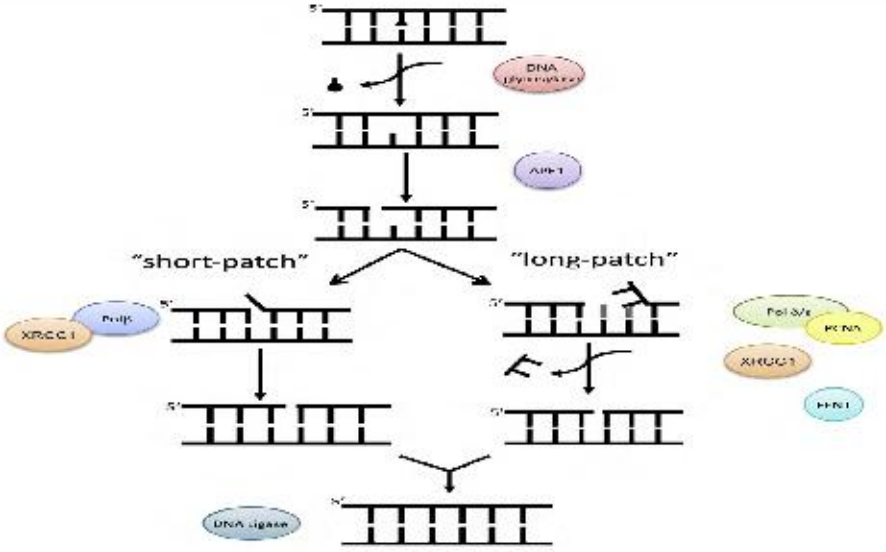
Abstract in Article

Title	Bioinformatic processing to identify SNP that potentially affect Ape1 function
Introduction	 The bae excision repair(BER) pathway is involved in the correction of DNA modifications These modifications may arise spontaneosly or from replication error or through chemical modification by oxidation or alkylation. first step : Removal of an inappropriate base by a DNA glycosylase second step : The abasic site is subsequently recognized by an apurinic/apymidinic(AP) endonuclease(Ape1) third step : Ape1 incises the phosphdiester backbone of DNA immediately 5' to lesion, leaving strand break with a normal 3'-hydroxyl group and a non-conentional 5'-abasic residue At this stage of the repair : The "Short-patch BER pathway proceeds with DNA polymerase(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 and requires and requires FEN1 to excise the flap-like structure produced by DNA polymerase strand displacement most frequently executed by DNApolymerase δ or ε) Finally, DNA ligase I or a complex of XRCC1 and Ligase III seal the nick and completes BER restoring DNA to its normal state.

Materials & Methods	1. Identification of variation within the Ape1 gene 1) To screen for amino acid substitutions in Ape1 ⇩ the most recent EST database using the tBlastn algorithm and SNP homepage at a NCBI(http://ncbi.nlm.nih.gov/SNP/index.html) 2) All scored variant DNA sequencing traces ⇩ downloading EST trace date from http://genome.wustl.edu/genomes and importing the traces in to the Genetic Annotation Initiative web server (http://www.chlc.org/gai/) 2. Predicting impact of amino acid substitutions in the Ape1 variants The possible impact of the amino acid substitutions in Ape1 variants ⇩ Using PolyPhen and SIFT software ⇩ Solvent accessible surface areas ofApe1residue ⇩ calculated by GETAREA using crystal structures (PDB ID: 1E9N, 1DE8, 1DE9) ⇩ Molecular models for the Ape1 variants ⇩ using Pymol ⇩ the potential structural changes of the different Ape1 variants

Results & Discussion

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)	I54T(0.00, 2.29)	N212H(0.00, 2.52)
K35Q(0.18, 1.27)	D124N(0.06, 0.04)	Q238K(0.26, 0.29)	Y45F (0.68, 1.56)	S66G(0.00, 2.36)	N212K(0.00, 2.83)
A38R(0.51, 1.27)	D148E(1.00, 0.46)	G239A(0.14, 0.54)	Q51H (0.05, 1.54)	L92Y(0.00, 1.73)	E217R (0.00, 1.42)
Q51S(0.36, 1.59)	D163N(0.03, 0.48)	F240L(0.06, 0.77)	K52E (0.86, 1.72)	P139Q(0.00, 2.58)	I218N (0.00, 2.84)
K58Q(0.58, 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, 0.13)	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)	Q182R (0.27, 1.71)	Y171N(0.00, 3.46)	N226E (0.00, 2.45)
L81H(0.13, 0.21)	L207A(0.22, 0.80)	N259K(0.01, 0.02)	S201F (0.02, 1.75)	G178V(0.02, 2.04)	N226G(0.00, 2.51)
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)	C241R(0.01, 0.53)	T232L (0.00, 1.93)	K197E(0.11, 0.07)	L256P (0.00, 2.15)
E87R(0.08, 1.18)	K224Q(0.00, 1.03)	L287W(0.00, 1.41)	H255A (0.31, 1.56)	V206C(0.00, 2.13)	Y264T (0.00, 3.42)
G115S(0.43, 0.76)	K227Q(0.07, 0.18)	K308Q(0.75, 0.45)	Y257L (0.00, 1.91)	V206G(0.00, 1.78)	C310W(0.00, 3.66)
A121S(0.43,0.11)	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)	

1) Eighty Ape1 variants are classified into three categories based on the PolyPhen prediction
2) SIFT and PolyPhen scores for each mutant(sift, PolyPhen)
3) Red for solvent/surface accessible, blue for intermediate, and black for buried sites

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

1) identified 80 unique amino acid substitution variants from the EST database
2) 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.
3) 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

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

1) The SIFT(sortingintolerantfromtolerant) and PolyPhen are used to predict such effects.

Results & Discussion	2) TheSIFT program uses multiple sequence alignment information to predict tolerant andd eleterious substitutions. SIFT scores which are less than 0.05 are deemed to be intolerant variations in the sequence and scores that are greatert han 0.2 are assigned to be tolerable substitutions. 3) PolyPhen makes predictions based on several sources of data including multiple sequence alignment - 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. - PSIC scores between the native and mutant, and big change (>2.0) in the scores indicates that the substitution is rarely observed in the protein. 3. Majority of the Apw1 variants that were predicted to be deleterious map to the core of the protein 1) The majoity of the mutations that were predicted to be damaging were internally located residues. (decreased hydrophobicity, introduction of charge effects ,volume chages)--> destabilize protein, prevent proper folding in the first place(Y171N, D210E,N212H,N212K) 2) a few of these "damaging" mutations(G178V,K197E,E217R, and N222Q) are located on the surface of Ape1 protein. G178V and N222Q are in the DNA binding loop and contact the DNA strand that contains the abasic site

4. Ape1 variants of residues lacalized at the surface are potentially damaging

1) There are a number of surfacd-localized Ape1 variants that were predicted to be potentially or probably damaging (Q51H,K52H,C138F,S201F,N222K,G178V,E217R,N222Q)

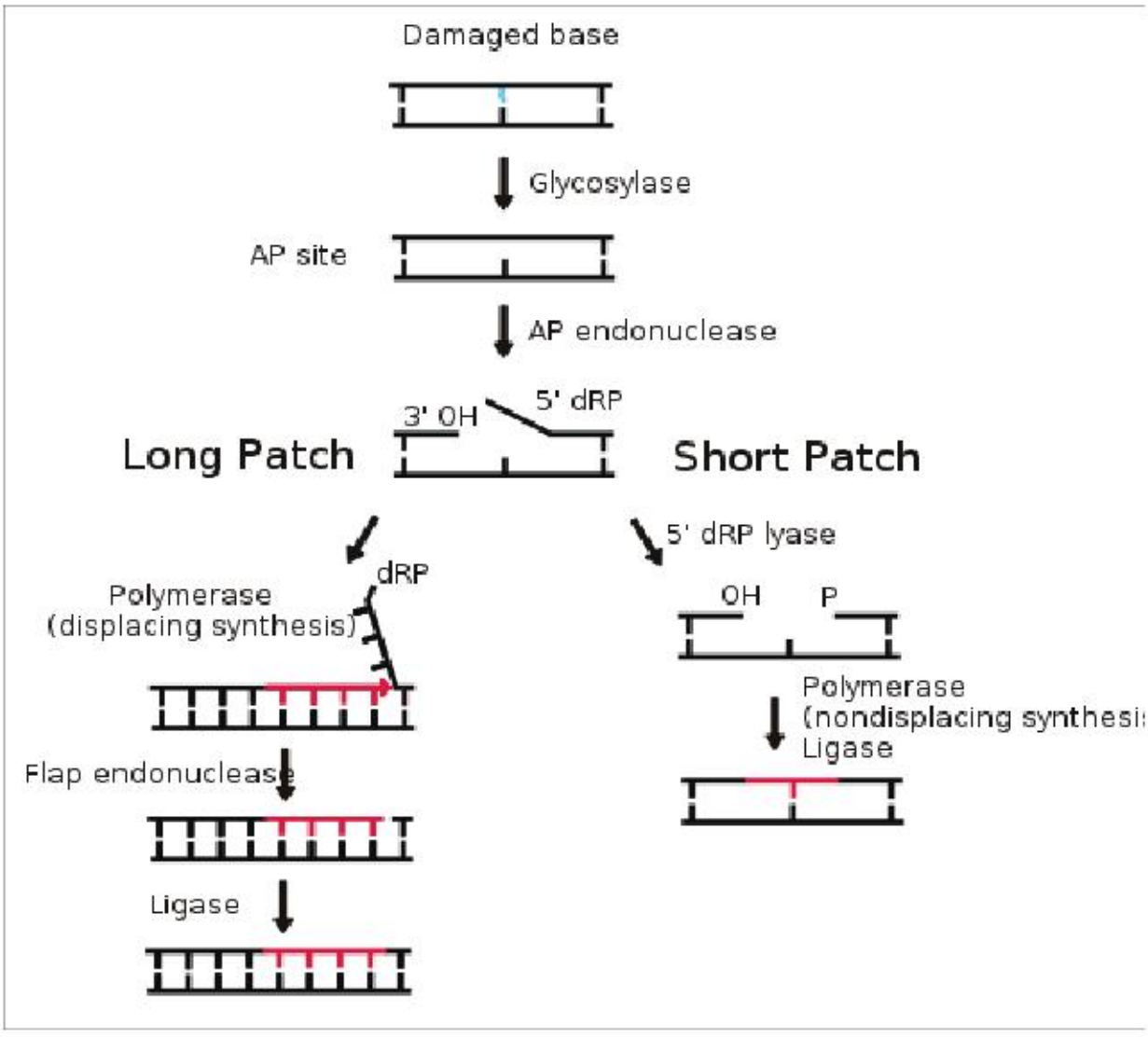
or SIFT-intolerant (D124N,D163N,R202P,K224R,K224Q,K227Q,N259K, and N259H)

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 variants of structure of the protein

3) Some of these mutated residues comstitute passible protein-protein interaction sites involved in siganling pathway (EX. Apw1 through uts N-termianl domain(residues 1- 35) has been showm to physiclly interact with the BER accessory protein XRCC1)

Defects in a variety of DNA repair pathways lead to cancer predisposition, and BER appears to follow this pattern. Deletion of BER genes increases the mutation rate in a variety of organisms, predicting that loss of BER could contribute to the development of cancer. Indeed, somatic mutations in Pol β have been found in 30% of human cancers, and some of these mutations lead to transformation when expressed in mouse cells. Mutations in the DNA glycosylase MYH are also known to increase susceptibility to colon cancer.

KEY WORDS BER ,APE1 cSNPS ,Disease susceptibility


Damaged base
Glycosylase
AP site
AP endonuclease
3' OH
5' dRP
Long Patch
Short Patch
5' dRP lyase
dRP
Polymerase
(displacing synthesis)
OH
P
Polymerase
(nondisplacing synthesi
Ligase
Flap endonuclease
Ligase