Contents lists available at ScienceDirect

Mutation Research/Genetic Toxicology and Environmental Mutagenesis

journal homepage: www.elsevier.com/locate/gentox
Community address: www.elsevier.com/locate/mutres

Bioinformatic processing to identify single nucleotide polymorphism that potentially affect Ape1 function

Eizadora T. Yu[a], Masood Z. Hadi[b,*]

[a] Biosystems Research Department, Sandia National Laboratories, Livermore, CA 94551-0969, USA
[b] Biomass Conversion Technologies Department, Sandia National Laboratories, Livermore, CA 94551-0969, USA

ARTICLE INFO

Article history:
Received 24 June 2010
Accepted 30 June 2010
Available online xxx

Keywords:
BER
APE1 cSNPS
Disease susceptibility
Variation

ABSTRACT

Inactivation of DNA damage response mechanisms is associated with several disease syndromes, including cancer, aging and neurodegeneration. A major corrective pathway for alkylation or oxidative DNA damage is base excision repair (BER). As part of an effort to identify variation in DNA repair genes, we used the expressed sequence tag (EST) database to identify amino acid variation in Ape1, an essential gene in the BER repair pathway. Nucleotide substitutions were considered valid only if the amino acid changes were observed in at least two independent EST sequencing runs (i.e. two independent EST reports). In total eighty amino acid variants were identified for the Ape1 gene. Using software tools SIFT and PolyPhen, which predict impacts of amino acid substitutions on protein structure and function, twenty-six variants were predicted by both algorithms to be deleterious to protein function. Majority of these intolerant mutations such as V206C and F240S, lie within the core of the protein and may affect the stability and folding of Ape1, or in the case of N212H, N212K, and Y171N, are close to the enzyme's active site and could drastically affect its function. A few of the intolerant mutations, i.e., G178V and E217R, are surface residues and are far from the active site, and as such, the predicted effect on Ape1 stability or function is not evident. These variants are reagents for further protein function studies and molecular epidemiology studies of cancer susceptibility.

1. Introduction

Whole genome association studies have recently been proposed as a powerful approach in order to detect numerous subtle genetic effects that may underlie susceptibility to genotoxic exposures as well as common diseases [1–3]. Unlike linkage studies, which look for co-inheritance of chromosomal regions with disease families, association studies look at differences in the frequency of genetic variation between unrelated individuals and controls. Such studies have been used to test the involvement of candidate genes in disease and to refine the location of disease genes in regions identified by linkage. Improved techniques for high throughput identification and genotyping of polymorphism in open reading frames offer the possibility of extending this approach to understand and characterize the function and susceptibility of the human genome.

The base excision repair (BER) pathway is involved in the correction of DNA modifications that arise either spontaneously or from attack by endogenous or exogenous sources of exposure [4,5]. These modifications may arise spontaneously or from replication errors or through chemical modification by oxidation or alkylation. Anti-cancer agents and various environmental mutagens generate many of these types of lesions. BER involves the concerted effort of several repair proteins that recognize and excise specific DNA damages, working to replace the damaged moiety with "normal" DNA (Fig. 1) [6,7]. Typically, the first step in BER involves the removal of an inappropriate base by a DNA glycosylase. The abasic site that is produced by DNA glycosylase activity is subsequently recognized by an apurinic/apyrimidinic (AP) endonuclease (Ape1), which incises the phosphodiester backbone of DNA immediately 5′ to the lesion, leaving a strand break with a normal 3′-hydroxyl group and a non-conventional 5′-abasic residue. At this stage of the repair, mammalian BER can be directed into one of two sub-pathways depending on the ends of the substrate. The "Short-patch BER" pathway proceeds with DNA polymerase β (Polβ) removing the 5′-abasic residue and filling in the single nucleotide gap. The alternative "long-patch" BER pathway entails the replacement of more than a single nucleotide (∼7–12 nucleotides), is PCNA-dependent (or stimulated) and requires and requires FEN1 to excise the flap-like structure produced by DNA polymerase strand displacement (most frequently executed by DNA polymerase δ or ξ). In either scenario, DNA Ligase I or a complex of XRCC1 and Ligase

Abbreviations: BER, base excision repair; Ape1, apurinic/apyrimidinicendonuclease.

[*] Corresponding author. Tel.: +1 925 294 4893; fax: +1 925 294 3020.
E-mail address: mzhadi@sandia.gov (M.Z. Hadi).

1383-5718/$ – see front matter © 2010 Published by Elsevier B.V.
doi:10.1016/j.mrgentox.2010.06.015

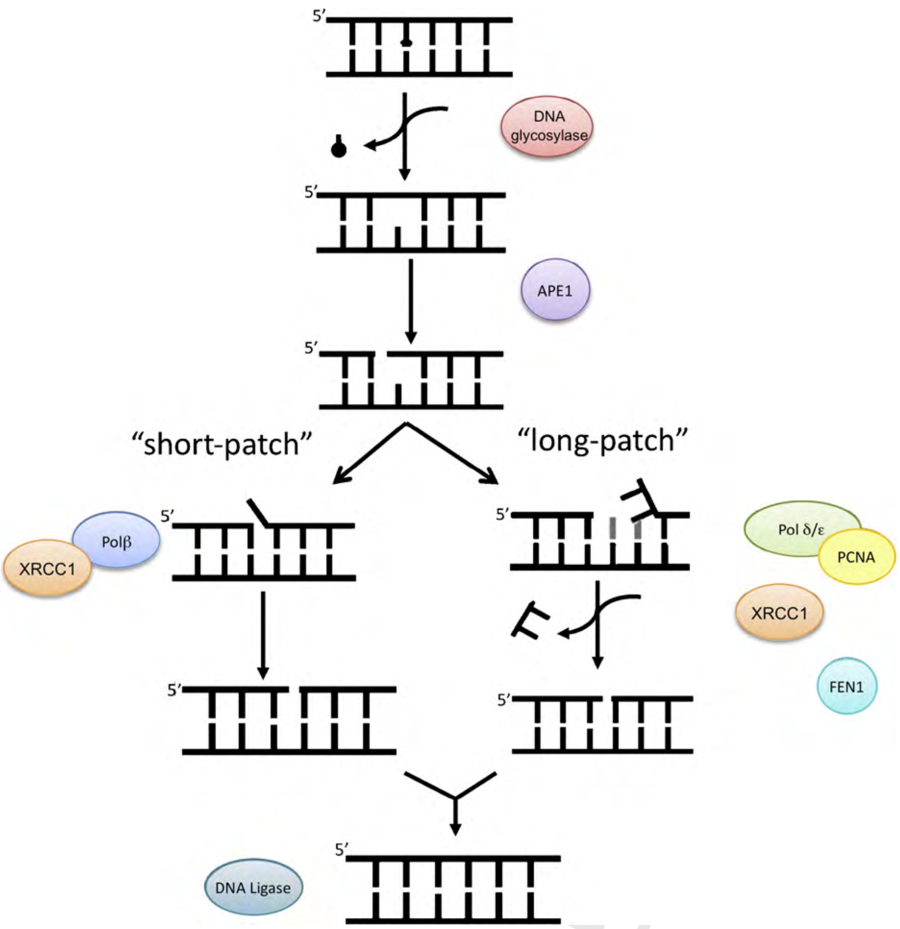


Fig. 1. Representation of the base excision repair pathway. In a highly co-ordinated fashion, damaged DNA in the form of oxidized or alkylated bases are recognized and removed by DNA glycosylase. The abasic site product is a substrate for AP endonuclease (Ape1), and is removed by incising the backbone 5′ to the lesion. Depending on various factors such as lesion type and cell cycle stage, the lesions are either repaired by "short-patch", where DNA Polβ removes the abasic residue and fills in the single nucleotide gap or "Long-patch" repair entails the replacement of 2–10 nucleotides by DNA δ/ε and requires PCNA and FEN1 proteins. XRCC1 has been shown to be an accessory protein for the sub-pathways. Finally, DNA ligase seals the nick and concludes the repair.

III seals the nick and completes BER restoring DNA to its normal state.

Given the known relationship of DNA repair to cancer and environmental exposures [3,4], the polymorphic variants identified have the potential to be population cancer risk factors because of the large number of individuals affected. Genes involved in DNA repair, such as those found in the BER pathway, are critical for protecting against mutations that lead to cancer and/or inherited genetic diseases [3,8–11]. Genes that are associated with an increased risk in sporadic cancer cases are referred to as "susceptibility" genes. Previous work to define the role of cancer susceptibility genes has often focused on variation in activity of carcinogen metabolizing enzymes with variant alleles that are associated with an increase in cancer risk [12,13]. The ability to measure DNA damage repair capacity *in vitro* has also provided insight into an individual's level of susceptibility [14–16]. The resequencing of DNA repair genes from several pathways have revealed variation in possible susceptibility genes [17,18], and several labs have demonstrated the feasibility of searching the public domain databases such as dbEST to identify single nucleotide polymorphisms (SNPs) [19–21]. To further molecular epidemiology studies that address the role of genetic variation of DNA repair genes in cancer susceptibility, we have screened the EST database to identify variation at the level of amino acid substitutions in the Ape1 gene. Single nucleotide polymorphisms resulting in coding errors (missense, in particular) were further examined using algorithms SIFT and PolyPhen to predict the impact of the amino acid substitution on enzyme structure [22–25]. These APE1 variants are candidates for future protein structure function studies as well as molecular epidemiology studies for understanding their role in disease susceptibility.

2. Materials and methods

2.1. Identification of variation within the Ape1 gene

To screen for amino acid substitutions in Ape1, we compared the amino acid sequence of Ape1 (accession #M92444) against the most recent EST database (build 130) using the tBlastn algorithm (http://www.ncbi.nlm.nih.gov/BLAST/) [26–28]. A total of 155 Ape1 or Ape1-related sequences were screened. Only those variants that were observed more than once were scored, as possible amino acid substitutions and the rest were not included in this study; nonsense (*) or ambiguous (B, Z, X, etc.) amino acid substitutions were also not considered for further analysis. Additional Ape1 amino acid variations were obtained from the SNP homepage at NCBI (http://www.ncbi.nlm.nih.gov/SNP/index.html). All scored variant DNA sequencing traces were verified by downloading the respective EST trace data from http://genome.wustl.edu/genomes and importing the traces in to the Genetic Annotation Initiative web server (http://www.chlc.org/gai/) for both amino acid substitution identification as well as SNP analysis.

2.2. Predicting impact of amino acid substitutions in the Ape1 variants

The possible impact of the amino acid substitutions in Ape1 variants were examined using PolyPhen and SIFT software [22–25]. The Ape1 amino acid sequence in GenBank accession #M92444 was used as wild type sequence. Solvent accessible surface areas of Ape1 residues were calculated by GETAREA [29] using the follow-

Table 1
Summary of Ape1SNPs within the exonicApe1 gene regions.

Gene region	dbSNP rs#ID	dbSNP allele	Allele frequency[a]	Wild type amino acid	Variant amino acid[b]	Codon position	Amino acid position
Exon 3	rs61757709	A/C	nd	K	Q	1	35
	rs34632023	G/A	0.025	G	E	2	39
	rs1048945	G/C	0.033	Q	H	3	51
	rs2307486	A/G	0.041	I	V	1	64
	rs61730854	C/T	nd	I	T	2	64
Exon 5	rs1130409	T/G	0.485	D	E	3	148
	rs1130410						
	rs33956927	G/A	0.027	G	R	1	241
	rs33956928						
	rs33956929						
	rs1803120	C/T	nd	P	S	1	311
	rs1803118	C/T	nd	A	V	2	317
	rs61757710	A/G	nd	T	T	3	233
	rs61757711						
	rs61757712						
	rs1065749	C/T	0.01	Y	Y	3	269
	rs4748	T/C	nd	L	L	1	286
	rs6172835	T/C	nd	Y	Y	3	315
	rs6172836						
	rs6172837						
	rs1130409	G/T			X		
	rs33956927	C/T			X		241
	rs4748	C/T		G	X		241
	rs4748	C/T		P	X		311

List of single nucleotide polymorphisms found in the Ape1 gene exon regions. SNP data taken from NCBI Entrez SNP database. Reference SNP (rs) ID numbers are listed along with the polymorphic allele.

[a] Allele frequency is noted except where it is not determined (nd).

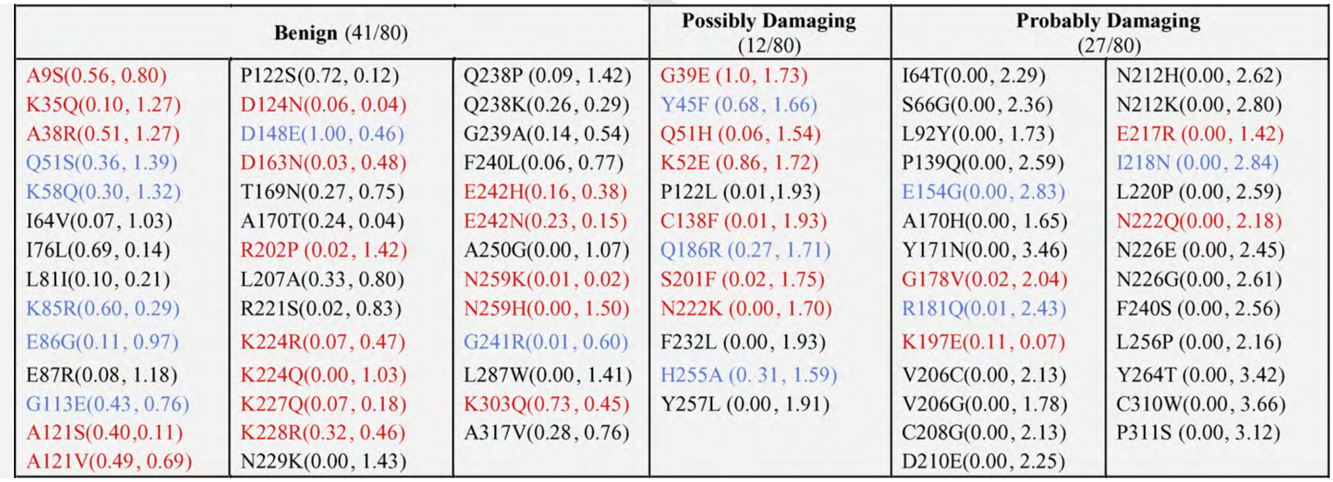
[b] Wild type and variant amino acid for the cSNPs are also provided, variations resulting in truncations are marked with X.

ing crystal structures (PDB ID: 1E9N, 1DE8, 1DE9) [30,31]. Molecular models for the Ape1 variants were built using Pymol [32] to visualize the potential structural changes of the different Ape1 variants.

3. Results and discussion

Genetic factors contribute to all human disease, conferring susceptibility, resistance, and efficacy of treatment as well as interactions with the environment. Gene identification technologies that emerged during the human genome project and since are less likely to successfully identify multiple genes underlying complex human diseases. However, association studies that are comparing the prevalence of markers in diseased individuals versus normal individuals can provide some clues to winnowing down the list of candidate markers that can be evaluated biochemically and subsequently in animal studies. The association of DNA repair pathways and genetic predisposition to cancer has been documented for nucleotide excision repair and mismatch repair [17]. Similarly genes involved in the BER pathway are the first line of defense against mutations that lead to cancer and other diseases. These mutations can be silent, or in other cases, cause alterations in transcription/translation (mutations in the promoter site or the

Table 2
Summary of Ape1 amino acid variants found in the EST database.



Benign (41/80)			Possibly Damaging (12/80)	Probably Damaging (27/80)	
A9S(0.56, 0.80)	P122S(0.72, 0.12)	Q238P (0.09, 1.42)	G39E (1.0, 1.73)	I64T(0.00, 2.29)	N212H(0.00, 2.62)
K35Q(0.10, 1.27)	D124N(0.06, 0.04)	Q238K(0.26, 0.29)	Y45F (0.68, 1.66)	S66G(0.00, 2.36)	N212K(0.00, 2.80)
A38R(0.51, 1.27)	D148E(1.00, 0.46)	G239A(0.14, 0.54)	Q51H (0.06, 1.54)	L92Y(0.00, 1.73)	E217R (0.00, 1.42)
Q51S(0.36, 1.39)	D163N(0.03, 0.48)	F240L(0.06, 0.77)	K52E (0.86, 1.72)	P139Q(0.00, 2.59)	I218N (0.00, 2.84)
K58Q(0.30, 1.32)	T169N(0.27, 0.75)	E242H(0.16, 0.38)	P122L (0.01,1.93)	E154G(0.00, 2.83)	L220P (0.00, 2.59)
I64V(0.07, 1.03)	A170T(0.24, 0.04)	E242N(0.23, 0.15)	C138F (0.01, 1.93)	A170H(0.00, 1.65)	N222Q(0.00, 2.18)
I76L(0.69, 0.14)	R202P (0.02, 1.42)	A250G(0.00, 1.07)	Q186R (0.27, 1.71)	Y171N(0.00, 3.46)	N226E (0.00, 2.45)
L81I(0.10, 0.21)	L207A(0.33, 0.80)	N259K(0.01, 0.02)	S201F (0.02, 1.75)	G178V(0.02, 2.04)	N226G(0.00, 2.61)
K85R(0.60, 0.29)	R221S(0.02, 0.83)	N259H(0.00, 1.50)	N222K (0.00, 1.70)	R181Q(0.01, 2.43)	F240S (0.00, 2.56)
E86G(0.11, 0.97)	K224R(0.07, 0.47)	G241R(0.01, 0.60)	F232L (0.00, 1.93)	K197E(0.11, 0.07)	L256P (0.00, 2.16)
E87R(0.08, 1.18)	K224Q(0.00, 1.03)	L287W(0.00, 1.41)	H255A (0. 31, 1.59)	V206C(0.00, 2.13)	Y264T (0.00, 3.42)
G113E(0.43, 0.76)	K227Q(0.07, 0.18)	K303Q(0.73, 0.45)	Y257L (0.00, 1.91)	V206G(0.00, 1.78)	C310W(0.00, 3.66)
A121S(0.40,0.11)	K228R(0.32, 0.46)	A317V(0.28, 0.76)		C208G(0.00, 2.13)	P311S (0.00, 3.12)
A121V(0.49, 0.69)	N229K(0.00, 1.43)			D210E(0.00, 2.25)	

Eighty Ape1 variants are classified into three categories: benign, possibly damaging, or probably damaging, based on the PolyPhen prediction. SIFT and PolyPhen scores for each mutant are also reported in parenthesis (SIFT and PSIC score differences, respectively). Surface accessibility areas of the wild type residues were calculated using the GETAREA program and the PDB file 1E9N. The accessibilities of the variants are highlighted (red for solvent/surface accessible, blue for intermediate, and black for buried sites). The crystal structure 1E9N does not contain coordinates for residues 1–43, and the relative accessibilities of these residues were derived from experimental data [31,52].

Please cite this article in press as: E.T. Yu, M.Z. Hadi, Bioinformatic processing to identify single nucleotide polymorphism that potentially affect Ape1 function, Mutat. Res.: Genet. Toxicol. Environ. Mutagen. (2010), doi:10.1016/j.mrgentox.2010.06.015

Table 3
Predicted amino acid variations in Ape1 resulting in probably damaging effects.

Mutation	Predicted effect on Ape1 protein structure and function
I64T	Hydrophobicity change at buried site.
S66G	Close contact with active site residue D210 (<5 Å). Loss of hydrogen bonding interactions that stabilize β core of protein.
L92Y A170H	Hydrophobicity change at buried site.
P139Q	Close contact with residue H116 (<5 Å).
E154G	Loss of hydrogen bonding interaction with R181 and R156. These arginine residues are near DNA binding site (<5 Å).
Y171N	Y171 is a catalytically important residue, and mutation will disrupt metal ligand and DNA binding contacts.
G178V	May potentially affect the position of the adjacent residue R177, which is crucial for binding DNA.
R181Q	Loss of hydrogen bonding interaction with E154. The mutation posits this residue closer to the scissile bond in the DNA backbone.
K197E	Loss of *in vivo* acetylation site (N-6 acetyl lysine at K197).
V206C	Charge change at buried site.
V206G Y264T	Cavity creation at buried site.
C208G	No obvious difference in structure, the prediction is based on sequence alignment. However, the side chain is near the redox-functional C65 residue (∼4 Å).
D210E	D210 is catalytically important residue. Mutation will disrupt annotated functional site.
N212HN212K	N212 is catalytically important residue. Mutation will disrupt annotated functional site.
E217R	Disruption of ligand binding site.
I218N	No obvious difference in structure, the prediction is based on sequence alignment.
L220P	Hydrophobicity change at buried site, and possible close contact with DNA (∼5 Å).
N222Q	Close contact with DNA (<5 Å) and N222 is implicated in DNA Polβ interaction site.
N226EN226G	Closest contact with DNA (<5 Å).
F240S	Decreased hydrophobicity and cavity creation at the buried site.
C310W	Hydrophobicity change and overpacking at buried site. Potential contact with active site residue H309.
P311S	Disruption of hydrophobic interactions with W67. Introduction of repulsive interactions with E87, C65, T313 (<4.7 Å)

A closer look at the structural effects of the mutations listed as probably damaging in Table 2. The substitutions were mapped to seventeenApe1 protein 3D structures currently found in the PDB. Parameters such as accessible surface area, hydrophobicity, charge and volume are calculated by PolyPhen and prediction rules on such properties can be found in the PolyPhen website (http://genetics.bwh.harvard.edu/pph/pph_help_text.html#OverviewStructure) [24,25,38]. Steric clashes within the protein subunits, ligands defined as heteroatoms and functional sites are also checked and potential contacts (distances between atoms are <5 Å) are also reported. All predicted substitutions were also manually checked with select 3D structures (PDB IDs: 1E9N, 1DE8, 1DE9) [30,31].

untranslated region) or within the protein structure itself (caused by missense mutations). Ape1 gene knockouts or truncations (deletions) have been shown to cause embryonic lethality in mice and triggered apoptosis in human cells [33,34], underscoring the importance of BER enzymes in viability and survival. However, other Ape1mutations are tolerated, and maybe in combination with other genetic variations within Ape1 or in other genes may affect how humans develop diseases. These Ape1 variants may be used as suitable biomarkers for detecting people at risk for certain diseases, and could potentially be used to identify them before the onset of disease progression. However, determining the clinical significance of every genetic variation and performing biochemical, epidemiology assays for each Ape1 variant can be a costly endeavor, and therefore, *in silico* pre-screening methods are very useful to tame this seemingly intractable problem [35].

3.1. *The sequence database presents a rich source of candidate Ape1 amino acid variants*

Bioinformatics constitutes an important methodology to rapidly identify sequence variations in databases. There are 88 SNPs in the NCBI SNP database that are associated with the Ape1 gene, 12 of which fall within the intronic regions, and could result in changes in mRNA processing and transcript half-life, stability and transcription regulation. Fifty-five are in exonic regions, of which 4 result in truncation of the protein and 9 translate into amino acid substitutions (Table 1). Searching the EST database (dbEST) using tBlastn also provided an effective method to rapidly screen and identify more amino acid variants [19]. A total of 155 Ape1 and Ape1-related sequences were screened and we have identified 80 unique amino acid substitution variants from the EST database (Table 2). Three of the variants (D148E, I64V, and G241R) were found in both EST and SNP databases. While the EST database presents an immense source of sequence variation information, sequence verification is crucial to be able to differentiate true positives versus sequencing errors. Resequencing and genotyping studies involving populations of cancer/disease patients have also produced other unique Ape1 that were not found in our EST database search results [18,36,37].

3.2. *The effects of Ape1 amino acid variants in protein structure are examined in silico using robust algorithms*

The SIFT (sorting intolerant from tolerant) [22,23] and PolyPhen programs (polymorphism phenotyping) [24,25] are generally used to predict such effects. The SIFT program uses multiple sequence alignment information to predict tolerant and deleterious substitutions. It compares evolutionarily conserved residues and evaluates specific amino acid location and their ability to tolerate replacement by different classes of amino acids. The SNP introduced amino acid is then compared to a range of tolerated amino acids based on its structure and the likelihood of the change disrupting protein structure is estimated. For example, SIFT scores which are less than 0.05 are deemed to be intolerant variations in the sequence, and scores that are greater than 0.2 are assigned to be tolerable substitutions. PolyPhen makes predictions based on several sources of data including multiple sequence alignment. However the multiple sequence alignment is generated by position-specific independent counts (PSIC) software which assigns a score that is indicative of the probability of a given amino acid occurring at a particular position against any random position [38]. PolyPhen also uses information about the structure of the protein, and in the analysis and the structural information includes position within the protein, surface or interior, contributions to well-defined structural elements, helices or sheets, or location within the active site. In addition, PolyPhen also considers known salient structural features and available structural data from the PDB [39,40], to predict the effect of the substitution on measurable physical parameters such as solvent accessibility area changes, charge effects, and changes in molecular contacts, especially with functional sites. Based on the calculated alignment score and differences in structural parameters, the algorithm assigns the mutation as being “benign”, “possibly damaging”,

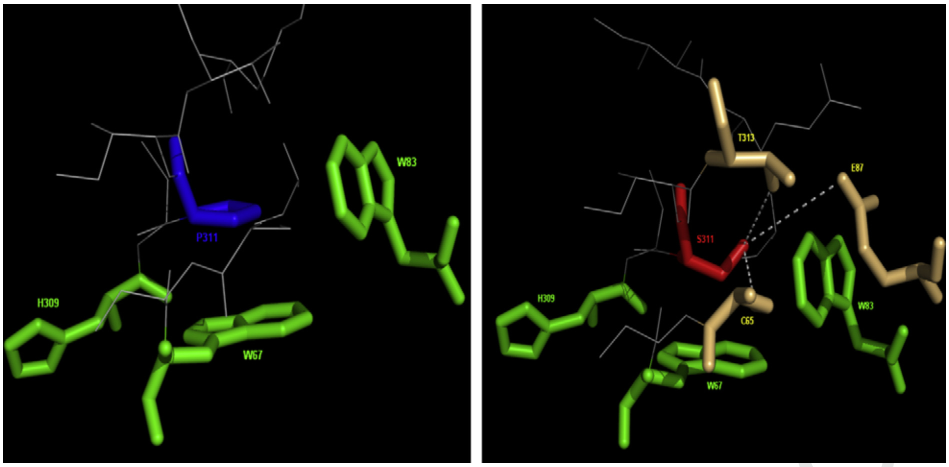


Fig. 2. Changes in local environment of Ape1 protein brought about by a P311S mutation. (A) The wild type proline residue (in blue) is flanked by hydrophobic residues W83 and W67 (in green). (B) Substitution of P311 residue to serine (red) presents a change in charge at the core, introducing repulsive interactions with residues E87, C65, and T313 (in orange), all of which lie within 5 Å of the S311 side chain. Figure created in Pymol using PDB 1DE8 [30,32].

and "probably damaging". PolyPhen reports its analysis in the PSIC scores between the native and mutant, and typically, a big change (>2.0) in the scores indicates that the substitution is rarely observed in the protein. Of the total eighty amino acid variants, SIFT predicted 42 (52.5%) to be intolerant mutations (with SIFT values <0.05), 24 (30%) to be tolerant (with SIFT values >0.2), the remaining 14 (17.5%) to be borderline or potentially intolerant mutations (Table 2). In contrast, PolyPhen predicted only 27 (33.75%) of these variants to be "probably damaging", 12 (15%) to be "possibly damaging", and 41 (51.25%) to be "benign" substitutions (Table 2). The basis for predicting the impact of mutations in these two algorithms are different, and we would expect that the outcomes to be in some ways, dissimilar. However, the mutations that overlap the two predictions should provide greatest reliability to behave similarly. Twenty Ape1 mutations are predicted to be both benign and tolerant by SIFT and PolyPhen, with the exception of one mutation (K197E), all the mutants that were predicted to be probably damaging were also predicted to be intolerant mutations (Table 2). Two cSNPs (I64T and P311S) were predicted to be deleterious mutations. A list of Ape1 amino acid substitutions with predicted deleterious effects are summarized in Table 3. Notably, there were more Ape1 variants that were consistently predicted to be intolerant than tolerant.

There are a few Ape1 variants that were predicted to be benign by PolyPhen but absolutely intolerant by SIFT (SIFT score = 0), i.e., K224Q, N229K, A250G, N259H, and L287W (Table 2). There is no direct way of evaluating the accuracy of these predictions made by SIFT and PolyPhen, as it is probable that the algorithms used different data sets. So, even if the predictions by these two programs were totally inconsistent for this set of mutants, these Ape1 variants still should be considered as candidates for SNP screening.

Frequently observed substitutions in dbEST (i.e., D148E, K224R, P122S, K227Q, G113E) were all predicted to be tolerable mutations. D148E is the most frequently occurring mutation found in the EST database and some clinical studies have shown that D148E mutation is unequivocally linked with certain cancers or at least have been observed in several cases [41–46]. However, purified, recombinant D148E protein has been shown to have no effect on the repair function of Ape1 [37]. It has been suggested that D148E mutant may have other reduced functions, i.e., in connection with possibly weakened protein–protein interactions involved in BER communications [37]. Four other cSNPS (K35Q, I64V, G241R, and A317V) were also predicted to be tolerable mutations. Similar to D148E, binding and incision activity of G241R mutant was biochemically assessed and found to be comparable to the wild type [37]. The rest of the cSNPs, G39E, Q51H, I64T and P311S were predicted to either have possible or probable damaging effects (Table 2). The change to threonine of I64 results in a hydrophobicity change at the core of the protein that could result in the destabilization of the β sheets. The G39E mutation alters the charge at a surface residue. In the case of the P311S variant, the serine residue presents a change in charge at the core, and could introduce repulsive interactions with residues E87, C65, and T313, all of which lie within 5 Å of P311 (Fig. 2). It is not known whether these Ape1 variants have impaired DNA binding and incision activity, but at least two of these mutations, Q51H and I64V, have been identified in cancer patients [36]. The molecular basis of this association is still unknown, and future biochemical characterization of the redox and DNA binding and incision activities of the cSNPs Q51H and I64V is warranted.

3.3. Majority of the Ape1 variants that were predicted to be deleterious map to the core of the protein

The single amino acid variations in Ape1 found from the database search map to 67 unique sites (residues) in the protein. Thirty-four of these sites are found in the interior of the protein and 21 are on the surface (Table 2). The remaining 12 had intermediate solvent accessible areas and were not categorized in either category. The majority of the mutations that were predicted to be damaging were internally located residues. This is not surprising as changes in the Ape1 protein core that result in decreased hydrophobicity, introduction of charge effects, and volume changes (i.e., cavity creation or over packing) may destabilize protein or worse, prevent proper folding in the first place (Table 3). There are a couple of variants that involve the critical catalytic residues and disrupt the active sites (Y171N, D210E, N212H, and N212K) [47,48]. The damaging effects of these variants are pretty evident, i.e., N212H and N212K mutations disrupt the functional site of Ape1, possibly competing for the magnesium ion cofactor, or making contacts with D210 (Fig. 3). The Y171N variant also has the potential to affect Ape1s repair function by disrupting the magnesium-binding site. The other mutations result in the absence of functional groups that participate in hydrogen bonding networks or hydrophobic interactions, and introduce alternative interacting networks which could also contribute to destabilizing the protein or altering its enzymatic activity (Table 3). K197 was identified as

Please cite this article in press as: E.T. Yu, M.Z. Hadi, Bioinformatic processing to identify single nucleotide polymorphism that potentially affect Ape1 function, Mutat. Res.: Genet. Toxicol. Environ. Mutagen. (2010), doi:10.1016/j.mrgentox.2010.06.015

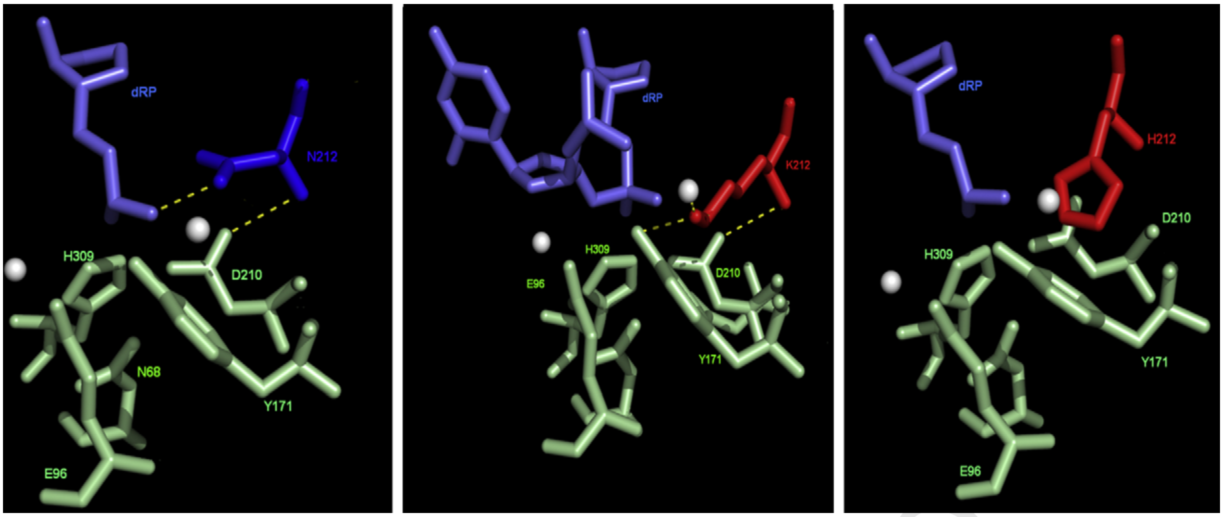


Fig. 3. Changes in local environment brought about by a N212K/N212H mutation. (A) The wild type asparagine residue (dark blue)contacts the phosphate backbone 5′ to the abasic DNA substrate (light blue) and is involved in coordinating the magnesium ion (white sphere). (B) Substitution of N212 residue tolysine (red) or (C) histidine (red) presents a disruption in metal coordination and potential steric clashes in the active site. Magnesium ions (white spheres) and nearby critical active site residues are shown (green). Figure created in Pymol using PDB 1DE8 [30,32].

an acetylation site in a recent proteomics study [49], and the K197E mutation precludes this post-translational modification, which could be important in protein-protein interactions and cellular signaling.

A few of these “damaging” mutations (G178V, K197E, E217R, and N222Q) are located on the surface of Ape1 protein. G187V and N222Q are in the DNA binding loops and contact the DNA strand that contains the abasic site. N222 has also been implicated in a putative interaction with DNA Polβ [50]. In a molecular dynamics (MD) simulation of the ternary complex of Ape1, DNA polymerase β and substrate DNA duplex, residue N222 was predicted to be one the residues involved in the interface between Ape1 and DNA Polβ. Whether these altered contacts actually destabilize the protein/DNA or protein/protein complex or alter redox and/or repair functions of Ape1 is unknown, and needs to be experimentally examined.

3.4. Ape1 variants of residues localized at the surface are potentially damaging

Unlike the set of probably damaging mutants, majority of which are localized in buried sites in the protein, there is no clear distinction in the accessibility/localization of the residues on the Ape1 variants that were predicted to be benign, as half were either surface or the buried sites. There are a number of surface-localized Ape1 variants that were predicted to be potentially or probably damaging (i.e., Q51H, K52E, C138F, S201F, N222K, G178V, E217R, N222Q) or SIFT-intolerant (i.e., D124N, D163N, R202P, K224R, K224Q, K227Q, N259K, and N259H) (Table 2). To some extent, as long as there are no major charge effects brought about by the amino acid substitutions, variants of surface-localized residues should not grossly affect the structure of the protein and should generally be predicted to be more tolerant mutations. Some of these mutated residues could constitute possible protein–protein interaction sites involved in signaling pathways. For example, Ape1 through its N-terminal domain (residues 1–35) has been shown to physically interact with the BER accessory protein XRCC1 [51], and the K35Q variant could play a role in this interaction. In addition, some of these Ape1 variants coincide with predicted interacting residues of Ape1 with DNA Polβ [50], which could be important in the hand-off of DNA substrates between the BER enzymes. Q186, Q238, R221, N222, and K224 account for up to 60% if the interface area in the MD simulations [50]. A recent footprinting study done on binary Ape1:DNA complexes as well as the ternary complex with DNA Polβ also implicatesApe1 residues K227 and K228 contacting the DNA duplex during BER progression, and the variants K227Q and K228R may have impaired and contribute to weakened interactions with the DNA substrate [52].

Most of the Ape1 amino acid substitutions found in dbEST were predicted to be tolerable mutations, however, it does not discount the fact that, there could be variations that when present in tandem or as multiple mutations, present undesirable changes in the molecular make-up of the enzyme. All in all, using our bioinformatics processing, we were able to find 3 of 9 validated Ape1 cSNPs together with over 80 new mutations that can be added to the expanding list of cSNPS that could be considered for future validation by biochemical, animal studies and clinical screening.

4. Conclusion

Molecular epidemiology and biochemical studies are the starting points for addressing the relationships of individual polymorphic variants to DNA repair capacity and cancer risk factor. These studies establish a framework for identifying and examining variation for follow-up functional impact. The timely identification of individuals that are susceptible to certain cancers or diseases is essential for cancer prevention. Biomarkers based on genetic variations in populations susceptible to certain diseases have been successfully identified and applied to screen for individuals at risk of developing certain types of cancer. Given that DNA repair and predisposition to cancer are intimately linked, developing simple approaches to identifying polymorphic variants of genes involved in DNA repair has the potential to quickly assessing plausible cancer risk factors. Screening public domain sequence databases represent a rapid way of identifying genetic variations in disease susceptibility genes. Moreover, with the wealth of sequences and 3D structural data available on DNA repair enzymes, bioinformatics programs that predict the putative effect of amino acid substitutions of the protein's structure and function can provide a rapid assessment of these variants. As a demonstration of this potential, we have used this combined bioinformatics approach to screen Ape1 sequences, and focused on sequence variations that resulted in non-synonymous polymorphisms. These variants constitute ratio-

Please cite this article in press as: E.T. Yu, M.Z. Hadi, Bioinformatic processing to identify single nucleotide polymorphism that potentially affect Ape1 function, Mutat. Res.: Genet. Toxicol. Environ. Mutagen. (2010), doi:10.1016/j.mrgentox.2010.06.015

nal candidate reagents for protein function studies and molecular epidemiology studies of cancer susceptibility. This process is also amenable to high throughput computation techniques and could be coupled with current sequencing technologies to better understand and characterize the function and susceptibility of the human genome.

Conflicts of interest

None.

Acknowledgments

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

References

[1] S. Istrail, G.G. Sutton, L. Florea, A.L. Halpern, C.M. Mobarry, R. Lippert, B. Walenz, H. Shatkay, I. Dew, J.R. Miller, M.J. Flanigan, N.J. Edwards, R. Bolanos, D. Fasulo, B.V. Halldorsson, S. Hannenhalli, R. Turner, S. Yooseph, F. Lu, D.R. Nusskern, B.C. Shue, X.H. Zheng, F. Zhong, A.L. Delcher, D.H. Huson, S.A. Kravitz, L. Mouchard, K. Reinert, K.A. Remington, A.G. Clark, M.S. Waterman, E.E. Eichler, M.D. Adams, M.W. Hunkapiller, E.W. Myers, J.C. Venter, Whole-genome shotgun assembly and comparison of human genome assemblies, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 1916–1921.

[2] E.C. Rouchka, W. Gish, D.J. States, Comparison of whole genome assemblies of the human genome, Nucleic Acids Res. 30 (2002) 5004–5014.

[3] E.D. Pleasance, R.K. Cheetham, P.J. Stephens, D.J. McBride, S.J. Humphray, C.D. Greenman, I. Varela, M.L. Lin, G.R. Ordonez, G.R. Bignell, K. Ye, J. Alipaz, M.J. Bauer, D. Beare, A. Butler, R.J. Carter, L. Chen, A.J. Cox, S. Edkins, P.I. Kokko-Gonzales, N.A. Gormley, R.J. Grocock, C.D. Haudenschild, M.M. Hims, T. James, M. Jia, Z. Kingsbury, C. Leroy, J. Marshall, A. Menzies, L.J. Mudie, Z. Ning, T. Royce, O.B. Schulz-Trieglaff, A. Spiridou, L.A. Stebbings, L. Szajkowski, J. Teague, D. Williamson, L. Chin, M.T. Ross, P.J. Campbell, D.R. Bentley, P.A. Futreal, M.R. Stratton, A comprehensive catalogue of somatic mutations from a human cancer genome, Nature 463 (2010) 191–196.

[4] D.M. Wilson, V.A. 3rd, Bohr, The mechanics of base excision repair and its relationship to aging and disease, DNA Repair (Amst.) 6 (2007) 544–559.

[5] T. Lindahl, Keynote: past, present, and future aspects of base excision repair, Prog. Nucleic Acid Res. Mol. Biol. 68 (2001) xvii–xxx.

[6] G.L Dianov, K.M. Sleeth, I.I. Dianova, S.L. Allinson, Repair of abasic sites in DNA, Mutat. Res. 531 (2003) 157–163.

[7] S. Mitra, I. Boldogh, T. Izumi, T.K. Hazra, Complexities of the DNA base excision repair pathway for repair of oxidative DNA damage, Environ. Mol. Mutagen. 38 (2001) 180–190.

[8] R.D. Kolodner, Mismatch repair: mechanisms and relationship to cancer susceptibility, Trends Biochem. Sci. 20 (1995) 397–401.

[9] M. Radman, I. Matic, J.A. Halliday, F. Taddei, Editing DNA replication and recombination by mismatch repair: from bacterial genetics to mechanisms of predisposition to cancer in humans, Philos. Trans. R. Soc. Lond. 347 (1995) 97–103.

[10] V.A. Bohr, DNA repair fine structure and its relations to genomic instability, Carcinogenesis 16 (1995) 2885–2892.

[11] T.D. Tlsty, A. Briot, A. Gualberto, I. Hall, S. Hess, M. Hixon, D. Kuppuswamy, S. Romanov, M. Sage, A. White, Genomic instability and cancer, Mutat. Res. 337 (1995) 1–7.

[12] F.J. Gonzalez, Genetic polymorphism and cancer susceptibility: fourteenth Sapporo Cancer Seminar, Cancer Res. 55 (1995) 710–715.

[13] M.R. Spitz, T.C. Hsu, X. Wu, J.J. Fueger, C.I. Amos, J.A. Roth, Mutagen sensitivity as a biological marker of lung cancer risk in African Americans, Cancer Epidemiol. Biomarkers Prev. 4 (1995) 99–103.

[14] K.J. Helzlsouer, E.L. Harris, R. Parshad, S. Fogel, W.L. Bigbee, K.K. Sanford, Familial clustering of breast cancer: possible interaction between DNA repair proficiency and radiation exposure in the development of breast cancer, Int. J. Cancer 64 (1995) 14–17.

[15] M.R. Spitz, R.S. McPherson, H. Jiang, T.C. Hsu, Z. Trizna, J.J. Lee, S.M. Lippman, F.R. Khuri, L. Steffen-Batey, R.M. Chamberlain, S.P. Schantz, W.K. Hong, Correlates of mutagen sensitivity in patients with upper aerodigestive tract cancer, Cancer Epidemiol. Biomarkers Prev. 6 (1997) 687–692.

[16] Q. Wei, L. Cheng, W.K. Hong, M.R. Spitz, Reduced DNA repair capacity in lung cancer patients, Cancer Res. 56 (1996) 4103–4107.

[17] M.R. Shen, I.M. Jones, H. Mohrenweiser, Nonconservative amino acid substitution variants exist at polymorphic frequency in DNA repair genes in healthy humans, Cancer Res. 58 (1998) 604–608.

[18] T. Xi, I.M. Jones, H.W. Mohrenweiser, Many amino acid substitution variants identified in DNA repair genes during human population screenings are predicted to impact protein function, Genomics 83 (2004) 970–979.

[19] K.H. Buetow, M.N. Edmonson, A.B. Cassidy, Reliable identification of large numbers of candidate SNPs from public EST data, Nat. Genet. 21 (1999) 323–325.

[20] Z. Gu, L. Hillier, P.Y. Kwok, Single nucleotide polymorphism hunting in cyberspace, Hum. Mutat. 12 (1998) 221–225.

[21] P. Taillon-Miller, Z. Gu, Q. Li, L. Hillier, P.Y. Kwok, Overlapping genomic sequences: a treasure trove of single-nucleotide polymorphisms, Genome Res. 8 (1998) 748–754.

[22] P.C. Ng, S. Henikoff, SIFT: predicting amino acid changes that affect protein function, Nucleic Acids Res. 31 (2003) 3812–3814.

[23] P.C. Ng, S. Henikoff, Predicting deleterious amino acid substitutions, Genome Res. 11 (2001) 863–874.

[24] S. Sunyaev, V. Ramensky, P. Bork, Towards a structural basis of human non-synonymous single nucleotide polymorphisms, Trends Genet. 16 (2000) 198–200.

[25] V. Ramensky, P. Bork, S. Sunyaev, Human non-synonymous SNPs: server and survey, Nucleic Acids Res. 30 (2002) 3894–3900.

[26] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.

[27] W. Gish, D.J. States, Identification of protein coding regions by database similarity search, Nat. Genet. 3 (1993) 266–272.

[28] M.S. Boguski, T.M. Lowe, C.M. Tolstoshev, dbEST—database for "expressed sequence tags", Nat. Genet. 4 (1993) 332–333.

[29] R. Fraczkiewicz, W. Braun, Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules, J. Comput. Chem. 19 (1998) 319–333.

[30] C.D. Mol, D.J. Hosfield, J.A. Tainer, Abasic site recognition by two apurinic/apyrimidinic endonuclease families in DNA base excision repair: the 3′ ends justify the means, Mutat. Res. 460 (2000) 211–229.

[31] P.T. Beernink, B.W. Segelke, M.Z. Hadi, J.P. Erzberger, D.M. Wilson 3rd, B. Rupp, Two divalent metal ions in the active site of a new crystal form of human apurinic/apyrimidinic endonuclease, Ape1: implications for the catalytic mechanism, J. Mol. Biol. 307 (2001) 1023–1034.

[32] W.L. DeLano, The PyMOL Molecular Graphics System, DeLano Scientific, Palo Alto, CA, USA, 2002, http://www.pymol.org.

[33] H. Fung, B. Demple, A vital role for Ape1/Ref1 protein in repairing spontaneous DNA damage in human cells, Mol. Cell 17 (2005) 463–470.

[34] T. Izumi, D.B. Brown, C.V. Naidu, K.K. Bhakat, M.A. Macinnes, H. Saito, D.J. Chen, S. Mitra, Two essential but distinct functions of the mammalian abasic endonuclease, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 5739–5743.

[35] A. MacAuley, W.C. Ladiges, Approaches to determine clinical significance of genetic variants, Mutat. Res. 573 (2005) 205–220.

[36] M. Pieretti, N.H. Khattar, S.A. Smith, Common polymorphisms and somatic mutations in human base excision repair genes in ovarian and endometrial cancers, Mutat. Res. 432 (2001) 53–59.

[37] M.Z. Hadi, M.A. Coleman, K. Fidelis, H.W. Mohrenweiser, D.M. Wilson 3rd, Functional characterization of Ape1 variants identified in the human population, Nucleic Acids Res. 28 (2000) 3871–3879.

[38] S.R. Sunyaev, F. Eisenhaber, I.V. Rodchenkov, B. Eisenhaber, V.G. Tumanyan, E.N. Kuznetsov, PSIC: profile extraction from sequence alignments with position-specific counts of independent observations, Protein Eng. 12 (1999) 387–394.

[39] H. Berman, K. Henrick, H. Nakamura, Announcing the worldwide Protein Data Bank, Nat. Struct. Biol. 10 (2003) 980.

[40] F.C. Bernstein, T.F. Koetzle, G.J. Williams, E.F. Meyer Jr., M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, The Protein Data Bank: a computer-based archival file for macromolecular structures, J. Mol. Biol. 112 (1977) 535–542.

[41] K. De Ruyck, M. Szaumkessel, I. De Rudder, A. Dehoorne, A. Vral, K. Claes, A. Velghe, J. Van Meerbeeck, H. Thierens, Polymorphisms in base-excision repair and nucleotide-excision repair genes in relation to lung cancer risk, Mutat. Res. 631 (2007) 101–110.

[42] T. Farkasova, S. Gurska, V. Witkovsky, A. Gabelova, Significance of amino acid substitution variants of DNA repair genes in radiosusceptibility of cervical cancer patients; a pilot study, Neoplasma 55 (2008) 330–337.

[43] J.J. Hu, T.R. Smith, M.S. Miller, H.W. Mohrenweiser, A. Golden, L.D. Case, Amino acid substitution variants of Ape1 and XRCC1 genes associated with ionizing radiation sensitivity, Carcinogenesis 22 (2001) 917–922.

[44] M. Kasahara, K. Osawa, K. Yoshida, A. Miyaishi, Y. Osawa, N. Inoue, A. Tsutou, Y. Tabuchi, K. Tanaka, M. Yamamoto, E. Shimada, J. Takahashi, Association of MUTYH Gln324His and APEX1 Asp148Glu with colorectal cancer and smoking in a Japanese population, J. Exp. Clin. Cancer Res. 27 (2008) 49.

[45] M. Manuguerra, G. Matullo, F. Veglia, H. Autrup, A.M. Dunning, S. Garte, E. Gormally, C. Malaveille, S. Guarrera, S. Polidoro, F. Saletta, M. Peluso, L. Airoldi, K. Overvad, O. Raaschou-Nielsen, F. Clavel-Chapelon, J. Linseisen, H. Boeing, D. Trichopoulos, A. Kalandidi, D. Palli, V. Krogh, R. Tumino, S. Panico, H.B. Bueno-De-Mesquita, P.H. Peeters, E. Lund, G. Pera, C. Martinez, P. Amiano, A. Barricarte, M.J. Tormo, J.R. Quiros, G. Berglund, L. Janzon, B. Jarvholm, N.E. Day, N.E. Allen, R. Saracci, R. Kaaks, P. Ferrari, E. Riboli, P. Vineis, Multifactor dimensionality reduction applied to a large prospective investigation on gene–gene and gene–environment interactions, Carcinogenesis 28 (2007) 414–422.

[46] A.F. Olshan, G.M. Shaw, R.C. Millikan, C. Laurent, R.H. Finnell, Polymorphisms in DNA repair genes as risk factors for spina bifida and orofacial clefts, Am. J. Med. Genet. 135 (2005) 268–273.

Please cite this article in press as: E.T. Yu, M.Z. Hadi, Bioinformatic processing to identify single nucleotide polymorphism that potentially affect Ape1 function, Mutat. Res.: Genet. Toxicol. Environ. Mutagen. (2010), doi:10.1016/j.mrgentox.2010.06.015

[47] J.P. Erzberger, D.M. Wilson 3rd, The role of Mg^{2+} and specific amino acid residues in the catalytic reaction of the major human abasic endonuclease: new insights from EDTA-resistant incision of acyclic abasic site analogs and site-directed mutagenesis, J. Mol. Biol. 290 (1999) 447–457.

[48] D.G. Rothwell, I.D. Hickson, Asparagine 212 is essential for abasic site recognition by the human DNA repair endonuclease HAP1, Nucleic Acids Res. 24 (1996) 4217–4221.

[49] C. Choudhary, C. Kumar, F. Gnad, M.L. Nielsen, M. Rehman, T.C. Walther, J.V. Olsen, M. Mann, Lysine acetylation targets protein complexes and co-regulates major cellular functions, Science 325 (2009) 834–840.

[50] A. Abyzov, A. Uzun, P.R. Strauss, V.A. Ilyin, An AP endonuclease 1–DNA polymerase beta complex: theoretical prediction of interacting surfaces, PLoS Comput. Biol. 4 (2008) e1000066.

[51] A.E. Vidal, S. Boiteux, I.D. Hickson, J.P. Radicella, XRCC1 coordinates the initial and late stages of DNA abasic site repair through protein–protein interactions, EMBO J. 20 (2001) 6530–6539.

[52] E. Yu, S.P. Gaucher, M.Z. Hadi, Probing conformational changes in Ape1 during the progression of base excision repair, Biochemistry (2010).