

Identification of 127 Amino Acid Substitution Variants in Screening 37 DNA Repair Genes in Humans¹

Harvey W. Mohrenweiser,² Tong Xi,
Johana Vázquez-Matías, and Irene M. Jones

Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, California 94550

Abstract

The repair of damaged DNA requires the function of multiple proteins in generally damage-specific, nonredundant pathways. The relationship of DNA repair to cancer susceptibility is obvious in “cancer families,” in which low frequency, high penetrance, loss-of-function variant alleles of genes with roles in the repair of damaged DNA have been associated with a high risk of disease. More important for the cancer incidence in the general population, many individuals exhibit reduced (60–75% of normal) repair capacity phenotypes that have been associated with several-fold increases in individual cancer risk. In a program to identify the molecular basis for the variation in repair capacity and the elevated cancer susceptibility, we have identified 127 amino acid substitution variants in resequencing 37 DNA repair genes in 36–164 unrelated individuals. Over 50% of the substitutions are exchanges of amino acid residues with dissimilar physical or chemical properties, at sites at which the common residue is identical in the human and mouse proteins. Five additional sequence changes resulting in proteins with altered termination of translation and one amino acid insertion variant were detected. The variant allele frequencies average 0.047, with individual variant allele frequencies ranging from <0.01 to 0.43. Homozygous variant individuals and individuals with multiple amino acid substitutions in a gene were observed. Most individuals exhibited variation in multiple genes in a repair pathway. Ten variant alleles accounted for 52% of the genetic variation among individuals, but a striking 23% of the total variation is associated with 108 variants with allele frequencies of less than 5%. Screening generally healthy individuals generates a catalogue of common variants that is a

resource for molecular epidemiology studies endeavoring to use a genotype to phenotype paradigm to estimate the role of genetic variation and individual susceptibility in disease risk from environmental and lifestyle exposures in the general population of the United States.

Introduction

Exposure of cells to environmental agents as well as the by-products of cellular metabolism results in extensive damage to DNA. The pattern of DNA damage is often complex but has characteristics associated with the damaging agent. Organisms have several, generally nonredundant, pathways for repairing different DNA lesions, *e.g.*, strand breaks, adducts, and oxidized bases, resulting from these exposures. The NER³ pathway removes UV-induced pyrimidine dimers and bulky DNA adducts associated with chemical exposures (1, 2). The DSB/RR pathway repairs strand breaks that are often associated with exposure to ionizing radiation and radiometric drugs or the result of incomplete repair of other damage (3–5). A third pathway, BER, directly processes damaged bases, such as oxidized bases, the most common spontaneous damage (6, 7). These lesions also result from exposure of cells to ionizing or UV radiation via elevations in the intracellular reactive oxygen levels (8). A fourth pathway, MMR, repairs replication errors (9, 10). As recently summarized, DNA repair in mammalian cells involves more than 80 genes with direct roles in the repair of DNA damage (11, 12). At least 40 additional genes with roles in DNA damage recognition and cell cycle checkpoint processes have important, although more indirect, roles in the repair of DNA damage (13, 14).

The important role of DNA repair in the maintenance of a normal cellular genotype and a cancer-free state is most obvious in “cancer families,” in which the presence of rare but highly penetrant variant alleles at a number of loci is associated with a high risk of cancer. A classic example is xeroderma pigmentosum, a prototype cancer gene syndrome associated with the development of UV-induced skin cancers resulting from the loss of function of a gene of the NER pathway (15). The association of defects in MMR with colon cancer is another example of the critical role of DNA repair in cancer prevention (9, 10). Other genes with direct or indirect roles in DNA repair and in which variant alleles are associated with elevated cancer risk, include *BRCA1*, *BRCA2*, *TP53*, *ATM*, and *NBS1* (16). Disruption of the function of genes with roles in the repair of damaged DNA is associated with increased sensitivity to DNA damaging agents and cancer proneness (17). These instances of inherited cancer predisposition have provided important models

Received 2/15/02; revised 6/28/02; accepted 6/28/02.

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.

¹ Work performed under the auspices of the United States Department of Energy by the University of California, Lawrence Livermore National Laboratory (contract W-7405-ENG-48) and supported in part by Interagency Agreement Y1-ES-8054-05 from National Institute of Environmental Health Sciences and National Cancer Institute Grant 1 U-1 CA 83180-03.

² To whom requests for reprints should be addressed, at Biology and Biotechnology Research Program, L-448, Lawrence Livermore National Laboratory, P. O. Box 0808, 7000 East Avenue, Livermore, CA 94550. Phone: (925) 423-0534; Fax: (925) 424-3130; E-mail: mohrenweiser1@llnl.gov.

³ The abbreviations used are: NER, nucleotide excision repair; DSB/RR, double-strand break/recombination repair; BER, base excision repair; MMR, mismatch repair; LLNL-IRB, Lawrence Livermore National Laboratory Institutional Review Board; UTR, untranslated region; UV, ultraviolet.

for increasing the understanding of DNA repair pathways and carcinogenic processes and the relationship of DNA repair to cancer risk. In terms of cancer incidence, these generally highly penetrant, low-activity disease-associated alleles are estimated to account for no more than 5% of the cancer cases in the general population. The remaining cancer cases or sporadic cancers are suggested to occur in individuals with combinations of common polymorphisms with often low penetrance, only marginally altered function, and weak effects giving rise to individuals with increased susceptibility to disease from environmental exposures and lifestyle factors.

Reduced DNA repair capacity is a polymorphic phenotypic trait. At least 10% of the individuals in the population have a capacity to repair DNA damage after *in vitro* exposure of lymphocytes to DNA-damaging agents, which is only 60–75% of the population mean (18). The reduced-repair-capacity phenotypes for damage induced by bleomycin, γ radiation, and benzo[*a*]pyrene-diolepoxide, classes of damage expected to be repaired by different pathways and, thus, different sets of genes, behave as independent traits (19). The genetic contribution to the interindividual differences in repair capacity ranges from 0.65 to 0.80 for these three traits, the traits being the ability to repair DNA damage induced by bleomycin, γ radiation, or benzo[*a*]pyrene-diolepoxide (20–22). A link between the reduced-DNA-repair-capacity phenotypes and cancer susceptibility is supported by data from a series of epidemiology studies. These studies have demonstrated that a reduced-repair-capacity phenotype is associated with an increased risk (odds ratios of 2–10) of developing tumors at several sites, including breast, lung, skin, liver, or head/neck [see review of Berwick and Vineis (23) and references cited therein]. A limited number of individuals have reduced capacity to repair damage induced by both bleomycin and benzo[*a*]pyrene-diolepoxide, damage expected to be repaired by different pathways. The number of individuals is consistent with the number expected, given the incidence of individuals with reduced capacity phenotypes for each agent. These individuals with reduced capacity in two pathways exhibit a higher risk of developing lung cancer than do individuals with a reduced capacity in only one pathway (24).

Accumulating evidence suggests that many diseases, including most cancer cases, result from low-level exposures and lifestyle factors in genetically susceptible individuals. The results presented here focus on identification of the common or polymorphic amino acid substitution variants existing in DNA repair genes in the general population. This extends the work of Shen *et al.* (25) to more genes in a larger number of individuals. This catalogue of variant alleles is a resource for molecular epidemiology studies using a genotype to phenotype or health consequence paradigm for associating DNA repair gene variants to repair capacity and cancer susceptibility.

Materials and Methods

The resequencing strategy, which involves sequencing of the same genomic region in multiple individuals to identify DNA sequence variation in DNA repair genes, has been described previously (25). It involves the direct sequencing of PCR products containing the individual exons of a gene plus adjacent intronic and noncoding regions. The PCR products include the splice sites and the 5' and 3' regions of the genes.

PCR Amplification of Exons. The PCR primers are designed using the Oligo Primer Analysis software (National Biosciences, Inc., Plymouth, MN). Appended to the 5' end of each PCR primer is the primer binding site for the forward or reverse

energy transfer (ET) DNA sequencing primer (Amersham Life Science, Inc., Cleveland, OH). PCR primers are matched so that the sense and the antisense PCR primers contain different sequencing primer binding sites. PCR primers are optimized as necessary by addition of DMSO. Primers were obtained from Sigma Genesys (The Woodlands, TX).

PCR primers are positioned so that amplification of the genomic sequence is initiated at least 75 nucleotides from the intron-exon boundary. This is sufficient distance for high-quality sequence data to be obtained before reaching the intron/exon splice site. The PCR products are ~400 bp (range, 300–450 bp), and, therefore, the entire fragment can be sequenced in both directions without developing new sequencing primers.

The GenBank accession numbers for the genomic sequences used for primer design are listed in Table 1. The locus designations are from the HUGO Gene Nomenclature Committee guidelines.⁴ Common aliases are included for a number of the genes.

DNA Sequencing. After PCR amplification, PCR products are diluted and used as substrate in sequencing reactions. Dye primer cycle sequencing reactions are performed according to manufacturer's instructions for the DYEnamic Direct cycle sequencing kit with the DYEnamic energy transfer primers (Amersham Life Science, Inc.) and loaded into an ABI Prism 377 stretch DNA sequencer (Applied BioSystems, Foster City, CA). The dye primer sequencing method yields generally uniform peak intensities, which facilitates identification of sequence variation in heterozygote individuals as the two comigrating peaks are of similar intensity, but ~50% of the intensity of the neighboring peaks in the chromatogram. All of the PCR products are sequenced in both directions, with the identification of the variant nucleotide in the sequencing of the reverse read providing evidence for the authenticity of the initial observation of sequence variation.

Sequence Analysis. The initial data analysis (lane tracking and base calling) is performed with the ABI prism DNA sequence analysis software (version 2.1.2). Chromatograms created by the ABI prism DNA sequence analysis software are imported into a Sun Microsystems UNIX workstation (Sun Microsystems Inc., Mountain View, CA). The chromatograms are reanalyzed with Phred (bases called and quality of sequence values assigned, version 0.961028), assembled with Phrap (version 0.960213) and the resultant data viewed with Consed (version 4.1).⁵ "PolyPhred" (version 2.1), a software package that uses the output from Phred, Phrap, and Consed, is used to identify single nucleotide substitutions in heterozygote individuals (26). All of the sequence variants identified by PolyPhred and the immediately surrounding region were inspected to confirm the existence of high-quality sequence reads in the region, before "marking" the nucleotide substitution or other sequence alteration as a variant. The consensus sequence for each gene is derived from the samples sequenced in this study and may differ from the specific sequence(s) in GenBank. The numbering of the nucleotides in the genomic sequence is consistent with the numbering in GenBank or the public domain draft sequence that is available. The common or wild-type allele is defined as the most common allele in the sample set sequenced rather than by the reference GenBank sequence. The GenBank genomic sequences for *ERCC1*, *LIG3*, *POLD1*, *RAD52*, and

⁴ Internet address: <http://www.gene.ucl.ac.uk/nomenclature/>.

⁵ Description and documentation for Phred, Phrap, and Consed may be obtained at <http://www.genome.washington.edu>.

Table 1 Summary of genes screened for variation

Gene	Alias	Samples screened ^a	Genomic sequence	cDNA sequence
<i>ADPRT</i>	<i>PARP</i>	I	<i>AC04143</i>	<i>J03473</i>
<i>APEX</i>	<i>APE1</i>	I, II	<i>M92444</i>	<i>M80261</i>
<i>CDK4</i>		II	<i>U37022</i>	<i>M14505</i>
<i>CDKN2A</i>	<i>INK4; p16</i>	II	<i>U12818–U12820</i>	<i>NM_000077</i>
<i>ERCC1</i>		II	<i>M63796</i>	<i>M13194</i>
<i>ERCC2</i>	<i>XPB</i>	II	<i>L47234</i>	<i>X52221</i>
<i>ERCC3</i>	<i>XPB</i>	IV	<i>AC027142</i>	<i>M31899</i>
<i>ERCC4</i>	<i>XPF</i>	II	<i>L76568</i>	<i>NM_005236</i>
<i>ERCC5</i>	<i>XPG</i>	IV	<i>D16305</i>	<i>NM_000123</i>
<i>FANCG</i>	<i>XRCC9</i>	I	<i>AC004472</i>	<i>U70310</i>
<i>FEN1</i>		III	<i>AC004770</i>	<i>L37374</i>
<i>LIG1</i>		I	<i>AC011466</i>	<i>M36067</i>
<i>LIG3</i>		I, III	<i>AC004223</i>	<i>U40671, X84740</i>
<i>MLH1</i>	<i>HNPCC</i>	III	<i>U40960TO–U40978</i>	<i>U07418</i>
<i>MRE11A</i>		I	<i>AP000786</i>	<i>NM_005590</i>
<i>MSH2</i>		III	<i>U41206–U41221</i>	<i>U03911</i>
<i>MSH3</i>		III	<i>D61397–D61419</i>	<i>U61981</i>
<i>MSH6</i>	<i>GTBP</i>	III	<i>AC006509</i>	<i>NM_000179</i>
<i>NBS1</i>		I	<i>AF069291.1</i>	<i>NM_002485</i>
<i>NTHL1</i>	<i>NTH1</i>	I	<i>AC005600</i>	<i>U79718</i>
<i>NUDT1</i>	<i>MTH1</i>	I	<i>D38591</i>	<i>D16581</i>
<i>PCNA</i>		II	<i>J04718</i>	<i>M15796</i>
<i>POLB</i>		I, II	<i>AH00541, and AF170802</i>	<i>NM_002690</i>
<i>POLD1</i>		I	<i>AC020909, and CITB-E1_2545M3</i>	<i>M80397</i>
<i>POLD2</i>		I	<i>AC006454</i>	<i>NM_006230</i>
<i>RAD23A</i>	<i>HHR23A</i>	II	<i>AD0000092</i>	<i>NM_005053</i>
<i>RAD23B</i>	<i>HHR23B</i>	IV	<i>AL137852</i>	<i>D21090</i>
<i>RAD50</i>		III	<i>AC004042A</i>	<i>U63139</i>
<i>RAD51</i>		I	<i>AF165088–AF165094</i>	<i>D14134</i>
<i>RAD52</i>		I	<i>AC004803</i>	<i>U27516</i>
<i>RAD54L</i>		I	<i>AL360086</i>	<i>NM_003579</i>
<i>XPA</i>		II	<i>AL442130, U16815</i>	<i>NM_000380</i>
<i>XPC</i>		IV	<i>AH009651</i>	<i>D21089</i>
<i>XRCC1</i>		I, II	<i>L34079</i>	<i>M36089</i>
<i>XRCC2</i>		III	<i>AC003109</i>	<i>Y08837</i>
<i>XRCC3</i>		II	<i>AF037222</i>	<i>NM_005432</i>
<i>XRCC4</i>		I	<i>AC034211, AC022416</i>	<i>NM_003401</i>

^a Sample sets are described in "Materials and Methods."

XRCC1 are in descending numbers because the genomic sequence for each of these genes is in the reverse orientation of the cDNA sequence.

Samples for Variation Screening. Four sets of samples were screened for variation. Table 1 relates the genes to the specific sample set screened and the number of unrelated individuals screened for variation for each gene. Table 1 also includes the GenBank number for the specific cDNA sequence used for assigning the location of the amino acid substitutions identified in these samples and listed in Table 2.

The majority of the genes have been screened for variation in 92 samples from the "DNA Polymorphism Discovery Resource" at the Coriell Institute for Medical Research (Group I). This resource was developed by the NIH to have a common set of samples available to investigators screening for common variants existing in the general population of the United States. The availability of lists of common polymorphisms in large numbers of candidate genes was expected to facilitate subsequent studies to relate genetic variation to disease risk (27). The samples (27) are from United States residents selected to represent the major ethnic groupings of the population, although the ethnic origin of specific individuals is unknown. The individuals in this sample set are from population groups as follows: European-American, 23; African-American, 23;

Mexican-American, 11; Native-American, 11; Asian-American, 23. Because the ethnicity of specific individuals is unknown, the estimated allele frequency data will not suggest potential differences among ethnic groups in the distribution of alleles. These samples cannot be associated with specific donors and were deemed to be exempt by the LLNL-IRB for human subjects research.

Because this variation screening effort was initiated before the establishment of the Polymorphism Discovery Resource, two other sample sets were screened for variation in early studies. The initial resequencing screened DNA from 36 unrelated individuals (Group II). Group II included 12 samples from unidentified individuals for whom no characteristics are known, although they are presumed to have been healthy at the time of sample collection in Ann Arbor, MI and are probably Caucasian. These are the same samples as screened by Shen *et al.* (25) in a preliminary search for variation in DNA repair genes. Because the samples cannot be associated with a donor, they were deemed to be exempt by the LLNL-IRB. Twenty-four additional samples in Group II were from Caucasians enrolled in a lung cancer study conducted at Johns Hopkins University, Baltimore, MD. Informed consent to use these samples for the study of the possible relationship of variation in DNA repair and cancer had been obtained and the study approved by the

Table 2 Amino acid substitution variants identified in DNA repair and repair-related genes

The common nucleotide followed by the variant nucleotide is enclosed in parentheses, and the codon for the amino acid residue is underlined and bold. The nine amino acid substitutions reported previously (25) are indicated with an asterisk.

Gene name	Exon	Codon	Common residue	Variant residue	Allele frequency	Mouse residue	cDNA sequence 5'→3'
BER							
<i>ADPRT</i>	4	188	Ala	Thr	0.006	Ser	TCCTT(G/A)CTACA
<i>ADPRT</i>	7	334	Val	Ile	0.011	Val	AGTGG(G/A)TAACC
<i>ADPRT</i>	8	383	Ser	Tyr	0.014	Ser	CTCCT(C/A)TGCTT
<i>ADPRT</i>	17	761	Val	Ala	0.18	Val	CAAGG(T/C)GGAAA
<i>ADPRT</i>	21	940	Lys	Arg	0.011	Lys	CAGCA(A/G)GTAC
<i>APE1</i>	3	51	Gln	His	0.03	Gln	GATCA(G/C)AAAAC
<i>APE1</i>	3	64	Ile	Val	0.01	Ile	TCAAG(A/G)TCTGC
<i>APE1</i>	5	148	Asp	Glu	0.33	Glu	GGCGA(T/G)GAGGA
<i>APE1</i>	5	241	Gly	Arg	0.01	Gly	GCTTC(G/A)GGGAA
<i>FEN1</i>		No variants					
<i>LIG1</i>	3	24	Ala	Val	0.01	Thr	GGAGG(C/T)ATCCA
<i>LIG1</i>	4	62	Arg	Trp	0.01	Gln	CGGCC(C/T)GGGTC
<i>LIG1</i>	9	249	Gly	Glu	0.01	Gly	GCCAG(G/A)GGGTC
<i>LIG1</i>	10	267	Asn	Ser	0.02	Asn	TTACA(A/G)TCCTG
<i>LIG1</i>	13	369	Val	Ile	0.01	Ile	AGTCC(G/A)TCCGG
<i>LIG1</i>	13	409	Arg	His	0.01	Cys	GTTC(G/A)CGACA
<i>LIG1</i>	16	480	Met	Val	0.01	Val	CAGCC(A/G)TGGTG
<i>LIG1</i>	20	614	Thr	Ile	0.01	Thr	GGTCA(C/T)ATCCT
<i>LIG1</i>	22	673	Glu	Asp	0.01	Gln	CGTGA(G/T)CCCCT
<i>LIG1</i>	22	677	Arg	Leu	0.01	Arg	TTCCC(G/T)GCCGC
<i>LIG3</i>	18	780	Arg	His	0.03	Cys	GTCCC(G/A)CAAGG
<i>LIG3</i>	19	811	Lys	Thr	0.01	Lys	TGCAA(A/C)GCCTT
<i>LIG3</i>	21	899	Pro	Ser	0.01	Thr	AGAAC(C/T)CTGCG
<i>NTHL1</i>	1	21	Arg	Trp	0.006	Arg	GGAGC(C/T)GGAGC
<i>NTHL1</i>	1	31	Gly	Val	0.006	Gly	CGGGG(G/T)GTGTA
<i>NTHL1</i>	1	33	Arg	Lys	0.006	Arg	GTGTA(G/A)GGAGG
<i>NTHL1</i>	3	176	Ile	Thr	0.005	Ile	GCTCA(T/C)CTACC
<i>NTHL1</i>	4	234	Ser	Leu	0.006	Ser	TGTGT(C/T)AGGCA
<i>NUDT1</i>	3	83	Val	Met	0.006	Val	TGGAC(G/A)TGCAT
<i>NUDT1</i>	4	135	Gly	Trp	0.006	Gly	TCCAC(G/T)GGTAC
<i>PCNA</i>		No variants					
<i>POLB</i>	1	8	Gln	Arg	0.01	Gln	GCCGC(A/G)GGAGA
<i>POLB</i>	7	137	Arg	Gln	0.006	Arg	TCAGC(G/A)AATTG
<i>POLB</i>	12	242	Pro	Arg	0.005	Pro	GCTTC(C/G)CAGTA
<i>POLD1</i>	1	19	Arg	His	0.12	Arg	GGCCC(G/A)TGGGG
<i>POLD1</i>	1	30	Arg	Trp	0.006	Ser	CACCT(C/T)GGCCA
<i>POLD1</i>	3	119	Arg	His	0.15	Arg	ATCCC(G/A)CGGCT
<i>POLD1</i>	4	173	Ser	Asn	0.05	Ser	CATCA(G/A)CCGGG
<i>POLD1</i>	4	177	Arg	His	0.003	Arg	CAGTC(G/A)CGGGG
<i>POLD1</i>	19	849	Arg	His	0.011	Arg	ACTGC(G/A)CCGCC
<i>POLD1</i>	26	1086	Arg	Gln	0.01	Arg	GGTGC(G/A)GAAGG
<i>POLD2</i>	8	303	Asn	Ser	0.005	Asn	CACCA(A/G)TTACA
<i>XRCC1</i>	3	72	Val	Ala	0.03	Val	GCTGG(T/C)GGGCA
<i>XRCC1</i>	5	161	Pro	Leu	0.005	Thr	GGCCC(C/T)GTCCC
<i>XRCC1</i>	5	173	Phe	Leu	0.005	Phe	CAGTT(C/G)CGTGT
<i>XRCC1</i>	6	194*	Arg	Trp	0.13	Arg	TCAGC(C/T)GGATC
<i>XRCC1</i>	9	280*	Arg	His	0.03	Arg	AACTC(G/A)TACCC
<i>XRCC1</i>	9	309	Pro	Ser	0.01	Ala	GACGA(C/T)CCCGA
<i>XRCC1</i>	10	399*	Arg	Gln	0.24	Arg	CTCCC(G/A)GAGGT
<i>XRCC1</i>	15	560	Arg	Trp	0.01	Arg	AGCGG(C/T)GGAAA
<i>XRCC1</i>	16	576	Tyr	Ser	0.01	Tyr	GGACT(A/C)TATGA
NER							
<i>ERCC1</i>		No variants					
<i>ERCC2</i>	8	199*	Ile	Met	0.01	Ile	TCAAT(C/G)CTGCA
<i>ERCC2</i>	8	201*	His	Tyr	0.01	His	TCCTG(C/T)ATGCC
<i>ERCC2</i>	10	312*	Asp	Asn	0.40	Asp	TGCC(G/A)ACGAA
<i>ERCC2</i>	20	616	Arg	Pro	0.01	Arg	CGGGC(G/C)GGCCG
<i>ERCC2</i>	23	751*	Lys	Gln	0.32	Gln	CGCTG(A/C)AGAGG
<i>ERCC3</i>		No variants					
<i>ERCC4</i>	7	379*	Pro	Ser	0.03	Pro	GCAAC(C/T)CAAAG
<i>ERCC4</i>	8	415	Arg	Gln	0.06	Arg	TGACC(G/A)AACAT
<i>ERCC5</i>	4	141	Asn	Asp	0.02	Asp	GAGAA(A/G)ACGAC
<i>ERCC5</i>	7	254	Met	Val	0.012	Met	AGGAA(A/G)TGAAT

Table 2 Continued

Gene name	Exon	Codon	Common residue	Variant residue	Allele frequency	Mouse residue	cDNA sequence 5'→3'
<i>ERCC5</i>	8	529	Cys	Ser	0.03	Arg	AACTT(G/C)TACAA
<i>ERCC5</i>	8	597	Val	Leu	0.011	Met	AGGCA(G/C)TAGAT
<i>ERCC5</i>	10	761	Ala	Thr	0.011	Ala	GGATC(G/A)CTGCT
<i>ERCC5</i>	15	1090	Glu	Asp	0.011	Asp	GGGGA(G/C)ACCTG
<i>ERCC5</i>	15	1104	Asp	His	0.18	Asp	GTGAA(G/C)ATGCT
<i>LIG1</i>		See above					
<i>PCNA</i>		See above					
<i>POLD1</i>		See above					
<i>POLD2</i>		See above					
<i>RAD23A</i>	5	200	Thr	Met	0.03	Thr	GCTCA(C/T)GGGAA
<i>RAD23B</i>	7	249	Ala	Val	0.10	Ala	TGGGG(C/T)TCCTC
<i>XPA</i>	6	256	Met	Ile	0.01	Met	GACAT(G/C)TACCG
<i>XPC</i>	1	16	Leu	Val	0.04	none	GCGAA(C/G)TGCGC
<i>XPC</i>	2	48	Leu	Phe	0.04	Ser	GCCTT(C/T)TCTCC
<i>XPC</i>	8	492	Arg	His	0.04	Arg	CCATC(G/A)TAAGG
<i>XPC</i>	8	499	Ala	Val	0.24	Ala	GCCAG(C/T)GGCAT
<i>XPC</i>	15	939	Lys	Gln	0.38	Lys	TTGAG(A/C)AGCTG
<i>DSB/RR</i>							
<i>LIG3</i>		See above					
<i>NBS1</i>	4	142	Asn	Ser	0.006	Asn	AAACA(A/G)TTGGA
<i>NBS1</i>	5	185	Gln	Glu	0.34	Glu	CAGTT(G/C)AGTCC
<i>NBS1</i>	6	196	Phe	Val	0.005	Phe	AAAGT(T/G)TTAC
<i>NBS1</i>	6	216	Gln	Lys	0.005	His	GACGG(C/A)AGGAA
<i>NBS1</i>	14	716	Asn	Asp	0.013	Asn	GAAAG(A/G)ATACA
<i>RAD50</i>	5	191	Thr	Ile	0.01	Thr	AGAAA(C/T)ACTTC
<i>RAD50</i>	16	884	Arg	His	0.024	Arg	ACGTC(G/A)TCAGC
<i>RAD50</i>	24	1239	Arg	Gln	0.01	Arg	TGACC(G/A)AGAAA
<i>RAD51</i>		No variants					
<i>XRCC2</i>	2	16	Ala	Ser	0.01	Ala	TCCTT(G/T)CCCCG
<i>XRCC2</i>	3	188	Arg	His	0.05	Arg	CTATC(G/A)CCTGG
<i>XRCC3</i>	7	241*	Thr	Met	0.43	Thr	GGCCA(C/T)GTGTC
<i>XRCC4</i>	1	12	Ser	Cys	0.01	Ser	TGTTT(C/G)TGAAC
<i>XRCC4</i>	2	75	Leu	Ser	0.006	Val	ATTGT(T/C)GTCCG
<i>XRCC4</i>	3	134	Ile	Thr	0.016	Ile	CACCA(T/C)TGCAG
<i>XRCC4</i>	3	137	Asn	Cys	0.005	Lys	GAAA(A/T/G)CAAGC
<i>XRCC4</i>	5	247	Ala	Ser	0.08	Ala	GGTTG(G/T)CTTCA
<i>Damage recognition and cell cycle checkpoints</i>							
<i>CDK4</i>	2	20	Val	Leu	0.03	Val	GGACA(G/T)TGTC
<i>CDKN2A</i>	2	148	Ala	Thr	0.05	Leu	ATGCC(G/A)CGGAA
<i>FANCG</i>	7	297	Thr	Ile	0.01	Ala	CACA(A/C/T)AGCAG
<i>FANCG</i>	8	330	Pro	Ser	0.01	Pro	TACTG(C/T)CACCA
<i>FANCG</i>	9	378	Ser	Leu	0.02	Ser	TAGCT(C/T)GGAGC
<i>FANCG</i>	10	464	Val	Phe	0.01	Ser	CCTGG(G/T)TTCAA
<i>RAD52</i>	4	70	Arg	Trp	0.005	Arg	GTCAT(C/T)GGGTA
<i>RAD52</i>	7	221	Gln	Glu	0.011	Glu	TGCAG(C/G)AGGTG
<i>RAD52</i>	8	287	Ser	Asn	0.05	His	GAAGA(G/A)TGAGG
<i>RAD54L</i>	4	74	Ile	Met	0.006	Ile	TTTAT(T/G)CGAAG
<i>RAD54L</i>	7	202	Arg	Cys	0.006	Arg	TTTTA(C/T)GCCAG
<i>RAD54L</i>	10	380	Arg	Gln	0.011	Arg	GGAGC(G/A)GCTGC
<i>RAD54L</i>	16	583	Ile	Thr	0.01	Ile	TCTCA(T/C)TGGGG
<i>MLH1</i>	8	213	Val	Met	0.04	Val	CAACC(G/A)TGGGAC
<i>MLH1</i>	8	219	Ile	Val	0.12	Ile	GCTCC(A/G)TCTTT
<i>MLH1</i>	11	325	Arg	Gln	0.01	Arg	GGAGC(G/A)GTGC
<i>MLH1</i>	11	326	Val	Ala	0.01	Val	GCGGG(T/C)GCAGC
<i>MLH1</i>	12	452	Thr	Ser	0.01	Leu	ATACA(A/T)CAAAG
<i>MLH1</i>	16	618	Lys	Ala	0.01	Lys	AGAAG(A/G)&(A/C)GGCTG
<i>MLH1</i>	16	618	Lys	Thr or Glu	0.01	Lys	AGAAG(A/G)or(A/C)GGCTG
<i>MLH1</i>	19	718	His	Tyr	0.01	His	TGGAA(C/T)ACATTG
<i>MSH2</i>	3	127	Asn	Ser	0.006	Asn	TGGCA(A/G)TCTCT
<i>MSH2</i>	3	170	Gln	Glu	0.006	Gln	CCATA(C/G)AGAGG
<i>MSH2</i>	6	319	Asp	Val	0.009	Asp	TGAAG(A/T)TACCA
<i>MSH2</i>	6	322	Gly	Asp	0.011	Gly	CACTG(G/A)CTCTC
<i>MSH2</i>	7	390	Leu	Phe	0.005	Leu	ACCGA(C/T)TTGCC

Table 2 Continued

Gene name	Exon	Codon	Common residue	Variant residue	Allele frequency	Mouse residue	cDNA sequence 5'→3'
MMR							
<i>MSH2</i>	13	735	Ile	Val	0.006	Ile	CTTCT(A/G)TCCTC
<i>MSH3</i>	8	429	Ala	Val	0.04	Met	AGAGG(C/T)GCTCA
<i>MSH3</i>	9	456	Tyr	Cys	0.006	Tyr	TGAAT(A/G)CAGCC
<i>MSH3</i>	10	514	Glu	Lys	0.05	Glu	AACCT(G/A)AGAAT
<i>MSH3</i>	13	597	Ser	Asn	0.014	Ser	ATCTA(G/A)TGTGT
<i>MSH3</i>	15	700	Phe	Leu	0.005	Phe	CTGAC(T/C)TCCTC
<i>MSH3</i>	21	931	Gly	Cys	0.006	Gly	GGATG(G/T)GTGCT
<i>MSH3</i>	21	940	Arg	Gln	0.10	Arg	AGGAC(G/A)GAGTA
<i>MSH3</i>	23	1036	Thr	Ala	0.30	Asp	CAGGC(A/G)CAGCA
<i>MSH6</i>	1	25	Ala	Val	0.003	Ala	CTCGG(C/T)CAGGG
<i>MSH6</i>	1	39	Gly	Glu	0.24	Gly	CCCCG(G/A)GGCCT
<i>MSH6</i>	4	396	Leu	Val	0.009	Leu	CTACA(C/G)TCTAT
<i>MSH6</i>	4	878	Val	Ala	0.016	Val	AGAAG(T/C)TGCTG
<i>MSH6</i>	6	1152	Val	Ile	0.009	Val	TAGCT(G/A)TAATG

Johns Hopkins University Institutional Review Board and the LLNL-IRB. Twelve of the samples are from cancer cases and 12 are from controls. No variants were identified in resequencing of lymphocyte DNA from the lung cancer cases that were not identified in other individuals also.

A subsequent sample set, that was also a predecessor to the Polymorphism Discovery Resource, included 72 individuals of African, Asian, or Caucasian origin selected for geographical diversity (Ref. 28; Group III). These samples are available from the Coriell Institute for Medical Research (Camden, NJ) and were also deemed to be exempt by the LLNL-IRB.

A fourth set of 46 samples (Group IV) has been used to screen for variation in genes of the NER pathway that may be associated with risk of melanoma. These individuals are of Caucasian origin and were selected because of the previous diagnosis of recurrent melanoma. The screening of these samples for sequence variation was approved by the Institutional Review Board of the Memorial Sloan-Kettering Cancer Center (New York, NY) and the LLNL-IRB.

All of the repair genes screened for variation are located on autosomes and, therefore, the number of chromosomes screened is twice the number of samples. Thus, in summary, Group I is 184 chromosomes, Group II is 72 chromosomes, Group III is 144 chromosomes, and Group IV is 92 chromosomes. *APEX*, *POLB*, and *XRCC1* were screened for variation in both the Group I and II sample sets or a total of 256 chromosomes, and *LIG3* was screened in the samples of Groups I and III or 328 chromosomes (Table 1).

Results

Summary of Amino Acid Substitution Variants. A total of 127 different single nucleotide polymorphisms resulting in amino acid substitutions were identified in the screening of 37 genes (Table 2). The sequence variation data and the individual genotypes for each individual for the genes screened in the Polymorphism Discovery Resource sample set (Group I) can be accessed on the Internet.⁶ Similar data for the Group III genes can also be accessed.⁷

A large number of different amino acid substitution variants were identified in several genes, including 10 different

variants for *LIG1*, 9 for *XRCC1*, and 8 for *MLH1*. Although none of the 10 *LIG1* variants exists at a frequency of over 0.02, the total variant allele frequency for the gene is 0.11. Similarly, the seven low-frequency variants of both *XRCC1* and *MLH1* exist at total variant allele frequencies of 0.10 for each gene. No amino acid substitution variants were detected in *RAD51*, *PCNA*, *FEN1*, *ERCC1*, and *ERCC3*. With the exception of *FEN1*, a small intronless gene, nucleotide substitutions were identified in both the exons and introns of these genes. In the collection of 127 amino acid substitution variants, 22 variant alleles (17%) exist at estimated frequencies of greater than 0.05. Two variant alleles existing at frequencies of at least 0.10 were identified in five genes (*ERCC2*, *XPC*, *MSH3*, *XRCC1*, and *POLD1*). The estimates of allele frequencies must be considered tentative, as relatively small numbers of chromosomes were screened for variation, although the estimated allele frequencies for several of the common polymorphisms are similar to frequencies observed in subsequent molecular epidemiology studies (29–31). The two nucleotide substitutions in the codon for amino acid residue 618 (Lys) of *MLH1* exist in the same individual. From the direct sequencing of the PCR product, it is not possible to differentiate between nucleotide substitutions on the same chromosome and substitutions on opposite chromosomes; thus, three possibilities for the variant amino acid residue exist, Ala if the substitutions are on the same chromosome or Thr and Glu if the substitutions are on different chromosomes.

Six additional variants with potential to disrupt protein structure were identified (Table 3). Two of the variants were identified in *RAD50*. The first is an insertion of three nucleotides resulting in the addition of a Gln residue after amino acid residue 363, with the sequence changing from Gln-Glu-His-Ile to Gln-Glu-(Gln)-His-Ile. The second variant is a nucleotide substitution in the codon for Gln at 826, generating a termination signal and resulting in deletion of ~30% of the protein. Each of these variants was observed only once, existing at an estimated allele frequency of <0.01. Nucleotide substitutions resulting in the generation of termination codons and synthesis of truncated proteins were identified at amino acid residue 333 of *MRE11A* (allele frequency of 0.009) and residues 346 (allele frequency of 0.033) and 415 (allele frequency of 0.041) of *RAD52*. A four-nucleotide duplication at the COOH terminus of *MSH6* was identified (allele frequency of 0.02). This insertion changed the amino acid sequence from Thr-Leu-Ile-Lys-

⁶ Internet address: <http://greengenes.llnl.gov/dpublic/secure/reseq/>.

⁷ Internet address: <http://manuel.niehs.nih.gov/egsnp/home.htm>.

Table 3 Summary of non-amino acid substitution variants disrupting protein structure

Gene name	Exon	Codon	Common residue	Variant residue	Allele frequency	Mouse residue	cDNA sequence 5'→3'
<i>RAD52</i>	10	346	Ser	Stop	0.033	Leu	GCCCT(C/A)GTCTA
<i>RAD52</i>	11	415	Tyr	Stop	0.041	Leu	AAATA(T/G)GATCC
<i>RAD50</i>	8	363	QEHI ^a	QE(Q)HI	0.005	QEHI	TCAAG(AAC/AACAAC)ATATC
<i>RAD50</i>	15	826	Gln	Stop	0.005	Gln	CTGTC(C/T)AAACA
<i>MRE11A</i>	8	333	Gln	Stop	0.009	Gln	CCATA(C/T)AAAGC
<i>MSH6</i>	16	1357	TLIKEL(STOP) ^b	TLID(STOP)	0.020	ALINGL(STOP)	GACT(TTGA/TTGATTGA)TTAAG

^a The wild-type human sequence is Gln-Glu-His-Ile, the variant sequence is Gln-Glu-Gln-His-Ile, and the mouse sequence is Gln-Glu-His-Ile.

^b The human wild-type sequence is Thr-Leu-Ile-Lys-Glu-Leu, the variant sequence is Thr-Leu-Ile-Asp, and the mouse sequence is Ala-Leu-Ile-Asn-Gly-Leu.

Table 4 Summary of variants identified in genes in different DNA repair pathways

Pathway	No. of genes	No. of variants	Variants/gene (range)	Allele frequency (average)
BER	12	49	0–10	0.033
NER	13	40	0–9	0.061
DSB/RR	8	22	0–5	0.054
MMR	4	28	6–8	0.038
Damage recognition and cell cycle checkpoint	5	15	1–5	0.02
Total ^a	37	133		
Average			3.6	0.047

^a *LIG1*, *POLD1*, *POLD2*, and *PCNA* have functions in both the BER and the NER pathways. *LIG3* functions in both BER and DSB/RR. These genes are counted in both pathways in the Table. The total is the number of different genes resequenced and the number of unique variants identified. The count includes both the protein truncation and the amino acid substitution variants.

Glu-Leu-stop to the variant sequence Thr-Leu-Ile-Asp-stop. No nucleotide substitutions were observed in the critical splice site consensus sequences (splice site plus or minus two nucleotides). In total, 133 different nucleotide sequence alterations with potential to disrupt protein function were identified in the screening of 37 genes.

The variation identified in screening 37 genes is summarized by repair pathway in Table 4. An average of 3.6 different variants were identified per gene. The totals in Table 4 are the number of different genes screened and unique variants observed. They are not the sum of the numbers for the individual pathways because several genes have roles in more than one repair pathway. The average variant allele frequency is 0.047 (Table 4). No major differences in average allele frequency for the genes of the different pathways were noted.

In addition to the 133 nucleotide substitution and insertion/deletion variants described in Tables 2 and 3, 96 nucleotide substitutions that did not result in amino acid substitutions were detected within the exons. Thus, ~60% of the nucleotide substitutions identified in the exons of these repair genes result in amino acid substitutions. In addition, 608 nucleotide substitutions have been identified in the adjacent intronic and 3'- and 5'-UTRs of these genes. Initial analysis does not provide evidence that any of the nucleotide substitutions in the 5'-UTR disrupt known regulatory sites.⁸ These sequence variation data are also available on the Internet.⁶

Characteristics of Variants. Twenty-eight of the 127 amino acid substitutions (22%) occurred at Arg residues, *i.e.*, Arg is the common allele. Twenty of the nucleotide substitutions in the Arg codon were at the G residue and 18 of the substitutions

were G to A (CGX to CAX), resulting in replacement of an Arg with either His (10 variants) or Gln (8 variants), depending on the third position nucleotide. His and Gln are amino acid residues with properties that differ from Arg. Six substitutions occurred at Pro residues, a residue imparting significant constraints on protein structure.

The nature of the amino acid residues involved in the interchange can be defined by criteria reflective of the physical and chemical properties of the respective members of the amino acid pair (32). Using the groupings of amino acid residues by similar chemical and physical characteristics used by Smith and Smith (33), 101 of the 127 variants (80%) involve the interchange of amino acid residues with dissimilar physical or chemical properties.

The evolutionary conservation of the amino acid residues at the site of the substitution is another characteristic suggestive of the potential for an amino acid substitution to impact protein function. Ninety-five (75%) of the 127 amino acid substitution variants exist at amino acid residues at which the common allele in the human protein is identical to the amino acid residue in the mouse protein. Seventy-nine or 62% of the substitutions involve the exchange of amino acid residues with dissimilar physical or chemical properties at residues at which the common allele in humans encodes the same amino acid residue as observed in the mouse protein. This number is reduced to 56 or 44% when the properties of the exchanged amino acid residues are scored by the Blosum 62 matrix (34). Over 50% of the 22 variants with allele frequencies of >0.05 involve the exchange of amino acid residues with dissimilar physical and/or chemical properties at residues at which the human and mouse proteins have the same amino acid residue, characteristics often associated with negative impact on protein function.

Contributions of Individual Alleles to Total Variation and the Complexity of Individual Genotypes. The relative contribution of variants existing at different allele frequencies to

⁸ Tong Xi, Johana Vázquez-Matías, and Harvey W. Mohrenweiser. Single nucleotide polymorphisms in the exonic and neighboring intronic and untranslated regions of 40 DNA repair and repair related genes in humans, manuscript in preparation.

Table 5 Contribution of alleles of different frequency to total variation

Allele frequency	Variant alleles (n)	Total variation (%)	Cumulative variation (%)
>0.40	2	14	14
0.30–0.399	5	27	41
0.20–0.299	3	11	52
0.10–0.199	7	16	68
0.05–0.099	8	8	76
0.02–0.049	22	11	88
<0.02	86	12	100

the total genetic variation among individuals is presented in Table 5. The seven variants existing at frequencies of >0.30 account for 41% of the total genetic variation among individuals for these genes, whereas the 116 variant alleles existing at individual frequencies of less than 10% account for 32% of the total variation in the population.

Homozygous variant individuals were identified for 14 of the 17 variant alleles existing at variant allele frequencies of 0.10 or greater. Two of the variant alleles not observed in the homozygous state were at loci (*XPC* and *RAD23B*) at which only 46 individuals were screened for variation.

Two or more different variant alleles were identified in 26 genes, and many individuals with two different variant alleles in a gene were observed. These individuals would express two different variant forms of a protein (if the substitutions were on different chromosomes) or a wild-type protein and a variant protein with two amino acid substitutions (if the substitutions were located on the same chromosome). For example, 14 of 36 individuals screened for variation in *ERCC2* exhibited variation at more than one residue. One individual was homozygous variant at both of the highly polymorphic sites (312Asn/Asn and 751Gln/Gln) in *ERCC2*, seven individuals were heterozygous at one site and homozygous variant at the other and six were heterozygous at both sites. Among 128 individuals screened for variation in *XRCC1*, eight individuals were homozygous variant at residue 399 (399Gln/Gln), and three were homozygous variant at residue 194 (194Trp/Trp). No individuals were homozygous at one of these two sites in *XRCC1* and also variant at the second site, which suggested that the variant alleles at *XRCC1* 194 and 399 were on different chromosomes. Twelve individuals were variant at one of these two sites and also at an additional site within the protein.

Complex genotypes were observed in compiling the variation for the genes of a repair pathway for an individual. The BER pathway, in which 49 different variants were identified in screening 12 of the 30 genes of this pathway, is illustrative. Only 3 of 90 individuals were homozygous wild-type at all 12 of the loci, whereas 5 individuals were identified with 5 variant alleles and 7 individuals had 6 variant alleles among the 12 BER genes screened thus far (Table 6). The average of 2.8 variant alleles per individual would extrapolate to ~7 variants in the average individual, from a pool estimated to include ~115 different variants, for the 30 genes of the BER pathway.

Seventy-five different combinations of alleles or pathway genotypes were observed in compiling the variation for these 12 BER genes. In addition to the three individuals who were wild-type at all of the 12 loci screened, five groups of either two or three individuals were heterozygous for the same single variant, whereas two pairs of individuals were heterozygous for the same 2 variant alleles. All of the 49 individuals with 3 or more variant alleles among these 12 genes had unique combination of single nucleotide polymorphisms. Examples of the

Table 6 Distribution of variant alleles per individual identified in screening 12 genes of the BER pathway in 90 individuals

No. of variant alleles/individual	No. of individuals
0	3
1	19
2	19
3	26
4	11
5	5
6	7

complexity of genotypes observed are presented for the 26 individuals with three variant alleles in Table 7. The data presented include only the 27 variant alleles existing in this specific subset of individuals.

Similar complex pathway genotypes were observed for the other repair pathways, although the data sets were not as complete because not all of the genes screened in these pathways were resequenced in the same sample set. [The individual genotype data (except for the limited data for Group II individuals) are available at <http://greengenes.llnl.gov/dpublic/secure/reseq/or> <http://manuel.niehs.nih.gov/egsnp/home.htm>.]

Discussion

Understanding the molecular and biological basis for the reduced-DNA-repair-capacity phenotype and the associated elevation in individual cancer risk will require that both the genes involved in the repair of the different classes of DNA damage be known and that the common variants in these genes be cataloged. Thus, this study focused on well-characterized genes of the different DNA repair pathways and the associated DNA damage recognition and cell cycle checkpoint processes. The ultimate goal is the identification of the common variants segregating in the United States population for all of the genes with roles in the repair of damaged DNA. This catalogue will serve as a resource for future biochemical and molecular epidemiology studies. The strategy used here, screening a limited number of individuals selected to represent a larger population, will not provide the resources for the study of genetically distinct populations or subpopulations with specific variation and potentially unique risks, e.g., the Ashkenazi Jewish population (35). Molecular epidemiology of genetically isolated populations will require additional screening to ensure ascertainment of variants specific to these populations. Although the ethnic composition of the set of samples screened for variation in this study is known (27), the ethnicity of specific individuals is unknown; thus, the estimated allele frequencies do not address the issue of ethnic differences in allele distribution or frequency. Future molecular epidemiology studies that address questions regarding the role of these variants in disease susceptibility will undoubtedly obtain data regarding potential differences in allele frequencies among different ethnic groups or substructured populations.

Extensive variation, 3.6 different amino acid substitution- and protein sequence altering-variants per gene with an average variant allele frequency of 0.047, was identified in the resequencing of the 37 genes in this screening effort. This is consistent with previous, more limited or focused studies of variation in DNA repair genes in "non-cancer family" individuals (36–38). The common variants of *ERCC2* (39), *ERCC4* (40), *RAD52* (41), *XPC* (42), *MSH3* (43), and *RAD54* (41) that

Table 7 Examples of genotypes observed in individuals with four variant alleles among 12 genes of the BER pathway

"1" designates the wild-type allele and "2" is the variant allele; heterozygote individuals are shaded light gray with a diagonal line, and homozygous variant individuals are shaded dark gray and boxed. The top line is the wild-type amino acid, line 2 is the variant residue, and the third line indicates the position of the substitution. Additional details regarding the substitutions are in Table 2.

	ADPRT	ADPRT	ADPRT	APEX	APEX	APEX	APEX	LIG1	LIG1	LIG1	LIG3	NTHL1	NTHL1	NTHL1	NUDT1	POLD1	POLD1	POLD1	POLD1	POLD1	POLD1	POLD2	XRCC1	XRCC1	XRCC1	XRCC1	XRCC1
cell	Ala	Val	Lys	Gln	Ile	Asp	Gly	Arg	Asn	Glu	Pro	Gly	Arg	Ile	Val	Arg	Arg	Arg	Ser	Arg	Arg	Asn	Pro	Phe	Arg	Arg	Tyr
line	Thr	Ala	Arg	His	Val	Glu	Arg	Trp	Ser	Asp	Ser	Val	Lys	Thr	Met	His	Trp	His	Asn	His	His	Ser	Leu	Leu	Trp	Gln	Ser
	188	761	940	51	64	148	241	62	267	673	899	31	33	176	83	19	30	119	173	177	849	303	161	173	194	399	526
76	11	11	11	11	11	11	11	11	11	11	11	11	12	11	11	11	11	11	11	11	11	11	11	11	11	11	11
04	12	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	12	12	11	11	11	11	11	11	11
32	11	11	11	11	11	11	11	11	11	12	11	11	11	11	11	11	11	11	12	12	11	11	11	11	11	11	11
39	11	12	11	11	11	11	12	11	11	11	11	11	11	12	11	11	11	11	11	11	11	11	11	11	11	11	11
61	11	12	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	12	11	11	11	12
19	11	12	11	11	11	11	11	11	11	11	11	11	11	11	12	12	11	11	11	11	11	11	11	11	11	11	11
16	11	12	11	11	11	11	11	11	11	11	11	12	11	11	11	11	11	11	11	11	11	11	11	11	11	12	11
18	11	11	11	11	11	11	11	12	11	11	11	11	11	11	11	11	11	12	11	11	11	11	11	11	11	12	11
07	11	12	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	12	11	12	11
28	11	12	11	11	11	11	11	11	11	11	11	11	11	11	11	12	11	11	11	11	11	11	11	11	11	12	11
86	11	12	11	12	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	12	11
60	11	11	11	12	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	22	11
37	11	11	11	11	11	12	11	11	11	12	11	11	11	11	11	11	11	11	11	11	11	12	11	11	11	11	11
35	11	11	11	11	12	11	11	11	11	11	11	11	11	11	12	11	11	11	11	11	11	11	11	11	11	11	11
10	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	12	22	11
48	11	11	12	11	11	12	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	12	11	11
71	11	11	11	11	11	12	11	12	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	12	11	11
41	11	11	11	11	11	12	11	11	12	11	11	11	11	11	11	11	11	11	12	11	11	11	11	11	11	11	11
47	11	12	11	11	11	12	11	11	11	11	11	11	11	11	11	11	11	11	12	11	11	11	11	11	11	11	11
27	11	11	12	11	11	12	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	12	11
43	11	11	11	11	11	12	11	11	11	11	11	11	11	11	11	11	11	11	11	11	12	11	11	11	11	12	11
25	11	11	11	11	11	12	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	22	11
22	11	12	11	11	11	22	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
31	11	12	11	11	11	22	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	12	11	11
52	11	11	11	11	11	22	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	12	11
50	11	11	11	11	11	22	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	12	12	11

were observed in this resequencing effort were identified in other studies also. The *RAD51* gene appears to be very invariant, because no amino acid substitution variants were identified in screening 92 individuals in this study or in 100 additional individuals (38) or in over 60 human tumor samples (44). Kato *et al.* (45) identified one amino acid-substitution variant in *RAD51* in 2 of 45 Japanese breast cancer patients, but this variant was not seen in genotyping 200 additional Japanese breast cancer and 100 colon cancer patients. As in this screen, no amino acid substitution variants were detected in screening another sample set for variants in *FEN1* and *PCNA* (46). Although the study presented in the present report focuses on nucleotide sequence variation that results in amino acid substitutions or that otherwise disrupts protein structure extensive additional sequence variation in both exons and also noncoding regions that could potentially impact gene expression was identified.⁸ The reported association of variation in the 5'-UTR of *RAD51* with the risk of breast and ovarian cancer in *BRCA1/2* mutation carriers is an example of the potential for variation in noncoding regions of genes to be important (47).

The average of 3.6 different variants per repair gene is slightly higher than the number of different variants observed in the systematic screening of other sets of candidate disease-susceptibility genes. For example, an average of 1.1–2.8 different amino acid substitution variants per gene were observed in sequencing 36 cardiovascular risk genes (48), 75 hypertension risk genes (49), or 106 common disease risk genes (50) in samples from multiethnic populations. In contrast to the extensive variation observed in screening multiethnic sample sets, only 0.4–0.5 different amino acid substitution variants per gene were identified in genes associated with the risk of ischemic heart disease (41 genes; Ref. 51) or rheumatoid arthritis (41 genes; Ref. 52) in the resequencing of DNA from 49 Japanese individuals. This lower level of variation in the Japanese sample could be related to a lower level of heterogeneity in that population. It is also consistent with the observation that the

total number of different variants detected is lowest when screening individuals of only one ethnic origin and is highest when individuals of African origin are included in a multiethnic sample set (53, 54). Similar results were observed in the present study, in which two variants of *APEX* and five variants of *XRCC1* that were not observed in the Caucasian samples of Group II were detected in screening the multiethnic Group I sample set. Thus, it is expected that the total number of different variants identified in these genes would have been even higher if the 11 repair genes that were screened for variation only in the Caucasian sample sets (Groups II and IV) had been screened in a multiethnic sample set.

The functional relevance of the extensive variation identified in DNA repair genes remains to be fully addressed, although initial data are becoming available. Seven variants of *APEX* identified in this study, in other reports (55, 56), or GenBank DNA sequence databases have been characterized as to activity in biochemical assays (57). Four of the variants retained only 10–60% of normal or wild-type activity. Three of the reduced activity variants involved the exchange of amino acid residues with dissimilar properties at residues that were identical in the human and mouse proteins. The fourth variant, which retained 60% of wild-type activity, was a replacement of Glu at residue 126 by Asp, residues with similar properties. Two of the variants retaining wild-type activity (Asp148Glu and Gly306Ala) were exchanges that would not be expected to disrupt structure by these limited criteria. The exchange of Arg at residue 241 for Gly (the mouse residue is Gly), residues with dissimilar characteristics, also retained wild-type activity. Using the characteristics of the amino acid residues, knowledge of evolutionary conservation and localization of the residues within the three-dimensional structure of the *APEX* protein, Hadi *et al.* (57) correctly predicted the impact of the substitutions for six of these seven variants. The exception was the Glu to Asp exchange at residue 126. The availability of the protein

structure obviously enhances the potential to correctly predict the impact of a substitution on protein activity.

In other studies to address the question of functional relevance, several variants have been associated with altered DNA repair capacity or level of damage from an exposure. Spitz *et al.* (58) reported that variant alleles at amino acid residues 312 and 751 of *ERCC2* were associated with a reduced capacity to repair damage induced by *in vitro* exposure of lymphocytes from individuals of a lung cancer cohort to benzo[*a*]pyrene diol epoxide. Homozygosity for a variant allele in either of two NER genes, *XPC* or *ERCC2*, was associated with the reduced capacity to repair UV-induced DNA damage as assayed by the host-reactivation assay in lymphocytes from a cohort of healthy subjects (59). Hu *et al.* (31) reported that the *APEX 148Glu* allele was associated with prolonged mitotic delay in lymphocytes exposed to ionizing radiation. In addition, women with at least three variant alleles of *APEX* and *XRCC1* were at increased likelihood of being ionizing-radiation sensitive. The *XRCC1 399Gln* variant has been associated with increased aflatoxin adducts (30) and with increased polyphenol adducts in the breast tissue of smokers (29). This variant has also been associated with increased levels of glycophorin A mutations in RBCs (30) and sister chromatid exchanges in lymphocytes from smokers (29). The data suggest that the approximately 40% of the population with this variant allele of *XRCC1* will have 20–50% more DNA damage than will individuals with the common allele after similar exposures.

A steadily increasing number of molecular epidemiology studies are reporting on the potential association of polymorphic DNA-repair gene variants with cancer risk. Over 30 studies reporting on the cancer risk associated with one or more of the variants of *APEX 148*, *ERCC2 312* and *751*, *XRCC1 194* and *399*, or *XRCC3 241* have been published. These manuscripts can be accessed using the appropriate gene symbol to search the National Center for Biotechnology Information *Entrez* search and retrieval system.⁹ These genotyping-risk-association studies can be summarized as follows: (a) the sample sizes are generally small, and usually the genotyping included only the variants in a single gene of a repair pathway and sometimes only a single variant in a gene; (b) most of these variant alleles have been reported to be associated with cancer risk in some studies; (c) other studies have reported an absence of association in cohorts with similar cancers or cohorts with other cancers; and (d) in some studies, the common or wild-type allele was associated with elevated risk.

The molecular epidemiology and biochemical characterization of individual variants are beginning to address the relationship of individual polymorphic variants to repair capacity and relevance as cancer risk factors. The cataloguing of variants in the genes in pathways with roles in ameliorating the negative consequences of exposures is a step in initiating these studies. It is important that a majority of the variants in each gene in a pathway be identified and be used as reagents for molecular epidemiology and biochemical function studies, because it is expected that the impact of most of the variants will be individually small (60, 61). It is usually assumed that selection against the alleles associated with elevated risk of complex diseases is generally small because they are expected to have only mildly deleterious impacts on function (62). As observed in the present study, the collective contribution of the low-frequency variants to the total variation among individuals is large, emphasizing the importance of using the full richness of this

genetic variation across all relevant genes and pathways to estimate individual susceptibility. The development of approaches for relating complex genotypes, interacting with exposure (63), to disease risk (64) will be required. In the end of days, the refinement of estimates of cancer risk will require extensive genotyping in studies using large cohorts with well-documented exposure histories and disease status to account for the manifold and subtle effects of gene-gene and gene-gene-environment interactions in determining individual susceptibility.

Acknowledgments

We gratefully acknowledge the assistance of Suzanne Duarte, Arlene Gonzales, and Karolyn Burkhart-Schultz in the sequencing and of Linda Ott, Mimi Yeh, and Tom Slezak in the informatics support. We appreciate the many discussions with Dr. David Wilson, III, and the assistance of Dr. Gloria Petersen in supplying the samples from Johns Hopkins University and of Dr. James Selkirk and the Environmental Genome Program of the National Institute of Environmental Health Sciences.

References

- Cleaver, J. E., Karplus, K., Kashani-Sabet, M., and Limoli, C. L. Nucleotide excision repair, "a legacy of creativity." *Mutat. Res.*, 485: 23–36, 2001.
- Batty, D. P., and Wood, R. D. Damage recognition in nucleotide excision repair of DNA. *Gene (Amst.)*, 241: 193–204, 2000.
- Thompson, L. H., and Schild, D. Homologous recombinational repair of DNA ensures mammalian chromosome stability. *Mutat. Res.*, 477: 131–153, 2001.
- Pfeiffer, P., Goedecke, W., and Obe, G. Mechanisms of DNA double-strand break repair and their potential to induce chromosomal aberrations. *Mutagenesis*, 15: 289–302, 2000.
- Khanna, K. K., and Jackson, S. P. DNA double-strand breaks: signaling, repair and the cancer connection. *Nat. Genet.*, 27: 247–254, 2001.
- Lindahl, T. Suppression of spontaneous mutagenesis in human cells by DNA base excision-repair. *Mutat. Res.*, 462: 129–135, 2000.
- Wilson, S. H. Mammalian base excision repair and DNA polymerase beta. *Mutat. Res.*, 407: 203–215, 1998.
- Demple, B., and Harrison, L. Repair of oxidative damage to DNA: enzymology and biology. *Annu. Rev. Biochem.*, 63: 915–948, 1994.
- Kolodner, R. D., and Marsischky, G. T. Eukaryotic DNA mismatch repair. *Curr. Opin. Genet. Dev.*, 9: 89–96, 1999.
- Hsieh, P. Molecular mechanisms of DNA mismatch repair. *Mutat. Res.*, 486: 71–87, 2001.
- Wood, R. D., Mitchell, M., Sgouros, J., and Lindahl, T. Human DNA repair genes. *Science (Wash. DC)*, 291: 1284–1289, 2001.
- Romen, A., and Glickman, B. W. Human DNA repair genes. *Environ. Mol. Mutagen.*, 37: 241–283, 2001.
- Orr-Weaver, T. L., and Weinberg, R. A. A checkpoint on the road to cancer. *Nature (Lond.)*, 392: 223–224, 1998.
- Weinert, T. DNA damage checkpoints update: getting molecular. *Curr. Opin. Genet. Dev.*, 8: 185–193, 1998.
- Cleaver, J. E. Xeroderma pigmentosum: the first of the cellular caretakers. *Trends Biochem. Sci.*, 26: 398–401, 2001.
- Hoeijmakers, J. H. Genome maintenance mechanisms for preventing cancer. *Nature (Lond.)*, 411: 366–374, 2001.
- Ishikawa, T., Ide, F., Qin, X., Zhang, S., Takahashi, Y., Sekiguchi, M., Tanaka, K., and Nakatsuru, Y. Importance of DNA repair in carcinogenesis: evidence from transgenic and gene targeting studies. *Mutat. Res.*, 477: 41–49, 2001.
- Grossman, L., Matanoski, G., Farmer, E., Hedayati, M., Ray, S., Trock, B., Hanfelt, J., Roush, G., Berwick, and Hu, J. J. DNA repair as a susceptibility factor in chronic diseases in human populations. In: M. Dizdaroglu and A. E. Karakaya (eds.), *Advances in DNA Damage and Repair*, pp. 149–167. New York: Kluwer Academic/Plenum Publishers, 1999.
- Wu, X., Gu, J., Amos, C. I., Jiang, H., Hong, W. K., and Spitz, M. R. A parallel study of *in vitro* sensitivity to benzo[*a*]pyrene diol epoxide and bleomycin in lung cancer cases and controls. *Cancer (Phila.)*, 83: 1118–1127, 1998.
- Wu, X., Spitz, M. R., de Andrade, M., Benowitz, N. L., and Swan, G. E. Genetic influence on mutagen sensitivity: a twin study. *Proc. Am. Assoc. Cancer Res.*, 41: 437, 2000.
- Cloos, J., Nieuwenhuis, E. J. C., Boomsma, D. I., Kuik, D. J., van der Sterre, M. L. T., Arwert, F., Snow, G. B., and Braakhuis, B. J. M. Inherited susceptibility

⁹ Internet address: <http://www.ncbi.nlm.nih.gov/Entrez/>.

- to bleomycin-induced chromatid breaks in cultured peripheral blood lymphocytes. *J. Natl. Cancer Inst. (Bethesda)*, *91*: 1125–1130, 1999.
22. Roberts, S. A., Spreadborough, A. R., Bulman, B., Barber J. B., Evans D. G., and Scott, D. Heritability of cellular radiosensitivity: a marker of low-penetrance predisposition genes in breast cancer. *Am. J. Hum. Genet.*, *65*: 784–794, 1999.
 23. Berwick, M., and Vineis, P. Markers of DNA repair and susceptibility to cancer in humans: an epidemiologic review. *J. Natl. Cancer Inst. (Bethesda)*, *92*: 874–897, 2000.
 24. Wu, X., Gu, J., Patt, Y., Hassan, M., Spitz, M. R., Beasley, R. P., and Hwang, L. Y. Mutagen sensitivity as a susceptibility marker for human hepatocellular carcinoma. *Cancer Epidemiol. Biomark. Prev.*, *7*: 567–570, 1998.
 25. Shen, M.-J., Jones, I., and Mohrenweiser H. W. Non-conservative amino acid substitutions exist at polymorphic frequency in DNA repair genes. *Cancer Res.*, *58*: 604–608, 1998.
 26. Nickerson, D. A., Tobe, V. O., and Taylor, S. L. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res.*, *25*: 2745–2751, 1997.
 27. Collins F. S., Brooks L. D., and Chakravarti, A. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.*, *8*: 1229–1231, 1998.
 28. Fritsche, E., Pittman, G. S., and Bell, D. A. Localization, sequence analysis and ethnic distribution of a 96-bp insertion in the promoter of the human *CYP2E1* gene. *Mutat. Res.*, *432*: 1–5, 2000.
 29. Duell, E. J., Wiencke, J. K., Cheng, T. J., Varkonyi, A., Zuo, Z. F., Ashok, T. D., Mark, E. J., Wain, J. C., Christiani, D. C., and Kelsey, K. T. Polymorphisms in the DNA repair genes *XRCC1* and *ERCC2* and biomarkers of DNA damage in human blood mononuclear cells. *Carcinogenesis (Lond.)*, *21*: 965–971, 2000.
 30. Lunn, R. M., Langlois R. G., Hsich, L. L., Thompson, C. L., and Bell, D. A. *XRCC1* polymorphisms: effects on aflatoxin B₁ DNA adducts and glycoprotein A variant frequency. *Cancer Res.*, *59*: 2557–2561, 1999.
 31. Hu, J. J., Smith, T. R., Miller, M. S., Mohrenweiser, H. W., Golden, A., and Case, D. Amino acid variants of *APE1* and *XRCC1* genes associated with ionizing radiation sensitivity. *Carcinogenesis (Lond.)*, *22*: 917–922, 2001.
 32. Risler, J. L., Delorme, M. O., Delacroix, H., and Henaut, A. Amino acid substitutions in structurally related proteins: a pattern recognition approach. *J. Mol. Biol.*, *204*: 1019–1029, 1998.
 33. Smith, R. A., and Smith, T. F. Automatic generation of primary sequence patterns from a set of related protein sequences. *Proc. Natl. Acad. Sci. USA*, *87*: 118–122, 1990.
 34. Henikoff, S., and Henikoff, J. G. Amino acid substitution matrices. *Adv. Protein Chem.*, *54*: 73–97, 2000.
 35. Struwing, J. P., Hartge, P., Wacholder, S., Baker, S. M., Berlin, M., McAdams, M., Timmerman, M. M., Brody, L. C., and Tucker, M. A. The risk of cancer associated with specific mutations of *BRCA1* and *BRCA2* among Ashkenazi Jews. *N. Engl. J. Med.*, *336*: 1401–1408, 1997.
 36. Ford, B. N., Ruttan, C. C., Kyle, V. L., Brackley, M. E., and Glickman, B. W. Identification of single nucleotide polymorphisms in human DNA repair genes. *Carcinogenesis (Lond.)*, *21*: 1977–1981, 2000.
 37. Butkiewicz, D., Rusin, M., Harris, C. C., and Chorazy, M. Identification of four single nucleotide polymorphisms in DNA repair genes: *XPA* and *XPB (ERCC3)* in Polish population. *Hum. Mutat.*, *15*: 577–578, 2000.
 38. Thorstenson, Y. R., Shen, P., Tusher, V. G., Wayne, T. L., Davis, R. W., Chu, G., and Oefner, P. J. Global analysis of ATM polymorphism reveals significant functional constraint. *Am. J. Hum. Genet.*, *69*: 396–412, 2001.
 39. Broughton, B. C., Steingrimsdottir, H., and Lehmann, A. R. Five polymorphisms in the coding sequence of the xeroderma pigmentosum group D gene. *Mutat. Res.*, *362*: 209–211, 1996.
 40. Fan, F., Liu, C., Tavare, S., and Arnheim, N. Polymorphisms in the human DNA repair gene *XPF*. *Mutat. Res.*, *406*: 115–120, 1999.
 41. Bell, D. W., Wahrer, D. C., Kang, D. H., MacMahon, M. S., FitzGerald, M. G., Ishioka, C., Isselbacher, K. J., Krainer, M., and Haber, D. A. Common nonsense mutations in *RAD52*. *Cancer Res.*, *59*: 3883–3888, 1999.
 42. Khan, S. G., Metter, E. J., Tarone, R. E., Bohr, V. A., Grossman, L., Hedayati, M., Bale, S. J., Emmert, S., and Kraemer, K. H. A new xeroderma pigmentosum Group C poly(AT) insertion/deletion polymorphism. *Carcinogenesis (Lond.)*, *21*: 1821–1825, 2000.
 43. Orimo, H., Nakajima, E., Yamamoto, M., Ikejima, M., Emi, M., and Shimada, T. Association between single nucleotide polymorphisms in the *hMSH3* gene and sporadic colon cancer with microsatellite instability. *J. Hum. Genet.*, *45*: 228–230, 2000.
 44. Ma, X., Qianren, J., Forsti, A., Hemminki, K., and Kumar, R. Single nucleotide polymorphism analyses of the human proliferating cell antigen (*PCNA*) and flap endonuclease (*FEN1*) genes. *Int. J. Cancer*, *88*: 938–942, 2000.
 45. Schmutte, C., Tomblin, G., Rhiem, K., Sadoff, M. M. Schmutzler, R., von Deimling, A., and Fishel, R. Characterization of the human *RAD51* genomic locus and examination of tumors with 15q14-15 loss of heterozygosity (LOH). *Cancer Res.*, *59*: 4564–4569, 1999.
 46. Kato, M., Yano, K., Matsuo, F., Salto, H., Katagiri, T., Kurumizaka, H., Yoshimoto, M., Kasumi, F., Akiyama, F., Sakamoto, G., Nagawa, H., Nakamura, Y., and Miki, Y. Identification of *RAD51* alteration in patients with bilateral breast cancer. *J. Hum. Genet.*, *45*: 133–137, 2000.
 47. Wang, W. W., Sprudle, A. B., Kolachana, P., Bove, B., Modan, B., Ebbers, S. M., Suthers, G., Tucker, M. A., Kaufman, D. J., Doody, M. M., Tarone, R. E., Daly, M., Levavi, H., Pierce, H., Chetrit, A., Yechezkel, G. H., Chenevix-Trench, G., Offit, K., Godwin, A. K., and Struwing, J. P. A single nucleotide polymorphism in the 5' untranslated region of *RAD51* and risk of cancer among *BRCA1/2* mutation carriers. *Cancer Epidemiol. Biomark. Prev.*, *10*: 955–960, 2001.
 48. Cambien, F., Poirier, O., Nicaud, V., Herrmann, S. M., Mallet, C., Ricard, S., Behague, I., Hallet, V., Blanc, H., Loukaci, V., Thillet, J., Evans, A., Ruidavets, J. B., Arveiler, D., Luc, G., and Tiret, L. Sequence diversity in 36 candidate genes for cardiovascular disorders. *Am. J. Hum. Genet.*, *65*: 183–191, 1999.
 49. Halushka, M., K. Fan, J. B., Bentley, K., Hsie, L., Shen, N., Weder, A., Cooper, R., Lipshutz, R., and Chakravarti, A. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat. Genet.*, *22*: 239–247, 1999.
 50. Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C. R., Lim, E. P., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G. Q., and Lander, E. S. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.*, *22*: 231–238, 1999.
 51. Ohnishi, Y., Tanaka, T., Yamada, R., Suematsu, K., Minami, M., Fujii, K., Hoki, N., Kodama, K., Nagata, S., Hayashi, T., Kinoshita, N., Sato, H., Kuzuya, T., Takeda, H., Hori, M., and Nakamura, Y. Identification of 187 single nucleotide polymorphisms (SNPs) among 41 candidate genes for ischemic heart disease in the Japanese population. *Hum. Genet.*, *106*: 288–292, 2000.
 52. Yamada, R., Tanaka, T., Ohnishi, Y., Suematsu, K., Minami, M., Seki, T., Yukioka, M., Maeda, A., Murata, N., Saiki, O., Teshima, R., Kudo, O., Ishikawa, K., Ueyosi, A., Tateishi, H., Inaba, M., Goto, H., Nishizawa, Y., Tohma, A. S., Ochi, T., Yamamoto, K., and Nakamura, Y. Identification of 142 single nucleotide polymorphisms in 41 candidate genes for rheumatoid arthritis in the Japanese population. *Hum. Genet.*, *106*: 293–297, 2000.
 53. Nickerson, D. A., Taylor, S. L., Weiss, K. M., Clark, A. G., Hutchinson, R. G., Stengard, J., Salomaa, V., Vartiainen, E., Boerwinkle, E., and Sing, C. F. DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nat. Genet.*, *19*: 233–240, 1998.
 54. Rieder, M. J., Taylor, S. L., Clark, A. G., and Nickerson, D. A. Sequence variation in the human angiotensin converting enzyme. *Nat. Genet.*, *22*: 59–62, 1999.
 55. Hayward, C., Colville, S., Swingler, R. J., and Brock, D. J. Molecular genetic analysis of the APEX nuclease gene in amyotrophic lateral sclerosis. *Neurology*, *52*: 1899–1901, 1999.
 56. Olkowiak, Z. L. AP endonuclease in patients with amyotrophic lateral sclerosis. *Neuroreport*, *26*: 239–242, 1998.
 57. Hadi, M., Coleman, M., Fidelis, K., Mohrenweiser, H. W., and Wilson, D. W., III. Functional characterization of Ape1 variants identified in the human population. *Nucleic Acids Res.*, *28*: 3871–3879, 2000.
 58. Spitz, M. R., Wu, X., Wang, Y., Wang, L. E., Shete, E. S., Amos, C. I., Guo, Z., Lei, L., Mohrenweiser, H., and Wei, Q. Modulation of nucleotide excision repair capacity by *XPD* polymorphisms in lung cancer patients. *Cancer Res.*, *61*: 1354–1357, 2001.
 59. Qiao, Y. L., Spitz, M. R., Sheng, H., Guo, Z., Shete, S., Hadeyati, M., Grossman, L., Kraemer, K. H., Mohrenweiser, H., and Wei, Q. Modulation of repair of ultraviolet damage in the host-cell reactivation assay by polymorphic *XPC* and *XPD/ERCC2* genotypes. *Carcinogenesis (Lond.)*, *23*: 295–299, 2002.
 60. Pritchard, J. K. Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet.*, *69*: 124–137, 2001.
 61. Tabor, H. K., Risch, N. J., and Myers, R. M. Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat. Rev. Genet.*, *3*: 1–7, 2002.
 62. Sunyaev, S. R., Lathe, W. C., III, Ramensky, V. E., and Bork, P. SNP frequencies in human genes: an excess of rare alleles and differing modes of selection. *Trends Genet.*, *16*: 335–337, 2000.
 63. Nelson, H. H., Kelsey, K. T., Mott, L. A., and Karagas, M. R. The *XRCC1* Arg399Gln polymorphism, sunburn, and non-melanoma skin cancer: evidence for gene-environment interaction. *Cancer Res.*, *62*: 152–155, 2002.
 64. Pharoah, P. D. P., Antoniou, A., Bobrow, M., Zimmern, R. L., Easton, D. F., and Ponder, B. A. J. Polygenic susceptibility to breast cancer and implications for prevention. *Nat. Genet.*, *31*: 33–36, 2002.

Cancer Epidemiology, Biomarkers & Prevention

AACR American Association
for Cancer Research

Identification of 127 Amino Acid Substitution Variants in Screening 37 DNA Repair Genes in Humans

Harvey W. Mohrenweiser, Tong Xi, Johana Vázquez-Matías, et al.

Cancer Epidemiol Biomarkers Prev 2002;11:1054-1064.

Updated version Access the most recent version of this article at:
<http://cebp.aacrjournals.org/content/11/10/1054>

Cited articles This article cites 62 articles, 12 of which you can access for free at:
<http://cebp.aacrjournals.org/content/11/10/1054.full#ref-list-1>

Citing articles This article has been cited by 19 HighWire-hosted articles. Access the articles at:
<http://cebp.aacrjournals.org/content/11/10/1054.full#related-urls>

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://cebp.aacrjournals.org/content/11/10/1054>.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.