNP in *hOGG1* and is weakly

1151
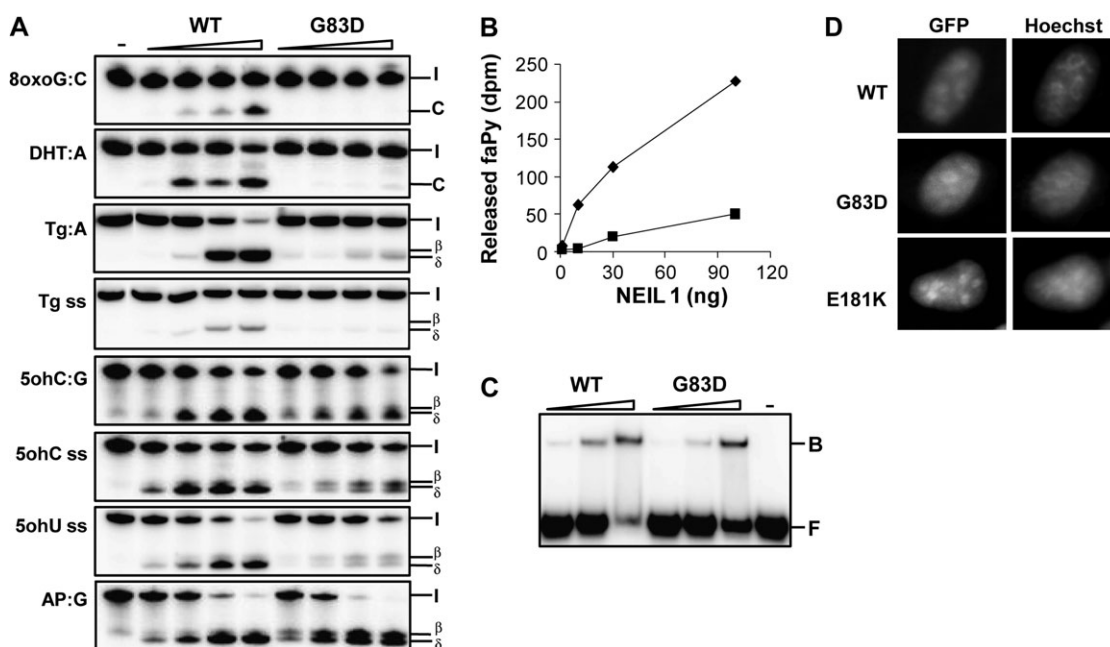
Fig. 4. Analysis of NEIL1 variants. (**A**) DNA glycosylase activity of G83D compared with WT NEIL1. Enzymes (2, 5, 10 and 20 ng) were incubated with different oligonucleotide substrates as indicated at 37°C for 30 min. The reaction products were separated by 20% polyacrylamide gel electrophoresis and visualized by phosphorimaging. (I = intact strand, C = cleavage product, β = β elimination, δ = δ elimination cleavage, ss = single strand). (**B**) FaPy DNA glycosylase activity of NEIL1 WT (◆) and G83D (■). Enzymes (3, 10, 30 and 100 ng) were assayed for removal of faPy from [^{3}H]-methyl-faPy-poly(dG·dC). (**C**) DNA binding properties of NEIL1 WT and G83D. NEIL1 WT and G83D (20, 50 and 100 ng) were incubated with 5ohC:G DNA on ice and DNA–protein complexes (B = bound substrate) were separated from free DNA (F) by 10% native polyacrylamide gel electrophoresis. Control lanes were without addition of protein. (**D**) Nuclear localization of NEIL1 G83D and E181K. Asynchronous growing HeLa S3 cells were transiently transfected with constructs expressing NEIL1-EGFP, NEIL1G83D-EGFP or NEIL1E181K-EGFP. Cells were imaged directly by fluorescence microscopy for EGFP detection. DNA was stained with Hoechst 33342.

associated with some cancers (14) and has reduced 8oxoG DNA glycosylase activity (39). Power calculations based on previous studies of this SNP [odds ratio of GG versus CC genotypes >2.0; (14)] emphasized limitations of the genetic association study design (0.11 for PSC-CCA versus PSC without CCA, 0.14 for PSC-CCA versus healthy controls; http://medipe.psu.ac.th/episoft/pssamplesize) to conclusively clarify whether this mutation may play a role in cholangiocarcinogenesis in PSC. Although comparable frequencies of S326C were observed in all three study populations, the possibility thus needs to be kept open that weak effects on CCA susceptibility from this variant may exist.

Two common *hMYH* mutations associated with colorectal adenoma polyposis are Y165C and G382D, which account for ~73% of all *hMYH* mutations reported to date (reviewed in ref. 40). Neither of these two mutations was found in our study. To date >80 different *hMYH* mutations have been described in MutY associated-polyposis patients (reviewed in ref. 41). Approximately 15% of these mutations have been subjected to functional studies and several of the mutants show reduced or no DNA glycosylase activity (41,42). The catalytically impaired hMYH variant R260Q was found in one patient carrying this mutation on one allele. The importance of the mutation for this patient is uncertain, as *hMYH* mutations associated with colon cancer are autosomal recessive. However, R260Q still binds DNA and may prevent access for WT hMYH to A:8oxoG pairs. The same mutation was also identified in a large multisite, population-based colorectal cancer case–control study (42,43) and was further analyzed by Ali *et al.* (42). Similarly as in this work, reduced A:8oxoG activity was found for R260Q. Our data show that R260Q has normal A:8oxoG binding, however, Ali *et al.* reported that the mutant has reduced A:8oxoG DNA-binding property. We do not know the reason for the discrepancy but different assay conditions could contribute. The catalytic impairment of hMYH R260Q was unexpected from a structural point of view as R260 is located on the surface of MutY pointing away

from the active site pocket (Figure 1A). From the structure, this residue is predicted to have no role in catalysis or in DNA binding as the distance from R260 to the adenine-binding pocket is >16 Å. Future screening for *hMYH* associated mutations in cancer should also include R260Q.

The C-terminal domain of MutY is required for 8oxoG recognition and deletion of this domain results in a protein with reduced binding affinity and DNA glycosylase activity for A:8oxoG but not for A:G (44,45). The H434 residue is a conserved residue within the C-terminal domain and is part of a loop that intercalates the DNA and makes contacts with the sugar-phosphate backbone 5′ to the oxidized guanine. The presence of aspartate in this position will result in a negative charge and repulsion of the DNA, as demonstrated by the lack of shift in the EMSAs for hMYH H434D. Interestingly, the H434D mutant has the same DNA glycosylase activity for A:8oxoG as WT hMYH; however, the ability to excise adenine from A:G mispairs is significantly reduced, supporting a previous observation that the C-terminal domain is not only involved in 8oxoG recognition but also affects binding and catalytic activities toward A:G mismatches (46).

The NEIL1 G83D mutant has been shown previously to lack DNA glycosylase activity toward Tg and faPy lesion but has β-elimination activity on AP DNA (47). Presumably, the gain of the acidic aspartate alters the surface of the active site of NEIL1 that is dominated by basic amino acid side chains in the WT enzyme. Here, we show that the DNA glycosylase activity for G83D is lost only for 8oxoG, DHT and Tg, substrates that all need to be in a context of double-stranded DNA for NEIL1 to cleave. For lesions known to be substrates for NEIL1 also when present in ssDNA (5ohC, 5ohU), G83D is as active as WT NEIL1. In addition to compromised DNA glycosylase activity for G83D for some substrates, the δ-elimination reaction is also affected. After incubation with G83D, about half of the substrate was cleaved by β,δ-elimination and the other half by β-elimination only.

This was most evident for the single-stranded substrates (5ohC, 5ohU) and for the AP DNA. It thus appears that introduction of an aspartic residue at this position directly or indirectly affects residues involved in δ-elimination. Roy *et al.* (47) reported that G83D yielded primarily β-elimination only. We do not know the reason for this discrepancy but might be due to different assay conditions, variation during expression and purification or the sequence context of the DNA substrate.

Residue G83 is part of a loop connecting β-strands 4 and 5, which in the *E.coli* EndoVIII structure includes the essential DNA-intercalating residue Leu70 (corresponding to Met81 in NEIL1). Modeling of aspartate in position 83 of NEIL1 suggests that this mutation alter the conformation of this loop, as the negatively charged aspartate side chain is pointing into a hydrophobic pocket in NEIL1. Significant rearrangement of this loop is expected in order to remove unfavorable steric and electrostatic conflicts following a mutation of Gly83 into Asp. Interestingly, according to our model, this loop is close to the non-lesion-containing strand, supporting the observed effect on double-stranded substrates only.

The NEIL1 G83D SNP was found in two PSC patients with CCA and both were heterozygotic for this polymorphism. As the other allele is healthy, the cells are probably to function with approximately half the level of normal NEIL1 activity. In tissue with severe oxidative stress, such as the bile ducts of PSC patients, a reduction in the DNA repair functions could lead to high levels of oxidative DNA damage and consequently, an increase in the incidence of gene mutations might be the result. Further, defect in one allele could be fatal if the other allele acquires mutations during initial stages of carcinogenesis. It has been estimated that a 20–35% reduction in repair capacity of cells is associated with a 4- to 6-fold higher risk of developing cancer (48).

An expression study has shown previously that the *NEIL1* messenger RNA level was reduced in about half of primary gastric cancer specimens examined indicating a role of NEIL1 in preventing cancer development (12). Variations in expression levels of *NEIL1* and also of the other DNA repair genes analyzed may also be of significance for CCA in PSC patient. However, the infiltrating nature of CCA tumors makes it difficult to obtain pure tumor material and such a study awaits further work.

In conclusion, by sequencing a panel of genes encoding proteins of the BER pathway in patients with PSC, PSC and CCA and healthy controls, we detected nine novel non-synonymous mutations in *hMYH*, *hOGG1* and *NEIL1*. Although no statistical significant differences in allele frequencies between any of the investigated groups were found, severe impairment of protein function was evident in extensive functional analysis of three mutations found in patients with PSC and PSC with CCA. In statistical terms, a causal relationship between either of these mutations and carcinogenesis and PSC is difficult to conclude. Nevertheless, our present findings add to increasing knowledge of the importance of rare, coding variants in complex diseases (37,49) and hint to the usefulness of massive sequencing in the search for genetic variants that influence the individual susceptibility to cancer development.

Supplementary material

Supplementary Tables 1 and 2 and Figures 1 and 2 can be found at http://carcin.oxfordjournals.org/

Funding

Acknowledgements

Conflict of Interest Statement: None declared.

References

1. Ames,B.N. *et al.* (1993) Oxidant are major contributor to cancer and aging. In: Halliwell,B. and Aruoma,O. (eds) *DNA and Free Radicals*. Ellis Horwood, New York, NY, pp. 1–18.
2. Wiseman,H. *et al.* (1996) Damage to DNA by reactive oxygen and nitrogen species: role in inflammatory disease and progression to cancer. *Biochem. J.*, **313**, 17–29.
3. Halliwell,B. *et al.* (1989) *Free Radicals in Biology and Medicine*. Oxford University Press, New York, NY.
4. Tchou,J. *et al.* (1993) Repair of DNA containing the oxidatively-damaged base, 8-oxoguanine. *Mutat. Res.*, **299**, 277–287.
5. Michaels,M.L. *et al.* (1992) The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). *J. Bacteriol.*, **174**, 6321–6325.
6. Olinski,R. *et al.* (1992) DNA base modifications in chromatin of human cancerous tissues. *FEBS Lett.*, **309**, 193–198.
7. Al-Tassan,N. *et al.* (2002) Inherited variants of MYH associated with somatic G:C—>T:A mutations in colorectal tumors. *Nat. Genet.*, **30**, 227–232.
8. Sieber,O.M. *et al.* (2003) Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in MYH. *N. Engl. J. Med.*, **348**, 791–799.
9. Slupphaug,G. *et al.* (2003) The interacting pathways for prevention and repair of oxidative DNA damage. *Mutat. Res.*, **531**, 231–251.
10. David,S.S. *et al.* (2007) Base-excision repair of oxidative DNA damage. *Nature*, **447**, 941–950.
11. Hazra,T.K. *et al.* (2007) Oxidative DNA damage repair in mammalian cells: a new perspective. *DNA Rep.*, **6**, 470–480.
12. Shinmura,K. *et al.* (2004) Inactivating mutations of the human base excision repair gene NEIL1 in gastric cancer. *Carcinogenesis*, **25**, 2311–2317.
13. Xu,J. *et al.* (2002) Associations between hOGG1 sequence variants and prostate cancer susceptibility. *Cancer Res.*, **62**, 2253–2257.
14. Goode,E.L. *et al.* (2002) Polymorphisms in DNA repair genes and associations with cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, **11**, 1513–1530.
15. Jones,S. *et al.* (2002) Biallelic germline mutations in MYH predispose to multiple colorectal adenoma and somatic G:C—>T:A mutations. *Hum. Mol. Genet.*, **11**, 2961–2967.
16. Paz-Elizur,T. *et al.* (2005) Repair of the oxidative DNA damage 8-oxoguanine as a biomarker for lung cancer risk. *Cancer Biomark.*, **1**, 201–205.
17. Paz-Elizur,T. *et al.* (2006) Reduced repair of the oxidative 8-oxoguanine DNA damage and risk of head and neck cancer. *Cancer Res.*, **66**, 11683–11689.
18. Chapman,R.W. *et al.* (1980) Primary sclerosing cholangitis: a review of its clinical features, cholangiography, and hepatic histology. *Gut*, **21**, 870–877.
19. Schrumpf,E. *et al.* (1994) Risk factors in primary sclerosing cholangitis. *J. Hepatol.*, **21**, 1061–1066.
20. Bergquist,A. *et al.* (2002) Hepatic and extrahepatic malignancies in primary sclerosing cholangitis. *J. Hepatol*, **36**, 321–327.
21. Wiesner,R.H. *et al.* (1989) Primary sclerosing cholangitis: natural history, prognostic factors and survival analysis. *Hepatology*, **10**, 430–436.
22. Chalasani,N. *et al.* (2000) Cholangiocarcinoma in patients with primary sclerosing cholangitis: a multicenter case-control study. *Hepatology*, **31**, 7–11.
23. Burak,K. *et al.* (2004) Incidence and risk factors for cholangiocarcinoma in primary sclerosing cholangitis. *Am. J. Gastroenterol.*, **99**, 523–526.
24. Melum,E. *et al.* (2008) Cholangiocarcinoma in primary sclerosing cholangitis is associated with NKG2D polymorphisms. *Hepatology*, **47**, 90–96.
25. Komichi,D. *et al.* (2005) Glycochenodeoxycholate plays a carcinogenic role in immortalized mouse cholangiocytes via oxidative DNA damage. *Free Radic. Biol. Med.*, **39**, 1418–1427.
26. Barker,D.L. *et al.* (2004) Two methods of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage panel. *Genome Res.*, **14**, 901–907.
27. Bai,H. *et al.* (2007) Functional characterization of human MutY homolog (hMYH) missense mutation (R231L) that is linked with hMYH-associated polyposis. *Cancer Lett.*, **250**, 74–81.
28. Bjoras,M. *et al.* (1997) Opposite base-dependent reactions of a human base excision repair enzyme on DNA containing 7,8-dihydro-8-oxoguanine and abasic sites. *EMBO J.*, **16**, 6314–6322.

29. Eide,L. *et al.* (1996) Base excision of oxidative purine and pyrimidine DNA damage in *Saccharomyces cerevisiae* by a DNA glycosylase with sequence similarity to endonuclease III from *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **93**, 10735–10740.

30. Morland,I. *et al.* (2002) Human DNA glycosylases of the bacterial Fpg/MutM superfamily: an alternative pathway for the repair of 8-oxoguanine and other oxidation products in DNA. *Nucleic Acids Res.*, **30**, 4926–4936.

31. Fromme,J.C. *et al.* (2004) Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. *Nature*, **427**, 652–656.

32. Bruner,S.D. *et al.* (2000) Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature*, **403**, 859–866.

33. Doublie,S. *et al.* (2004) The crystal structure of human endonuclease VIII-like 1 (NEIL1) reveals a zincless finger motif required for glycosylase activity. *Proc. Natl Acad. Sci. USA*, **101**, 10284–10289.

34. Zharkov,D.O. *et al.* (2002) Structural analysis of an *Escherichia coli* endonuclease VIII covalent reaction intermediate. *EMBO J.*, **21**, 789–800.

35. Hazra,T.K. *et al.* (2002) Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. *Proc. Natl Acad. Sci. USA*, **99**, 3523–3528.

36. Dou,H. *et al.* (2003) Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2. *J. Biol. Chem.*, **278**, 49679–49684.

37. Lesage,S. *et al.* (2002) CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. *Am. J. Hum. Genet.*, **70**, 845–857.

38. McCarthy,M.I. *et al.* (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.*, **9**, 356–369.

39. Vodicka,P. *et al.* (2007) Association of DNA repair polymorphisms with DNA repair functional outcomes in healthy human subjects. *Carcinogenesis*, **28**, 657–664.

40. Sampson,J.R. *et al.* (2005) MutYH (MYH) and colorectal cancer. *Biochem. Soc. Trans.*, **33**, 679–683.

41. Cheadle,J.P. *et al.* (2007) MUTYH-associated polyposis—from defect in base excision repair to clinical genetic testing. *DNA Repair*, **6**, 274–279.

42. Ali,M. *et al.* (2008) Characterization of mutant MUTYH proteins associated with familial colorectal cancer. *Gastroenterology*, **135**, 499–507.

43. Cleary,S.P. *et al.* (2009) Germline MutY human homologue mutations and colorectal cancer: a multisite case-control study. *Gastroenterology*, **136**, 1251–1260.

44. Noll,D.M. *et al.* (1999) The C-terminal domain of the adenine-DNA glycosylase MutY confers specificity for 8-oxoguanine.adenine mispairs and may have evolved from MutT, an 8-oxo-dGTPase. *Biochemistry*, **38**, 6374–6379.

45. Li,X. *et al.* (2000) The C-terminal domain of MutY glycosylase determines the 7,8-dihydro-8-oxo-guanine specificity and is crucial for mutation avoidance. *J. Biol. Chem.*, **275**, 8448–8455.

46. Li,L. *et al.* (2003) The C-terminal domain of *Escherichia coli* MutY is involved in DNA binding and glycosylase activities. *Nucleic Acids Res.*, **31**, 3038–3049.

47. Roy,L.M. *et al.* (2007) Human polymorphic variants of the NEIL1 DNA glycosylase. *J. Biol. Chem.*, **282**, 15790–15798.

48. Mohrenweiser,H.W. *et al.* (1998) Variation in DNA repair is a factor in cancer susceptibility: a paradigm for the promises and perils of individual and population risk estimation? *Mutat. Res.*, **400**, 15–24.

49. Walsh,T. *et al.* (2008) Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*, **320**, 539–543.