Uncovering antigenic property of protein through mutation finding in DNA repair sequences

Project Report submitted in partial fulfillment of the requirement for the degree of

Bachelor of Technology

In

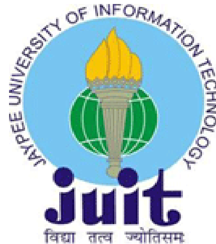
Bioinformatics

under the Supervision of

Dr. Tiratha Raj Singh

Submitted By

Khushboo Jindal (121521)



Jaypee University of Information Technology,

Waknaghat, Solan – 173234,

Himachal Pradesh

TABLE OF CONTENTS

Chapter No.	Topics	Page
	Certificate from the Supervisor	II
	Acknowledgement	III
	Abstract	IV
	List of Figures	V
Chapter-1	Introduction	1
Chapter-2	Review of literature	
Chapter-3	Methodology	
Chapter-4	Results and Discussion	
Chapter-5	Conclusion	
	References	

CERTIFICATE

This is to certify that the work titled —" Uncovering antigenic property of protein through mutation finding in DNA repair sequences", submitted by Khushboo Jindal (121521) in partial fulfillment for the award of degree of Bachelor of Technology in Bioinformatics to Jaypee University of Information Technology, Waknaghat, Himachal Pradesh has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.

Date:

Dr. Tiratha Raj Singh
Assistant Professor (Senior Grade)
Department of Biotechnology and Bioinformatics
Jaypee University of Information Technology

ACKNOWLEDGEMENT

The project in this report is an outcome of continual work over and intellectual support from various sources. It is therefore almost impossible to express adequately the debts own to many persons who have been instrumental in imparting this work a successful status. It is however a matter of pleasure to express my gratitude and appreciation to those who have been contributing to bring about this project. We are extremely glad to make use of this opportunity to express a deep sense of gratitude to Dr. Tiratha Raj Singh, Dr. D.K. Bastola, Ishwor Thapa, Ankush Bansal, Ankita Shukla, and Ashwani Kumar, for lending us stimulating suggestions, innovative quality, guidance and creative thinking. I am grateful to all them for their constant support and constructive attitude. Dr. Tiratha Raj Singh provided constructive criticism, constant encouragement and valuable advice at all stages of project.

Words are inadequate to express our deep sense of gratitude to our classmates and friends for their constant support and keeping our spirit up throughout the project and their companionship during our stay in JUIT.

Last but not the least, I am grateful to my family for their unflinching support and guidance during the course of our study. I am lucky and proud of them.

………………..

Khushboo Jindal

Date:

ABSTRACT

The body's immune system develops antibodies that recognize and bind to "antigenic sites", which are regions found on a virus surface proteins. The distinct surface features of an antigen is its antigenic determinant. Antigenic molecules, normally "large" biological polymers, usually present surface features that can act as points of interaction for specific antibodies. Antigenic Sites show the statistically significant. For example in regards to influenza viruses, an antigenic site is the area on the protein spike protruding from their lipid bilayer that carries the receptor binding sites. This is supposed to trigger an immune response when your body detects these sites by binding antibodies to them. Each antibody is specifically produced by the immune system to match an antigen after cells in the immune system come into contact with it; this allows a precise identification of the antigen and the initiation of a tailored response. The triplet nature of the genetic code means that base changes within coding sequence can have several different outcomes. In this we did mutations so as to change in one DNA base pair that results in the substitution of one amino acid for another in the protein made by a gene. All the mutations and their impacts on ADNA repair systems have been checked and verified for their specific roles in various mechanisms involved.

LIST OF FIGURES

1. **Fig.**1. Base Excision Repair
2. **Fig.**2. Dna Mismatch Repair Pathways
3. **Fig**.3. Nucleotide Excision Repair
4. **Fig**.4. Double strand break repair
5. **Fig**.5. Reactome Enrichment analysis
6. **Fig**.6. Genes involved in cell cycle
7. **Fig.**7. Functional annotation of genes
   Fig.7(a), Fig.7(b), Fig.7(c), Fig.7(d)
8. **Fig**.8. Blast results

CHAPTER 1

INTRODUCTION

A great revolution and landmark accomplishments were attained in the field of biomedical research in 2003 during the era of The Human Genome Project (HGP): when the entire human genome was sequenced and the genes were mapped under high-end collaborative efforts of many countries [1]. For the comprehensive understanding of human genome, this collective endeavor dealt with major goals such as providing an accurate sequence of 3.2 billion DNA base pairs and identification of the estimated 20,000-25,000 human genes. The project also intended for establishing new tools and innovative technologies for analyzing the data and making it accessible worldwide. Moreover, it invoked Ethical, Legal, and Social Implications (ELSI) program for dealing with bioethics concerning the individual and society issues [2]. Ultimately, the project has proved a promising platform offering information on structure, function and organization of complete set of human genes. Till date, The HGP is considered as a commendable joint initiative by the scientific community, pooling its skills and resources to achieve a common goal which served as a perfect example stating a quote by a Japanese poet, Ryunosuke Satoro, *"Individually, we are one drop. Together, we are an ocean"*. After this crucial epoch, the complete human genome sequence was ultimately accessible after 50 years of attaining the structure of DNA (by James Watson and Francis Crick in 1953) [3].

The human genome is susceptible to numerous endogenous and exogenous damages resulting in genomic instability. DNA damage occurs at the rate of 10,000 to 1 million molecular lesions per cell each day. These damages are majorly responsible for altering the primary structure of the double helix by chemically modifying bases and often introduce unusual chemical bonds or bulky adducts that are not easily sustained in the standard double helix [4]. Once, the cell determines these diverse forms of damages, a coordination of cellular responses comprising of transcriptional activation, cell cycle control, apoptosis, senescence and DNA repair processes are triggered [5]. These DNA lesions not only impact the genome but if not repaired or aberrantly repaired enhance the likelihood of developing diseases such as multiple form of cancers, neurological abnormalities, immunodeficiency and premature aging. There are basically two kinds of classification for the damages; one depends on the type of factors or agents causing the DNA lesions and the other on the consequences of damages; discussed in the following sections.

There are numerous intrinsic as well external agents that contribute to the impairment in DNA where intrinsic factors include aberrations in metabolic and other cellular processes and external

factors comprises of the environmental agents such as ionizing radiations and genotoxic compounds [6].These damages are often introduced in DNA due to replication errors and other aberrant cellular metabolic processes like DNA replication, recombination and repair . It also includes damages resulting from reacting oxygen and nitrogen species (ROS and RNS respectively) such as superoxide anions, hydrogen peroxide and hydroxyl radicals engendered as the byproducts from normal metabolic processes like lipid peroxidation and oxidative respiration. The DNA suffering from endogenous damages consequently result in bulky adduct formation, hydrolysis (deamination, depurination, and depyrimidination of bases), oxidation (generation of 8-oxo-7,8-dihydroguanine and DNA strand interruptions), mismatches (resulting due to errors in replication) and alkylation (frequently due to 7-methylguanine, 1-methyladenine, 6-O-methylguanine) of bases [7]. All these chemical modifications interfere with the normal cellular processes and destabilize the integrity of genome.

There is a huge diversity in environmental agents causing damage to DNA such as UV radiations, X-rays, Gamma rays, plant toxins, thermal disruption, viruses, certain aromatic and genotoxic compounds . All these external factors are alleged to alter the structure of DNA and produce aberrations in DNA such as formation of free radicals, cyclobutane pyrimidine dimers (CPDs) and pyrimidine–pyrimidone-(6-4)-photoproducts . Additionally, due to exogenous damage, the formed DNA adducts include oxidized bases, alkylated phosphotriesters and cross-linked DNA [8]. Depurination and single-strand breaks (SSBs) in DNA are also perceived at the elevated temperatures. Since, there are a variety of DNA damaging agents, also the outcome of these damages are extremely versatile leading to oxidation, alkylation and hydrolysis of bases including deamination, depurination and depyrimidination. Other prominent DNA damages include mismatches, bulky adduct formation, pyrimidine dimers and cross-linking of DNA. Further, the formation of these adducts trigger processes such as cell-cycle arrest or cell death, transcriptional program activation, apoptosis and DNA repair. Thus, to counteract the deleterious effect of these damages and to maintain the integrity of genome, different mechanisms exist for repairing DNA in a precise manner. A few lesions often escape repair process or remain unrepaired, leading to irreversible mutations in DNA that further alters the cellular phenotype and enhances the risk of oncogenesis and other associated diseases .

All the organisms ranging from prokaryotes to eukaryotes are equipped with DNA repair processes that deals with diverse form of lesions and prevent the genome from permanent mutations [9] . Each day approximately 1016–1018 DNA repair events take place in a healthy adult containing 1014 cells. The process of DNA repair involves intricate interactions transpiring in a highly systematic behavior where the major steps comprise of recognizing the damage and initiating signaling process, recruitment of repair proteins, processing of lesions, resynthesis of double strand DNA (dsDNA) and finally ligation is achieved. A single DNA repair mechanism can never handle the plethora of lesions , therefore there exists several DNA repair mechanisms that individually tackle the specialized damages . DNA repair pathways are broadly classified into 4 major classes based on their ability to recognize and remove different damages namely, direct reversal of damage (DRD), single-strand damage, double-strand damage and translesion synthesis (TLS) [10]. Each of these mechanisms incorporates a wide range of proteins, enzymes and follows different approaches for repairing the damaged DNA. Thus, DNA repair is an intricate process which confiscates all these diverse form of lesions via different repair mechanisms and maintains the genome integrity.

The DRD mechanism mainly focuses on lesions occurring in only one of the four bases and does not ruptures the phosphodiester backbone of DNA for repairing damages. The mechanism doesn't require a DNA template to chemically reverse the damage. A wide range of damages such as CPDs and alkylation of bases are directly removed via this mechanism. The CPDs resulting as a consequence of an abnormal covalent linkage between adjacent pyrimidine bases formed due to UV light irradiation are repaired directly by photoreactivation i.e. reversing the damage using photolyase enzyme. This photoreactivation reaction was the first DNA repair process to be discovered in the bacteriophage in 1949. Cells are also capable of reversing the methylation of guanine residue, i.e. O6-methylguanine (O6-MeG) caused due to either environmental alkylating agents or endogenously by S-adenosylmethionine which acts as a methyl donor in many cellular reactions [11]. The spontaneous addition of methyl group is one of the major factors responsible for point mutations in humans. Most of the lesions formed by alkylation of DNA are repaired by O-6-methylguanine-DNA methyltransferase encoded by methyl guanine methyl transferase (MGMT) gene. This enzyme repairs the damage by stoichiometrically transferring the alkyl group from O-6 position to the cysteine residue of

enzyme. The carried process is an expensive stoichiometric reaction since the enzyme is irreversibly inactivated.

CHAPTER 2

REVIEW OF LITERATURE

Our cells are constantly exposed to insults from endogenous and exogenous agents that can introduce damage into our DNA and generate genomic instability. Many of these lesions cause structural damage to DNA and can alter or eliminate fundamental cellular processes, such as DNA replication or transcription. DNA lesions commonly include base and sugar modifications, single and double-strand breaks, DNA-protein cross-links, and base-free sites [12]. To counteract the harmful effects of DNA damage, cells have developed a specialized DNA repair system, which can be subdivided into several distinct mechanisms based on the type of DNA lesion. These processes include base excision repair, mismatch repair, nucleotide excision repair, and double-strand break repair, which comprise both homologous recombination and non-homologous end-joining.

It is estimated that each of the ∼1013 cells within the human body incurs tens of thousands of DNA-damaging events per day. DNA exclusively serves as the repository for the genetic information in each living cell and its integrity and stability are of much greater consequence than other cellular components, such as RNA and proteins. DNA damage can interfere with essential cellular processes, such as transcription or replication, and can compromise the viability of the cell. Specific DNA lesions can also induce mutations that cause cancer or other diseases as well as contribute to the aging process [13]. Thus, cells have evolved a network of DNA repair mechanisms to remove different types of DNA damage. Regardless of the type of lesion and the mechanism required for its repair, cells initiate a highly coordinated cascade of events—collectively known as the DNA damage response (DDR)—that senses the DNA damage, signals its presence, and mediates its repair. DNA, like any other molecule, is subject to chemical reactions. DNA damage may result from either intrinsic or extrinsic agents. In general,the vast majority ofDNAmodifications are endogenous in origin.The simplest form of endogenous DNA damage is spontaneous hydrolysis.

DNA Repair Mechanism

To compensate for the many types of DNA damage that occur, cells have developed multiple repair mechanisms wherein each corrects a different subset of lesions. At a minimum, most would agree that mammalian cells utilize five major DNA repair mechanisms: base excision repair (BER), mismatch repair (MMR), nucleotide excision repair (NER), and double-strand break repair, which includes both homologous recombination (HR) and non-homologous end joining (NHEJ) [14]. APE1 hydrolyzes the phosphodiester backbone immediately 5'to the AP site, creating a single-strand break flanked by 3' OH and 5' deoxyribose phosphate (5' dRP) termini. Alternatively, some DNA glycosylases have an associated AP lyase activity and are also capable of cleaving AP sites via a β-elimination reaction to produce 3' phospho-α, β-unsaturated aldehyde and 5' phosphate at the margins of the break.The well-characterized DNA repair mechanisms include DRD, base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR), homologous recombination repair (HRR), non-homologous end joining

(NHEJ) and TLS, these incorporates varied set of genes, enzymes and pathways for repairing the DNA [15].

Direct Reversal of Damage (DRD)

The DRD mechanism mainly focuses on lesions occurring in only one of the four bases and does not ruptures the phosphodiester backbone of DNA for repairing damages. The mechanism doesn’t require a DNA template to chemically reverse the damage. A wide range of damages such as CPDs [16] and alkylation of bases are directly removed via this mechanism. The CPDs resulting as a consequence of an abnormal covalent linkage between adjacent pyrimidine bases formed due to UV light irradiation are repaired directly by photoreactivation i.e. reversing the damage using photolyase enzyme [17]. This photoreactivation reaction was the first DNA repair process to be discovered in the bacteriophage in 1949. Cells are also capable of reversing the methylation of guanine residue, i.e. O6-methylguanine (O6-MeG) caused due to either environmental alkylating agents or endogenously by S-adenosylmethionine which acts as a methyl donor in many cellular reactions. The spontaneous addition of methyl group is one of the major factors responsible for point mutations in humans. Most of the lesions formed by alkylation of DNA are repaired by O-6-methylguanine-DNA methyltransferase encoded by methyl guanine methyl transferase (MGMT) gene. This enzyme repairs the damage by stoichiometrically transferring the alkyl group from O-6 position to the cysteine residue of enzyme. The carried process is an expensive stoichiometric reaction since the enzyme is irreversibly inactivated. DRD mechanism [18]. Due to UV exposure, the dimers formed in DNA by covalent binding of two consecutive thymine bases instead of normal base pairing are removed by photoreactivation process in which the DNA photolyases recognize the bends in DNA. On excitation with blue light (>300 nm), photolyases change the conformation and break the dimer apart; making the DNA free from damage. The other damage reversal is the removal of methyl group from guanine residue via MGMT [19].

Base Excission Repair

BER, as the name implies, is the predominant mechanism responsible for the repair of damaged DNA bases that, in contrast to NER do not significantly distort the overall structure of the DNA. BER is described as a highly coordinated pathway of consecutive enzymatic reactions. However, several distinct BER sub-pathways occur, which are contingent on the type of damage encountered at the onset as well as throughout the BER process. BER is typically initiated by the series of lesion-specific DNA glycosylases [20] that remove the damaged base by cleaving the N-glycosidic bond linking the base to its corresponding deoxyribose, leading to the production of an AP or a basic site. At least twelve DNA glycosylases have been identified to date, each acting upon a single or small number of partially overlapping base lesions . Despite their structural

diversity, all DNA glycosylases utilize a base-flipping mechanism in which the target base is 'flipped' to an extra helical position for excision from DNA [21]. The resultant AP site is both an intermediate product of BER and a highly prevalent DNA lesion produced by spontaneous base loss. In either case, AP sites are generally repaired by apurinc/apyrmidinic endonuclease 1 (APE1), the second enzyme in the canonical BER pathway [21]. Overall, BER is a multistep process that requires the sequential activity of several proteins and consists of numerous entry points based on the type of damage encountered.

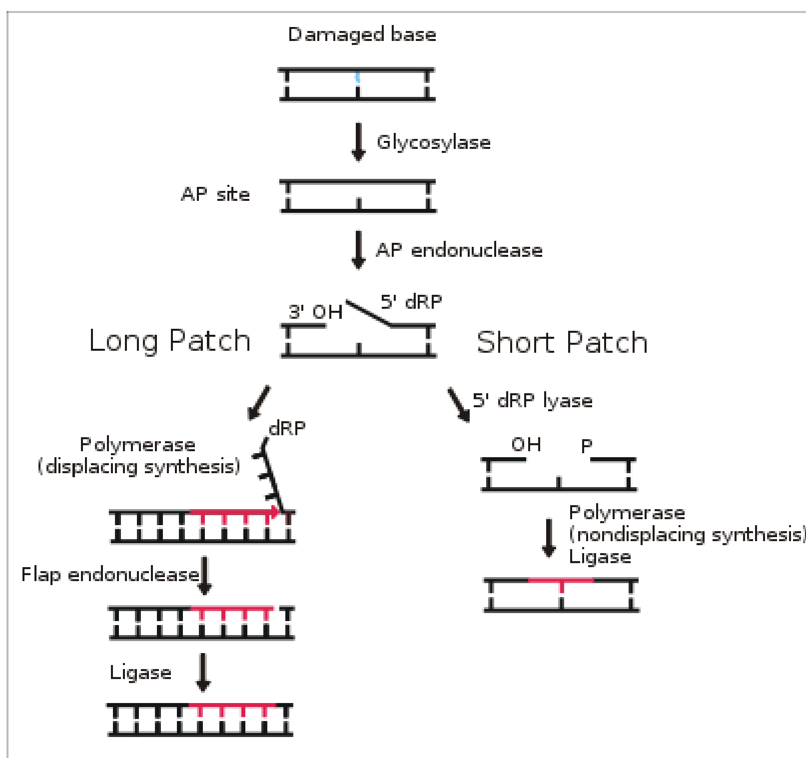


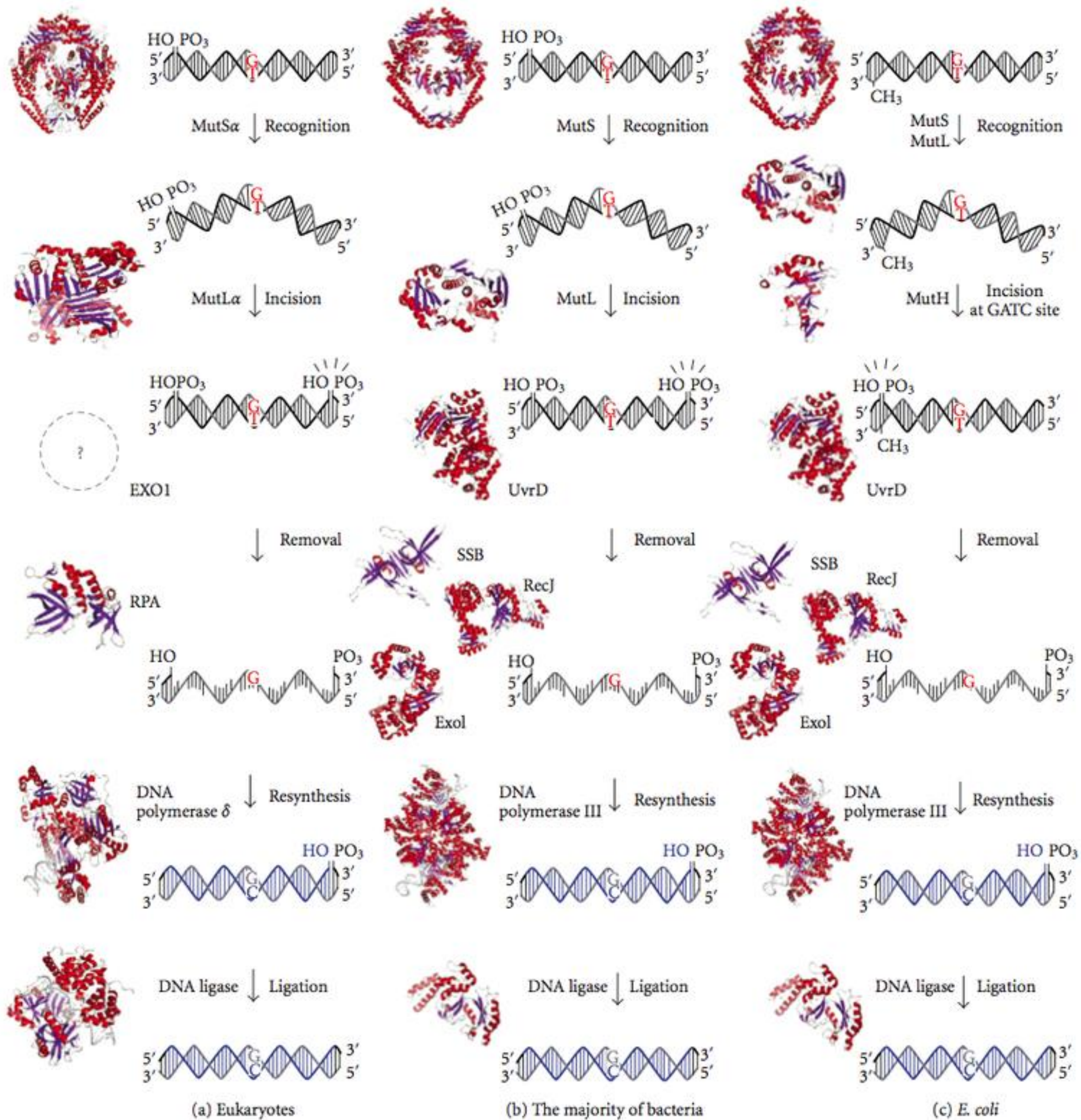
Figure1: Basic steps of base excision repair

Fig 1: Image adapted from :
https://en.wikipedia.org/wiki/Base_excision_repair#/media/File:BER_basic_pathway.svg

Mismatch Repair (MMR)

The MMR system plays an essential role in post-replication repair of misincorporated bases that have escaped the proofreading activity of replication polymerases. In addition to mismatched bases, MMR proteins also correct insertion/deletion loops (IDLs) that result from polymerase slippage during replication of repetitive DNA sequences [22]. The significance of this pathway is corroborated by the fact that MMR deficient cells are said to display a mutator phenotype, which

is characterized by invariably microsatellite instability and an elevated mutation frequency. More importantly, germline mutations inMMRgenes are predisposed to a variety of cancers,including hereditary non-polyposis colon cancer, also known as Lynch syndrome. The MMR pathway can be divided into three principle steps: a recognition step where mispaired bases are recognized, an excision step where the error-containing strand is degraded resulting in a gap, and a repair synthesis step, where the gap is filled by the DNA resynthesis. The MMR process is highly conserved from E.coli to humans. The canonical human MMR pathway is carried out by two major protein complexes, which are so called MutS and MutL, based on their homology to the E.coli MMR proteins.While MutS is responsible for mismatch recognition, MutL couples the recognition of the mispaired bases by the MutS complexes to downstreamMMRevents, which lead to the removal of the strand containing the error. In mammalians, the initial mismatch recognition step is fulfilled by two MutS activities that function as heterodimers. The MSH2-MSH6 heterodimer, also known as MutSα, preferentially recognizes base-base mismatches and small IDLs of one or two nucleotides, while MutSβ, the heterodimer of MSH2 and MSH3 recognizes larger IDLs. Formation of the MutSDNA complex is followed byATP-dependent recruitment of MutL homolog (MLH)complexes [23]. Three MutL activities have been identified and, like MutS, also function as heterodimeric complexes. MutLα, a heterodimer of MLH1 and PMS2, which contains the primary MutL activity (~90 %) in humans and supports the repair initiated by both MutSα and MutSβ. The two additional MutL heterodimers consist ofMLH1/PMS2 (MutLβ) and MLH1/MLH3 (MutLγ), which may play minor roles in MMR.



Figures2 : Diagram of DNA mismatch repair pathways. The first column depicts mismatch repair in eukaryotes, while the second depicts repair in most bacteria. The third column shows mistmatch repair, to be specific in E. coli.

Fig2: Image adapted from :
https://en.wikipedia.org/wiki/DNA_mismatch_repair#/media/File:DNA_mismatch_repair.png

Nucleotide Excision Repair (NER)

NER is a highly versatile repair pathway that can recognize and remove a wide variety of bulky, helix-distorting lesions from DNA. The most significant of these lesions are pyrimidine dimers,

such as cyclobutane pyrimidine dimers (CPD) and 6–4 photoproducts, which are produced by the UV component of sunlight. Another noteworthy substrate of NER is cisplatin-DNA intrastrand crosslinks. NER is mediated by the sequential assembly of repair proteins at the site of the DNA lesion [24]. While mechanistically similar to BER, the NER pathway is more complex, requiring some thirty different proteins to carry out a multi-step ‘cut-and-patch’-like mechanism. These steps involve DNA damage recognition, local opening of the DNA helix around the lesion, excision of a short single-strand segment of DNA spanning the lesion, and sequential repair synthesis and strand ligation. The biological importance of NER is supported by the fact that defects in NER cause several human genetic disorders, including xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy, which are all characterized by extreme sun sensitivity. In addition, these diseases demonstrate overlapping symptoms associated with cancer, developmental delay, immunological defects, neurodegeneration, and premature aging.

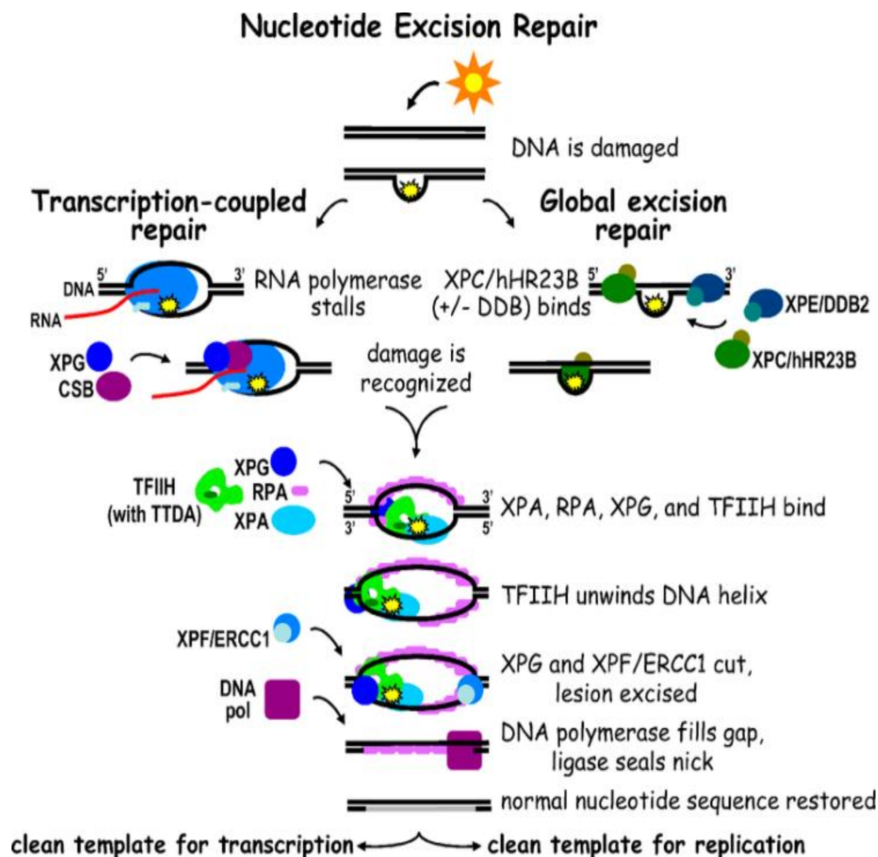


Figure3: Diagram of both the TC-NER and GG-NER pathways. The two pathways differ only in initial DNA damage recognition.

Fig3: Image adapted from :
https://en.wikipedia.org/wiki/Nucleotide_excision_repair#/media/File:Nucleotide_Excision_Repair-journal.pbio.0040203.g001.png

Double-Strand Break Repair

Double-strand breaks (DSBs) are amongst the most biologically hazardous types of DNA damage. For instance, a single unrepaired DSB is often sufficient to cause cell death. In addition, inaccurate repair can lead to deletions or chromosomal aberrations, events that associated with the development of cancer or other genomic instability syndromes. Thus, the repair of DSBs is both critical for cell survival and maintenance of genome integrity [25]. The two main mechanisms by which mammalian cells repair DSBs are homologous recombination (HR) and non-homologous end-joining (NHEJ). These two repair systems differ in their requirement for a homologous template DNA and in the fidelity of DSB repair. HR-directed repair is largely an error-free mechanism as it utilizes the genetic information contained in the undamaged sister chromatid as a template. In contrast, NHEJ is normally error-prone and involves elimination of DSBs by direct ligation of the broken ends [26]. NHEJ is reasoned to be the predominant pathway in mammalian cells operating in all phases of the cell cycle, while HR is restricted to the late-S and G2 phases.

Figure4 : DNA damage resulting in multiple broken chromosomes

Fig4: Image adapted from
https://en.wikipedia.org/wiki/DNA_repair#/media/File:Brokechromo.jpg

Homologous Recombination (HR)

HR can be conceptually divided into three phases: presynapsis, synapsis, and postsynapsis. During presynapsis, the DNA ends surrounding the DSB are processed through 5' to 3' end resection to generate molecules with 3' single-stranded tails. The heterotrimeric MRN complex (Mre11-Rad50-Nbs1) together with CtIP (RBBP8) [27] are responsible for the initiation of resection in which the 5' ends on either side of the DSB are trimmed back to create short 3' overhangs of single-strand DNA. The second step in the 5' to 3' resection is presumably continued by the combined action of BLM helicase (Bloom syndrome, RecQ helicase-like) and Exo1 exonuclease. Following end resection, single-stranded DNA tails are bound by RPA to remove disruptive secondary structures thatwould otherwise obstruct binding of Rad51 recombinase. RPA is subsequently replaced by Rad51 in conjunction with several mediator proteins, such as Rad52, BRCA2, and a group of proteins known as Rad51 paralogs (RAD51B, RAD51C, RAD51D, XRCC2, and XRCC3). The Rad51-coated single-stranded DNA tail, also referred to as the Rad51 nucleoprotein filament, then executes the DNA sequence homology search, which is the central reaction of HR. Once the homologous DNA has been identified, Rad51 mediates DNA strand invasion reaction, wherein the damaged DNA strand invades the template DNA duplex (i.e., sister chromatid) [28]. Next, DNA synthesis from the 3' end of the invading strand is carried out by DNA polymerase η followed by successive ligation by DNA ligase I to yield a four-way junction intermediate structure known as a Holliday junction.

Non- Homologous End-Joining (NHEJ)

The molecular mechanism of NHEJ is mediated by a relatively small number of essential factors that are sequentially recruited to DSB sites. The initial step in the NHEJ process entails recognition and binding of the Ku70/Ku80 heterodimer (Ku) to the exposed DNA termini of the DSB. Structurally, Upon binding to DNA, the Ku-DNA complex recruits the catalytic subunit of DNA-dependent protein kinase (DNA-PKcs) to generate the so-called DNA-PK holoenzyme, which exhibits protein kinase activity [29]. The recruitment of DNA-PKcs induces an inward translocation of Ku along the DNA, allowing DNA-PKcs to contact DNA termini. More importantly, the binding of the DNA-PKcs molecules on opposing DSB ends promotes synapsis or tethering of the two DNA molecules. Synapsis of DNA-PKcs also results in autophosphorylation of DNA-PKcs, which allows the DNA termini to become accessible. Like most DNA repair processes, depending on the type and complexity of the DSB break, DNA ends may require modification prior to ligation.

The biological significance of DNA repair mechanisms is underscored by the fact that their deregulation can contribute to the initiation and progression of cancer. On the other hand, DNA repair can confer resistance to front line cancer treatments (i.e. chemotherapy and radiation), which rely on the generation of DNA damage to kill cancer cells. Thus, the sensitivity of cancer

cells to DNA damaging agents is most likely related to intrinsic deficiencies in DNA repair mechanisms. The capacity of cancer cells (or cancer stem cells) [30] to recognize DNA damage and initiate DNA repair is a key mechanism for therapeutic resistance or recurrence.

Translesion synthesis (TLS)

Although cells can deal with almost all forms of lesions with the aid of highly complex DNA repair systems but circumstantially when the lesion still exists before initiation of replication, it can block the replication machinery and ultimately lead to cell death. In such situations, there is a need for dedicated mechanisms for tolerating the DNA damage without mediating repair of a lesion. These DNA damage tolerance processes still maintain the integrity and survive damages to genome but often lead to mutations [31]. In normal repair systems, the DNA polymerases cannot bypass these lesions and the replication halts (if not repaired) but in the damage tolerance process or TLS, there are specialized polymerases recruited on the damage called translesion polymerases.These translesion polymerases replace the stalled replicative polymerases at the 3'-OH end of a primed DNA template in a process called polymerase switching. After following synthesis over the stretch contacting lesion, the process is reversed and replicative polymerases again resume the synthesis. These translesion polymerases are usually error-prone and have implications in a variety of cancers. There are a few specialized polymerases that accurately bypass specific classes of DNA lesions whereas others bypass the same lesion with high error rates. Thus, recruitment of inappropriate specialized polymerase during TLS could also result in mutations or genome rearrangements.

The key mediators for the signaling are ATM and ATR kinases that facilitate repair process via their downstream targets and induce cell cycle arrest. These processes have gained a lot of attention for cancer therapy since resistance to genotoxic therapies has been associated to damage response signaling and targeting the process will prove vital [32]. Over recent years, the role of DNA repair pathways and abnormalities in them have provided insights in understanding the development of numerous diseases such as aging, cancer , neurological aberrations and also their therapeutic relevance. In view of the significance of DNA repair in human diseases, we have compiled a list of diseases occurring due to mutations or other abnormalities in DNA repair system in the upcoming section. The objectives designed for my research work also focus on some of the diseases and the role of associated DNA repair mechanisms and pathways.

Diseases Specific to DNA Repair

Genetic instability leading to carcinogenesis is known to stimulate by DNA damages and errors created by the DNA machinery. These damages unless repaired lead to mutations in DNA repair genes, impacting the phenotypic consequences and hence give rise to numerous human genetic

diseases by escalating the predisposition to a variety of cancers. Defects in all the repair mechanisms have been reported to associate with several disorders therefore appropriate understanding of these intricate mechanisms is essential to comprehend the underlying human genetic diseases. In the following section, we provide a brief description for a few DNA repair associated disorders along with the respective biomarkers.

Xeroderma Pigmentosum (XP)

XP is a rare autosomal genetic disorder in which extreme sensitivity to UV radiations is observed. It results due to aberrations or mutations in NER associated genes (Ddb2, Ercc2, Ercc3, Ercc4, Ercc5, Xpa and Xpc) and Polh, which corresponds to TLS mechanism. The normal ability of cell to repair UV induced damages (thymine dimers) is affected in XP due to aberrations in these genes . The disorder is mainly manifested by symptoms such as freckles, dark patches, corneal ulcerations and often neurological abnormalities. XP increases the susceptibility to develop skin cancer and other form of cancers i.e. cancer on eyes, lips and ears. XP is predominantly found in the Japanese population and prevalent in both males and females with equal probabilities. The two DNA repair mechanisms thus contributing towards XP are NER, specifically GG-NER and TLS repair mechanisms.

Cockayne syndrome (CS) CS is a genetic autosomal recessive disorder characterized by growth retardation (short stature), neurological abnormalities, premature aging and sensitivity to sunlight. There are four major variations of CS i.e. CS Type I, II, III and XP-CS where CS type I to III are classified on the basis of severity and age of onset of disease whereas in XP-CS, person suffers from both XP and CS. Mutations in two DNA repair genes, i.e. ERCC6, also known as Cockayne syndrome complementation group type B (CSB) and ERCC8, also referred as Cockayne syndrome complementation group type A (CSA) are responsible for CS. Both the genes are implicated in TC-NER, the mechanism activated in highly transcribed region. Mutations in ERCC6 gene alone contributes to ~70% of CS cases. If either of the genes is mutated or altered, DNA is not repaired and the damage is accumulated hence stimulating cell death.

Werner Syndrome (WS)

WS is exemplified by premature aging and is also known as adult progeria. The clinical manifestations of the disorder includes juvenile bilateral cataracts, mask-like face, bird-like nose, diabetes mellitus, atherosclerosis and osteoporosis. It has been reported that mutations in WRN gene forms the basis for WS since the mutation leads to production of an abnormally short nonfunctional werner protein. The growth failure observed in WS is due to the altered werner protein which enforces the cells to either divide slowly or stop dividing as compared to normal.

This mutation also causes damage to accumulate in the genome thus impairing the normal cellular activities. The normal WRN gene encodes a protein whose central domain resembles members of the RecQ helicases that play an important role in repairing DSBs i.e. mainly implicated in HRR and NHEJ mechanisms. Thus, mutations in the gene may impact genetic stability and lead to development of WS.

Fanconi Anemia (FA)

FA is another DNA repair associated autosomal recessive disorder characterized by congenital defects such as short stature, growth failure and abnormalities in skin, ears, kidneys and eyes. Leukemia and bone marrow failure are common traits of the disease i.e. the person's capability to produce blood cells is distorted. FA is an outcome of mutations in 15 DNA repair genes i.e. BRCA2, BRIP1, FANCA, FANCB, FANCC, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCL, FANCM, PALB2, RAD51C and SLX4 involved in FA pathway [33]. When the DNA replication halts due to DNA damage, the FA pathway is activated which recruits certain proteins at the site of damage which repair damages and replication can then continue. Thus, these genes play vital roles in repairing DNA and maintaining integrity of genome. Mutations in these genes have adverse impact on the damage accumulation and finally contribute to cell death and malignancies.

Systematic Perspective for DNA Repair Mechanisms

A high fidelity of transmission of genetic information is absolutely necessary in ensuring normal life development. Genomic stability is essential for the maintenance of life. Eukaryotic cells exposed to DNA damaging agents also activate important defensive pathways by inducing multiple proteins involved in DNA repair. Specifically, the p53 (or TP53) protein which is involved in many anti-cancer and apoptosis mechanisms is only present in animals while it is absent in plants and fungi. This protein plays an important role in human as it is an activator of components of the NER pathway. In addition, extensive studies over the last two decades have generated a wealth of information on the DNA repair systems and pathways of cell. We now have an improved understanding of DNA damaging factors and the myriad of mechanisms by which cells protect their genomic integrity. DNA damage response can be described as a collaborative effect of DNA repair systems and cell-cycle regulation [34]. Proteins involved in DNA damage checkpoints are found to perform similar functions in regular cell-cycle control processes. Some proteins such as p53 and Rad51 have regulatory functional effects over multiple pathways, establishing the linkages between different pathways in the network.

Evolution in the field of human genomics and continuous innovations and developments in the areas such as genetics, high-throughput techniques, comparative genomics, systems biology and bioinformatics has added a new dimension to the biomedical research. There are evidences of decreased lifespan and increased cancer incidence in experimental animals with genetic

deficiency in DNA repair. Inherited mutations that affect DNA repair genes are strongly associated with high cancer risk in humans. HNPCC is strongly associated with specific mutations in genes involved in MMR pathway. BRCA1 and BRCA2, two well known mutations conferring a hugely increased risk of breast cancer on carrier and both are associated with a large number of DNA repair pathways. Recently, nine mutations (3 nonsense, 5 missense and 1 affecting mRNA splicing) in RECQL gene implicated in double-strand repair are reported in early breast cancer progression which now serves as a marker for its screening. Other severe human genetic disorders associated with DNA damage and repair pathways (especially NER) include XP and CS.

Eventually, the exhaustive mechanistic perspective on these DNA repair mechanisms will help in enhanced realization of the disposition to many human hereditary diseases such as cancer and aging. The in silico study on DNA repair process and its implicated mechanisms will not only uncover vital clues regarding oncogenesis [35] and age-related diseases but will also provide innovative paradigms for genetic susceptibility, prevention, diagnosis and the rational therapy.

CHAPTER 3

METHODOLOGY

The following steps are performed to do the work for the chromosome number 9.

Step 1: 1: Download coding sequences

Using UCSC Table Browser, download the RefSeq mRNA sequences into a file called chr9.fna

Step 1b: Translate to amino acid sequences

transeq chr1.fna chr1.faa

Step 2: Finding antigenic sites

antigenic -sequence chr1.faa -outfile chr1.ag.txt

From this output (chr9.antigenic), we extracted the liens with Sequence and HitCount and present it as two columns.

Sequence: hg38_refGene_NM_001005484_1 from: 1 to: 306

HitCount: 9

cat chr1.ag.txt | grep -A1 "# Sequence:" | cut -f3 -d' '|

grep -v "\-\-" | sed 'N;s/\n/ /' > chr1.ANTIGENIC1

Step 3: Mutate gene to allow more of Leucine, Cysteine and Valine

perl /var/www/spring2016/project/mutate.pl chr9.fna

/var/www/spring2016/project/possibleMut.lonly.txt >

chr9.lonly.fna

perl /var/www/spring2016/project/mutate.pl chr9.fna

/var/www/spring2016/project/possibleMut.conly.txt >

chr9.conly.fna

perl /var/www/spring2016/project/mutate.pl chr9.fna

/var/www/spring2016/project/possibleMut.vonly.txt >

chr9.vonly.fna

transeq chr9.lonly.fna chr9.lonly.faa

transeq chr9.conly.fna chr9.conly.faa

transeq chr9.vonly.fna chr9.vonly.faa

Step 4: Finding antigenic sites in mutated sequences

antigenic -sequence chr1.lonly.faa -outfile

chr9.lonly.ag.txt

antigenic -sequence chr1.conly.faa -outfile

chr9.conly.ag.txt

antigenic -sequence chr1.vonly.faa -outfile

chr9.vonly.ag.txt

cat chr1.lonly.ag.txt | grep -A1 "# Sequence:" | cut -f3 -

d' '| grep -v "\-\-" | sed 'N;s/\n/ /' > chr1.l.ANTIGENIC2

cat chr1.conly.ag.txt | grep -A1 "# Sequence:" | cut -f3 -

d' '| grep -v "\-\-" | sed 'N;s/\n/ /' > chr1.c.ANTIGENIC2

cat chr1.vonly.ag.txt | grep -A1 "# Sequence:" | cut -f3 -

d' '| grep -v "\-\-" | sed 'N;s/\n/ /' > chr1.v.ANTIGENIC2

Step 5: Statistical analysis

In this step, we obtain the list of genes, which have at-least five more

antigenic sites in the mutated protein.

The following commands to find the genes with increased

antigenic sites.

```
orig<-read.table(file="chr1.ANTIGENIC1",sep=" ",header=F)
mutC<-read.table(file="chr1.c.ANTIGENIC2",sep=" ",header=F)
mutL<-read.table(file="chr1.l.ANTIGENIC2",sep=" ",header=F)
mutV<-read.table(file="chr1.v.ANTIGENIC2",sep=" ",header=F)
mutLHigh<-which(mutL[,2]-orig[,2]>=5)
mutCHigh<-which(mutC[,2]-orig[,2]>=5)
mutVHigh<-which(mutV[,2]-orig[,2]>=5)
unionHigh<-union(mutLHigh,union(mutCHigh,mutVHigh))
genes<-unique(sort(orig[unionHigh,1]))
write.table(file="genesWithHigherAnitgenicSites.txt",genes,
quote=F,row.names=F,col.names=F)
```

Get only the mRNA accession IDs to convert it to gene symbols.

```
cat genesWithHigherAnitgenicSites.txt | cut -f3,4 -d'_' >
genesWithHigherAnitgenicSites.acc.txt
```

(Check the content of output file.)

Next step is to convert, mRNA accession ids to gene symbols. Open the following UniProt URL:

http://www.uniprot.org/uploadlists/

Here, in “1. Provide your identifiers”, paste your list of mRNA accession numbers obtained above (genesWithHigherAnitgenicSites.acc.txt)

In “2. Select options”, select RefSeq Nucleotide as “From” Option and UniProtKB as “To” option and click on “GO”.

In the result page, click on and select Gene names (primary) as shown below. Then click on save.

In the UniProt interface, you will now see Gene Names (primary) column. In the

URL, add &format=tab so that the output is now in tabular format. Save this

output and use cut command to get the gene name column into a file called,

chr1.higherAntigenic.genes.txt

Step 6: Functional annotation

Use this file, chr9.higherAntigenic.genes.txt , as input to R functional annotation

using GO and Reactome. this file should replace up or down regulated gene list file. The remaining process to get significant GO terms and reactome pathways remains the same.

Functional Enrichment:

We have performed the functional enrichment of list of up-regulated and down-regulated genes obtained. We are using Gene Ontology, Reactome/KEGG pathway enrichment.

<u>Three Major steps for enrichment</u>

• Get an input list of differentially expressed genes

• Convert the gene names to Entrez Gene ID

- Use a function (ID2EG) in R

• Functional enrichment

o Call enrich function

- GOFunction
- EnrichPathway
- Pathview

A. <u>GO Enrichment of genes of interest</u>

For GO enrichment in R, the Bio-conductor provides a package called GOFunction.

Step 1: Gather Up/Down regulated genes

Up and Down lists are saved in /var/www/spring2016/week14/

Step 2: Gene Name to Entrez ID conversion

```
> up<-read.table(file="/var/www/spring2016/week14/up.list.txt")

> down<-read.table(file="/var/www/spring2016/week14/down.list.txt")
```

Step 2a: Gather entrez ids of the gene

```
> library("pathview")

> upsymbol2eg<-id2eg(as.character(up[,1]),category="symbol",org="Hs")

> entrez_up<-upsymbol2eg[,2]

> downsymbol2eg<-id2eg(as.character(down[,1]),category="symbol",org="Hs")

> entrez_down<-downsymbol2eg[,2]
```

Step 2b: Gather entrez ids for the background genes

```
> library("org.Hs.eg.db")

> px <- org.Hs.egACCNUM

> rGenes<-as.numeric(mappedkeys(px))
```

Step3: Perform GO Enrichment

Enrichment

• Iteratively test the enrichment of each GO term one by one for the given list of genes

• Hypergeometric tests to find if the subpopulation of GO terms is over or under represented in a given list of genes.

```
library("GOFunction")

> sigBPUpTerm<-GOFunction(as.numeric(unique(entrez_up)),rGenes,

        organism="org.Hs.eg.db",ontology="BP",fdrmethod="BY",filename="upBP

SigTerm")
```

The ontology="BP" gives Biological process ontology enrichment.

Alternatively, we can change "BP" to "MF" and "CC" for molecular function and

cellular component enrichment.

B. Reactome Enrichment in R

```
> library(“pathview”)

> library(“ReactomePA”)

> x <- enrichPathway(gene = entrez_up, pvalueCutoff = 0.05, readable=T)

> head(summary(x))
```

C. Pathway Visualization

The next step is to overlay the up-regulated genes in known Cell-Cycle Pathway

```
> library(“pathview”)

> pv.out <- pathview(gene.data = entrez_up, pathway.id = "04110",

species = "hsa", out.suffix = "cell_cycle")
```

CHAPTER 4

RESULTS AND DISCUSSION

Here we performed varied analysis process to find role of genes involved in DNA repair process, especially involved in antigenic property of a protein. To check antigenic property we retrieve 35 raw genes from UCSC browser in R and find 8 genes which are actively involved in DNA repair process and annotate those genes in broad term of gene ontology represented as biological process, molecular function and cellular components whose analysis are shown in following figures. To check similarity and evolutionary study of those actively participated gene of DNA repair. Not much similarity found among all those genes and parallaely multiple sequence alignment give some hint of those genes in their evolutionary conservation role which was further validated when check the result in conseq server.

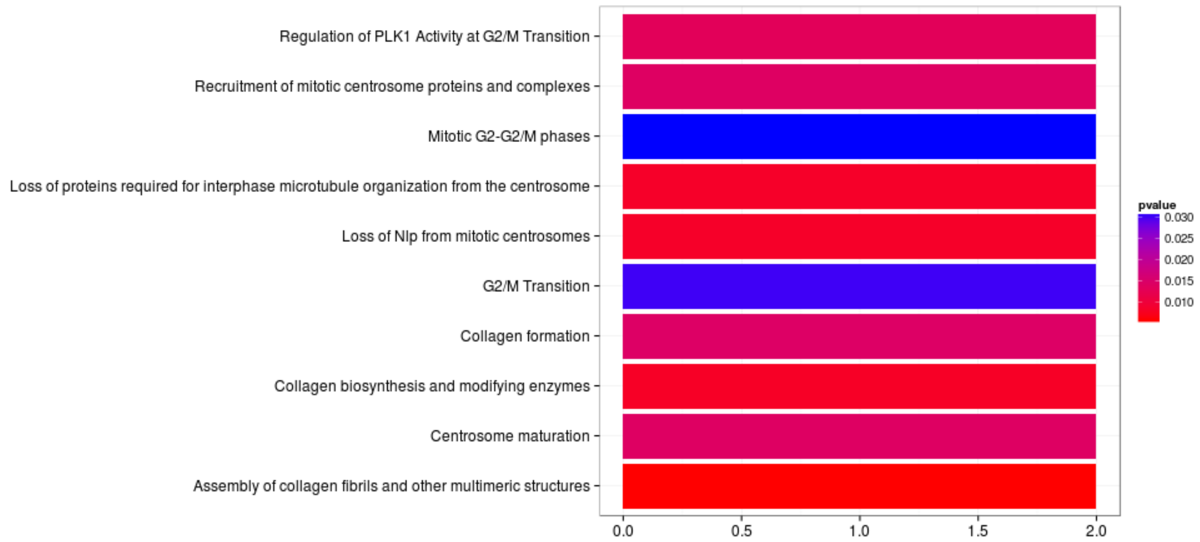


Figure5: Performing the Reactome Enrichment analysis gives the barplot of genes involved in cell cycle

Gene names	Gene ontology (biological process)
COL5A1	axon guidance [GO:0007411]; blood vessel development [GO:0001568]; cell adhesion [GO:0007155]; cell migration [GO:0016477]; collagen biosyntheti
ODF2	G2/M transition of mitotic cell cycle [GO:0000086]; mitotic cell cycle [GO:0000278]; multicellular organismal development [GO:0007275]; organell
SURF6 SURF-6	ribosome biogenesis [GO:0042254]
HNRNPK HNRPK	acute-phase response [GO:0006953]; aging [GO:0007568]; camera-type eye development [GO:0043010]; cellular protein metabolic process [GO:0044267]
SH3GL2 CNSA2 SH3D2A	antigen processing and presentation of exogenous peptide antigen via MHC class II [GO:0019886]; axon guidance [GO:0007411]; central nervous syst
GOLGA2	asymmetric cell division [GO:0008356]; cellular protein metabolic process [GO:0044267]; centrosome organization [GO:0051297]; COPII vesicle coat
TJP2 X104 ZO2	apoptotic process [GO:0006915]; cellular component disassembly involved in execution phase of apoptosis [GO:0006921]; establishment of endotheli
ATP6V1G1 ATP6G ATP6G1	cell redox homeostasis [GO:0045454]; cellular iron ion homeostasis [GO:0006879]; insulin receptor signaling pathway [GO:0008286]; ion transmembr
PRUNE2 BMCC1 BNIPXL C	apoptotic process [GO:0006915]; polyphosphate catabolic process [GO:0006798]
NOL8 C9orf34 NOP132	DNA replication [GO:0006260]; positive regulation of cell growth [GO:0030307]; rRNA processing [GO:0006364]
CDK5RAP2 CEP215 KIAA1	brain development [GO:0007420]; centriole replication [GO:0007099]; centrosome organization [GO:0051297]; chromosome segregation [GO:0007059]; e
RNF20 BRE1A	histone H2B ubiquitination [GO:0033523]; histone monoubiquitination [GO:0010390]; negative regulation of cell migration [GO:0030336]; negative r
IFT74 CCDC2 CMG1	cilium assembly [GO:0042384]; determination of left/right symmetry [GO:0007368]; heart development [GO:0007507]; intraciliary transport involved
COL27A1 KIAA1870	extracellular matrix organization [GO:0030198]; growth plate cartilage chondrocyte development [GO:0003431]
AKAP2 KIAA0920 PRKA2	

Figure6: Genes involved in cell cycle.

Reactome pathway-based analysis. These proteins and genes that are shown in the barplot mutated are higher in Cell Cycle. These genes are SH3GL2, CNSA2, SH3D2A involved in immune response.These

genes play a significant role in the immune response. SH3GL2 is involved in Signal Transduction Pathway. Signal transduction occurs when an extracellular signaling molecule activates a specific receptor located on the cell surface or inside the cell.CNSA2 is involved in Disease. We get total 35 genes after mutation that shows more antigenic sites. Interpretation were drawn from above differential gene expression was that most of the gene involved in cell cycle regulation process and cell organelle metabolism which are represented by P-value in range of 0.010 to 0.030 shown in figure1. Role of genes involved in repair process shown in figure 2.

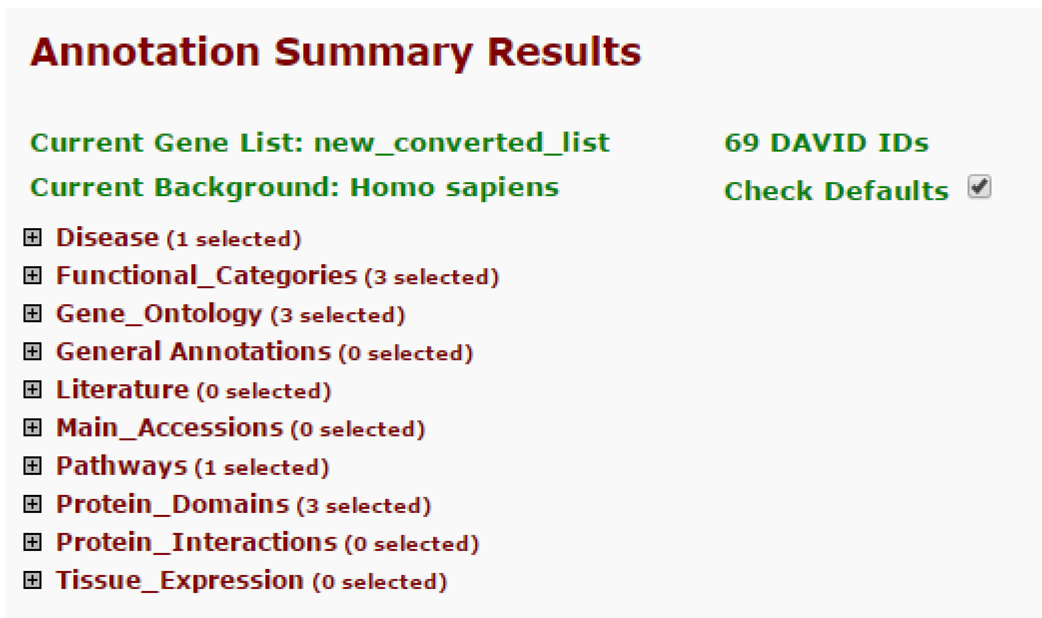


Figure7: Functional annotation of genes and their involvement are summarized majorly in term of disease, GO, Pathways and domain studies comprises 1, 3, 1, and 3 respectively; shown in following tables.

Disease (1 selected)

☐ GENETIC_ASSOCIATION_DB_DISEASE	1.4%	1	Chart	
☐ GENETIC_ASSOCIATION_DB_DISEASE_CLASS	1.4%	1	Chart	
☑ OMIM_DISEASE	2.9%	2	Chart	

Figure7(a): Disease selected

Category	%	Count	
COG_ONTOLOGY	31.9%	22	Chart
PIR_SEQ_FEATURE	1.4%	1	Chart
SP_COMMENT_TYPE	34.8%	24	Chart
SP_PIR_KEYWORDS	46.4%	32	Chart
UP_SEQ_FEATURE	36.2%	25	Chart

Figure7(b): ShowingPercentage contribution

Category	%	Count	
GOTERM_BP_1	27.5%	19	Chart
GOTERM_BP_2	27.5%	19	Chart
GOTERM_BP_3	26.1%	18	Chart
GOTERM_BP_4	26.1%	18	Chart
GOTERM_BP_5	23.2%	16	Chart
GOTERM_BP_ALL	27.5%	19	Chart
GOTERM_BP_FAT	27.5%	19	Chart
GOTERM_CC_1	44.9%	31	Chart
GOTERM_CC_2	44.9%	31	Chart
GOTERM_CC_3	44.9%	31	Chart
GOTERM_CC_4	44.9%	31	Chart
GOTERM_CC_5	44.9%	31	Chart
GOTERM_CC_ALL	44.9%	31	Chart
GOTERM_CC_FAT	40.6%	28	Chart
GOTERM_MF_1	33.3%	23	Chart
GOTERM_MF_2	33.3%	23	Chart
GOTERM_MF_3	17.4%	12	Chart
GOTERM_MF_4	14.5%	10	Chart

Figure7(c) :

Category	%	Count	
EC_NUMBER	4.3%	3	Chart
KEGG_PATHWAY	8.7%	6	Chart
PANTHER_PATHWAY	7.2%	5	Chart
REACTOME_PATHWAY	4.3%	3	Chart

Figure7(d) : Showing results including KEGG Pathway

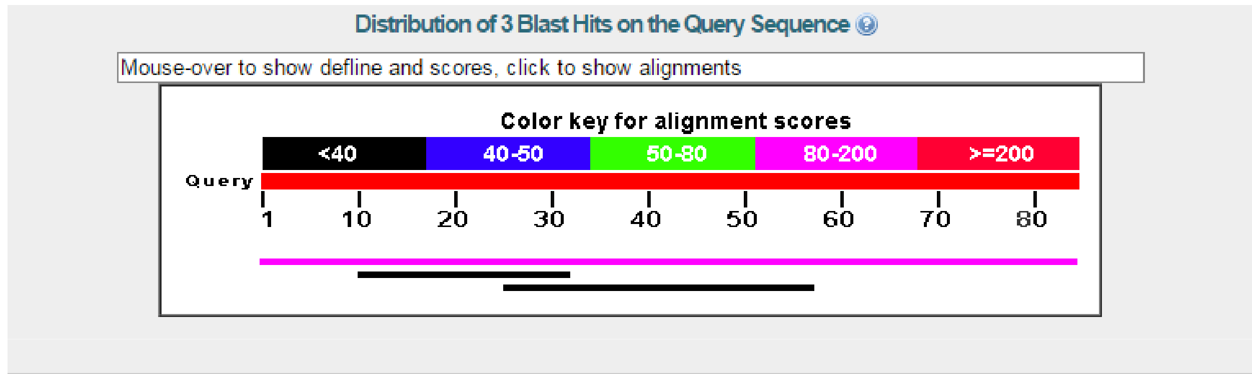


Figure8 : Showing results of Blast

Protein Blast hit of those primary 8 gene sequence involve in DNA repair process show about more than 60 % sequence similarly when compare with non redundant sequence database of ncbi and parallely perform multiple sequence alignment of those sequence to check consevness in those sequence with all default parameter setting. Result are shown below:

```
CLUSTALW Result

[clustalw.aln][clustalw.dnd][readme]
 Exec


CLUSTAL 2.1 Multiple Sequence Alignments


Sequence type explicitly set to Protein
Sequence format is Pearson
Sequence 1: gi|444738435|emb|CCQ43283.1|     84 aa
Sequence 2: gi|38014150|gb|AAH08760.3|      591 aa
Sequence 3: gi|959027357|gb|JAP01810.1|     206 aa
Sequence 4: gi|119608461|gb|EAW88055.1|     361 aa
Sequence 5: gi|78070721|gb|AAI07743.1|      372 aa
Sequence 6: gi|32140760|ref|NP_116277.2|   1860 aa
Start of Pairwise alignments
Aligning...



Sequences (1:2) Aligned. Score: 15.4762
Sequences (1:3) Aligned. Score: 14.2857
Sequences (1:4) Aligned. Score: 13.0952
Sequences (1:5) Aligned. Score: 14.2857
Sequences (1:6) Aligned. Score: 14.2857
Sequences (2:3) Aligned. Score: 13.1068
Sequences (2:4) Aligned. Score: 11.0803
Sequences (2:5) Aligned. Score: 11.2903
Sequences (2:6) Aligned. Score: 30.9645
```

```
Sequences (3:4) Aligned. Score: 15.534
Sequences (3:5) Aligned. Score: 13.1068
Sequences (3:6) Aligned. Score: 15.0485
Sequences (4:5) Aligned. Score: 13.2964
Sequences (4:6) Aligned. Score: 13.0194
Sequences (5:6) Aligned. Score: 12.9032
Guide tree file created:   [clustalw.dnd]

There are 5 groups
Start of Multiple Alignment

Aligning...
Group 1:                     Delayed
Group 2:                     Delayed
Group 3:                     Delayed
Group 4:                     Delayed
Group 5:                     Delayed
Alignment Score -127

CLUSTAL-Alignment file created  [clustalw.aln]

clustalw.aln

CLUSTAL 2.1 multiple sequence alignment


gi|38014150|gb|AAH08760.3|        ------------------------------------
--------------
gi|32140760|ref|NP_116277.2|
MGAGSARGARGTAAAAAARGGGFLFSWILVSFACHLASTQGAPEDVDILQ
gi|444738435|emb|CCQ43283.1|      ------------------------------------
--------------
gi|959027357|gb|JAP01810.1|       ------------------------------------
--------------
gi|119608461|gb|EAW88055.1|       ------------------------------------
--------------
gi|78070721|gb|AAI07743.1|        ------------------------------------
--------------


gi|38014150|gb|AAH08760.3|        ------------------------------------
--------------
gi|32140760|ref|NP_116277.2|
RLGLSWTKAGSPAPPGVIPFQSGFIFTQRARLQAPTGTVIPAALGTELAL
gi|444738435|emb|CCQ43283.1|      ------------------------------------
--------------
gi|959027357|gb|JAP01810.1|       ------------------------------------
--------------
gi|119608461|gb|EAW88055.1|       ------------------------------------
--------------
gi|78070721|gb|AAI07743.1|        ------------------------------------
--------------
```

```
gi|38014150|gb|AAH08760.3|        ------------------------------------
--------------
gi|32140760|ref|NP_116277.2|
VLSLCSHRVNHAFLFAVRSQKRKLQLGLQFLPGKTVVHLGSRRSVAFDLD
gi|444738435|emb|CCQ43283.1|      ------------------------------------
--------------
gi|959027357|gb|JAP01810.1|       ------------------------------------
--------------
gi|119608461|gb|EAW88055.1|       ------------------------------------
--------------
gi|78070721|gb|AAI07743.1|        ------------------------------------
--------------


gi|38014150|gb|AAH08760.3|        ------------------------------------
--------------
gi|32140760|ref|NP_116277.2|
MHDGRWHHLALELRGRTVTLVTACGQRRVPVLLPFHRDPALDPGGSFLFG
gi|444738435|emb|CCQ43283.1|      ------------------------------------
--------------
gi|959027357|gb|JAP01810.1|       ------------------------------------
--------------
gi|119608461|gb|EAW88055.1|       ------------------------------------
--------------
gi|78070721|gb|AAI07743.1|        ------------------------------------
--------------


gi|38014150|gb|AAH08760.3|        ------------------------------------
--------------
gi|32140760|ref|NP_116277.2|
KMNPHAVQFEGALCQFSIYPVTQVAHNYCTHLRKQCGQADTYQSPLGPLF
gi|444738435|emb|CCQ43283.1|      ------------------------------------
--------------
gi|959027357|gb|JAP01810.1|       ------------------------------------
--------------
gi|119608461|gb|EAW88055.1|       ------------------------------------
--------------
gi|78070721|gb|AAI07743.1|        ------------------------------------
--------------


gi|38014150|gb|AAH08760.3|        ------------------------------------
--------------
gi|32140760|ref|NP_116277.2|
SQDSGRPFTFQSDLALLGLENLTTATPALGSLPAGRGPRGTVAPATPTKP
gi|444738435|emb|CCQ43283.1|      ------------------------------------
--------------
gi|959027357|gb|JAP01810.1|       ------------------------------------
--------------
```

gi|119608461|gb|EAW88055.1| ------------------------------------
----
gi|78070721|gb|AAI07743.1| ------------------------------------
----

gi|38014150|gb|AAH08760.3| ------------------------------------
----
gi|32140760|ref|NP_116277.2|
QRTSPTNPHQHMAVGGPAQTPLLPAKLSASNALDPMLPASVGGSTRTPRP
gi|444738435|emb|CCQ43283.1| ------------------------------------
----
gi|959027357|gb|JAP01810.1| ------------------------------------
----
gi|119608461|gb|EAW88055.1| ------------------------------------
----
gi|78070721|gb|AAI07743.1| ------------------------------------
----

gi|38014150|gb|AAH08760.3| ------------------------------------
----
gi|32140760|ref|NP_116277.2|
AAAQPSQKITATKIPKSLPTKPSAPSTSIVPIKSPHPTQKTAPSSFTKSA
gi|444738435|emb|CCQ43283.1| ------------------------------------
----
gi|959027357|gb|JAP01810.1| ------------------------------------
----
gi|119608461|gb|EAW88055.1| ------------------------------------
----
gi|78070721|gb|AAI07743.1| ------------------------------------
----

gi|38014150|gb|AAH08760.3| ------------------------------------
----
gi|32140760|ref|NP_116277.2|
LPTQKQVPPTSRPVPARVSRPAEKPIQRNPGMPRPPPPSTRPLPPTTSSS
gi|444738435|emb|CCQ43283.1| ------------------------------------
----
gi|959027357|gb|JAP01810.1| ------------------------------------
----
gi|119608461|gb|EAW88055.1| ------------------------------------
----
gi|78070721|gb|AAI07743.1| ------------------------------------
----

gi|38014150|gb|AAH08760.3| ------------------------------------
----
gi|32140760|ref|NP_116277.2|
KKPIPTLARTEAKITSHASKPASARTSTHKPPPFTALSSSPAPTPGSTRS

gi|444738435|emb|CCQ43283.1| ------------------------------------
----
gi|959027357|gb|JAP01810.1| ------------------------------------
----
gi|119608461|gb|EAW88055.1| ------------------------------------
----
gi|78070721|gb|AAI07743.1| ------------------------------------
----

gi|38014150|gb|AAH08760.3| ------------------------------------
----
gi|32140760|ref|NP_116277.2|
TRPPATMVPPTSGTSTPRTAPAVPTPGSAPTGSKKPIGSEASKKAGPKSS
gi|444738435|emb|CCQ43283.1| ------------------------------------
----
gi|959027357|gb|JAP01810.1| ------------------------------------
----
gi|119608461|gb|EAW88055.1| ------------------------------------
----
gi|78070721|gb|AAI07743.1| ------------------------------------
----

gi|38014150|gb|AAH08760.3| ------------------------------------
----
gi|32140760|ref|NP_116277.2|
PRKPVPLRPGKAARDVPLSDLTTRPSPRQPQPSQQTTPALVLAPAQFLSS
gi|444738435|emb|CCQ43283.1| ------------------------------------
----
gi|959027357|gb|JAP01810.1| ------------------------------------
----
gi|119608461|gb|EAW88055.1| ------------------------------------
----
gi|78070721|gb|AAI07743.1| ------------------------------------
----

gi|38014150|gb|AAH08760.3| --------------------------------
MGPPGPPGPRGPSGAPGA
gi|32140760|ref|NP_116277.2|
SPRPTSSGYSIFHLAGSTPFPLLMGPPGPKGDCGLPGPPGLPGLPGIPGA
gi|444738435|emb|CCQ43283.1| ------------------------------------
----
gi|959027357|gb|JAP01810.1| ------------------------------------
----
gi|119608461|gb|EAW88055.1| ------------------------------------
----
gi|78070721|gb|AAI07743.1| ------------------------------------
----

gi|38014150|gb|AAH08760.3| DGPQGPPGGIGNPGAVGEKGEPGEAGEP--------
-------GLPGEGG
gi|32140760|ref|NP_116277.2|
RGPRGPPGPYGNPGLPGPPGAKGQKGDPGLSPGKAHDGAKGDMGLPGLSG
gi|444738435|emb|CCQ43283.1| ------------------------------------
----
gi|959027357|gb|JAP01810.1| ------------------------------------
----
gi|119608461|gb|EAW88055.1| ------------------------------------
----
gi|78070721|gb|AAI07743.1| ------------------------------------
----

gi|38014150|gb|AAH08760.3| PPGPKGERGEKGESGPS-------------------
----
gi|32140760|ref|NP_116277.2|
NPGPPGRKGHKGYPGPAGHPGEQGQPGPEGSPGAKGYPGRQGLPGPVGDP
gi|444738435|emb|CCQ43283.1| ------------------------------------
----
gi|959027357|gb|JAP01810.1| ------------------------------------
----
gi|119608461|gb|EAW88055.1| ------------------------------------
----
gi|78070721|gb|AAI07743.1| ------------------------------------
----

gi|38014150|gb|AAH08760.3| ------------------------------------
----
gi|32140760|ref|NP_116277.2|
GPKGSRGYIGLPGLFGLPGSDGERGLPGVPGKRGKMGMPGFPGVFGERGP
gi|444738435|emb|CCQ43283.1| ------------------------------------
----
gi|959027357|gb|JAP01810.1| ------------------------------------
----
gi|119608461|gb|EAW88055.1| ------------------------------------
----
gi|78070721|gb|AAI07743.1| ------------------------------------
----

gi|38014150|gb|AAH08760.3| ------------------------------------
----
gi|32140760|ref|NP_116277.2|
PGLDGNPGELGLPGPPGVPGLIGDLGVLGPIGYPGPKGMKGLMGSVGEPG
gi|444738435|emb|CCQ43283.1| ------------------------------------
----
gi|959027357|gb|JAP01810.1| ------------------------------------
----
gi|119608461|gb|EAW88055.1| ------------------------------------
----

gi|78070721|gb|AAI07743.1| ------------------------------------
----

gi|38014150|gb|AAH08760.3| ------------------------------------
----
gi|32140760|ref|NP_116277.2|
LKGDKGEQGVPGVSGDPGFQGDKGSQGLPGFPGARGKPGPLGKVGDKGSI
gi|444738435|emb|CCQ43283.1| ------------------------------------
----
gi|959027357|gb|JAP01810.1| ------------------------------------
----
gi|119608461|gb|EAW88055.1| ------------------------------------
----
gi|78070721|gb|AAI07743.1| ------------------------------------
----

gi|38014150|gb|AAH08760.3|
GAAGPPGPKGPPGDDGPKGSPGPVGFPGDPGPPGEPGPAGQDGPPGDKG-
gi|32140760|ref|NP_116277.2|
GFPGPPGPEGFPGDIGPPGDNGPEGMKGKPGARGLPGPRGQLGPEGDEGP
gi|444738435|emb|CCQ43283.1| ------------------------------------
----
gi|959027357|gb|JAP01810.1| ------------------------------------
----
gi|119608461|gb|EAW88055.1| -
MASLLAKDAYLQSLAKKICSHSAPEQQARTRAGKTQGSETAGP------
gi|78070721|gb|AAI07743.1|
MASNHKSSAARPVSRGGVGLTGRPPSGIRPLSGNIRVATAMPPG------

gi|38014150|gb|AAH08760.3| --------------------------
DDGEPGQTGSPGPTGE--------
gi|32140760|ref|NP_116277.2|
MGPPGAPGLEGQPGRKGFPGRPGLDGVKGEPGDPGRPGPVGEQGFMGFIG
gi|444738435|emb|CCQ43283.1| ------------------------------------
----
gi|959027357|gb|JAP01810.1| ------------------------------------
----
gi|119608461|gb|EAW88055.1| ------------------------------------
----
gi|78070721|gb|AAI07743.1| ------------------------------------
----

gi|38014150|gb|AAH08760.3| ------------------------------------
-------PGPSGPP
gi|32140760|ref|NP_116277.2|
LVGEPGIVGEKGDRGMMGPPGVPGPKGSMGHPGMPGGMGTPGEPGPQGPP
gi|444738435|emb|CCQ43283.1| ------------------------------------
----

gi|959027357|gb|JAP01810.1| ------------------------------------
----
gi|119608461|gb|EAW88055.1| ------------------------------------
------------PK
gi|78070721|gb|AAI07743.1| ------------------------------------
----------TARP

gi|38014150|gb|AAH08760.3|
GKRGPPGPAGPEGRQGEKGAKGEAGLEGPPGKTGPIGPQGAPGKPGPDGL
gi|32140760|ref|NP_116277.2|
GSRGPPGMRGAKGRRGPRGPDGPAGEQGSRGLKGPPGPQGRPGRPGQQGV
gi|444738435|emb|CCQ43283.1| ------------------------------------
----
gi|959027357|gb|JAP01810.1| ------------------------------------
----
gi|119608461|gb|EAW88055.1|
KKRKKTQKKFRKREEKAAEHKAKSLGEKSPAASGARRPEAAKEEAAWASS
gi|78070721|gb|AAI07743.1|
GSRGCPIGTGGVLSSQIKVAHRPVTQQGLTGMKTGTKGPQRQILDKSYYL

gi|38014150|gb|AAH08760.3|
RGIPGPVGEQGLPGSPGPDGPPGPMGPPGLPGLKGDSGPKG---------
gi|32140760|ref|NP_116277.2|
AGERGHLGSRGFPGIPGPSGPPGTKGLPGEPGPQGPQGPIGPPGEMGPKG
gi|444738435|emb|CCQ43283.1| ------------------------------------
----
gi|959027357|gb|JAP01810.1| --GQKEAEERHKKLIHGSLWPTWRLRPP--------
----
gi|119608461|gb|EAW88055.1|
SAGNPADGLATEPESVFALDVLRQRLHEKIQEARGQGSAKELS-------
gi|78070721|gb|AAI07743.1|
GLLRSKISELTTEVNKLQKGIEMYNQENSVYLSYEKRAET----------

gi|38014150|gb|AAH08760.3| ---------------------------------
EKGHPGLIGLIGPPGEQ
gi|32140760|ref|NP_116277.2|
PPGAVGEPGLPGEAGMKGDLGPLGTPGEQGLIGQRGEPGLEGDSGPMGPD
gi|444738435|emb|CCQ43283.1| ------------------------------------
----
gi|959027357|gb|JAP01810.1| ------------------------------------
---TTGKHIVFGSD
gi|119608461|gb|EAW88055.1| ------------------------------------
PAALEKRRRRKQER
gi|78070721|gb|AAI07743.1| -----------------------------------
LAVEIKELQGQLADY

gi|38014150|gb|AAH08760.3| GEKGDRGLPGPQGSSGPKGEQGITGPSGPIGPP---
----

```
gi|32140760|ref|NP_116277.2|
GLKGDRGDPGPDGEHGEKGQEGLMGEDGPPGPPGVTGVRGPEGKSGKQGE
gi|444738435|emb|CCQ43283.1|       ------------------------------------
--------------
gi|959027357|gb|JAP01810.1|        DEEEKQEVVSAVKESKKPLFQDSQSED---------
--------------
gi|119608461|gb|EAW88055.1|        DRKKRKRKELRAKEKARKAEEATEAQEVVEATP---
--------------
gi|78070721|gb|AAI07743.1|         NMLVDKLNTNTEMEEVMNDYNMLKAQNDRETQS---
--------------


gi|38014150|gb|AAH08760.3|                       ----------------------
GPPGLPGPPGPKGAKGSSGPTGPK----
gi|32140760|ref|NP_116277.2|
KGRTGAKGAKGYQGQLGEMGVPGDPGPPGTPGPKGSRGSLGPTGAPGRMG
gi|444738435|emb|CCQ43283.1|       ------------------------------------
--------------
gi|959027357|gb|JAP01810.1|        ------------------------------------
--------------
gi|119608461|gb|EAW88055.1|        ------------------------------------
--------------
gi|78070721|gb|AAI07743.1|         ------------------------------------
--------------


gi|38014150|gb|AAH08760.3|         ------------------------------------
--------------
gi|32140760|ref|NP_116277.2|
AQGEPGLAGYDGHKGIVGPLGPPGPKGEKGEQGEDGKAEGPPGPPGDRGP
gi|444738435|emb|CCQ43283.1|       ------------------------------------
--------------
gi|959027357|gb|JAP01810.1|        ------------------------------------
--------------
gi|119608461|gb|EAW88055.1|        ------------------------------------
--------------
gi|78070721|gb|AAI07743.1|         ------------------------------------
--------------


gi|38014150|gb|AAH08760.3|              -------------------------------
GEAGHPGPPGPPGPPGEVI
gi|32140760|ref|NP_116277.2|
VGDRGDRGEPGDPGYPGQEGVQGLRGKPGQQGQPGHPGPRGWPGPKGSKG
gi|444738435|emb|CCQ43283.1|       ------------------------------------
--------------
gi|959027357|gb|JAP01810.1|        ------------------------------------
--------------
gi|119608461|gb|EAW88055.1|            --------------------------------
EGACTEPREPPGLIFNKV
gi|78070721|gb|AAI07743.1|              -------------------------------
LDVIFTERQAKEKQIRSVE
```

```
gi|38014150|gb|AAH08760.3|                      QPLP-----------------
IQASRTRRNIDASQLLDDGNGENYVDYAD
gi|32140760|ref|NP_116277.2|
AEGPKGKQGKAGAPGRRGVQGLQGLPGPRGVVGRQGLEGIAGPDGLPGRD
gi|444738435|emb|CCQ43283.1|       ------------------------------------
--------------
gi|959027357|gb|JAP01810.1|        ------------------------------------
--------------
gi|119608461|gb|EAW88055.1|        EVSE--------------------------------
--------------
gi|78070721|gb|AAI07743.1|         EEIEQ-------------------------------
--------------


gi|38014150|gb|AAH08760.3|         G-----------------------------------
--------------
gi|32140760|ref|NP_116277.2|
GQAGQQGEQGDDGDPGPMGPAGKRGNPGVAGLPGAQGPPGFKGESGLPGQ
gi|444738435|emb|CCQ43283.1|       ------------------------------------
--------------
gi|959027357|gb|JAP01810.1|        ------------------------------------
--------------
gi|119608461|gb|EAW88055.1|        ------------------------------------
--------------
gi|78070721|gb|AAI07743.1|         ------------------------------------
--------------


gi|38014150|gb|AAH08760.3|         ------------------------------------
--------------
gi|32140760|ref|NP_116277.2|
LGPPGKRGTEGRTGLPGNQGEPGSKGQPGDSGEMGFPGMAGLFGPKGPPG
gi|444738435|emb|CCQ43283.1|       ------------------------------------
--------------
gi|959027357|gb|JAP01810.1|        ------------------------------------
--------------
gi|119608461|gb|EAW88055.1|        ------------------------------------
--------------
gi|78070721|gb|AAI07743.1|         ------------------------------------
--------------


gi|38014150|gb|AAH08760.3|         ------------------------------------
--------------
gi|32140760|ref|NP_116277.2|
DIGFKGIQGPRGPPGLMGKEGIVGPLGILGPSGLPGPKGDKGSRGDWGLQ
gi|444738435|emb|CCQ43283.1|       ------------------------------------
--------------
gi|959027357|gb|JAP01810.1|        ------------------------------------
--------------
```

gi|119608461|gb|EAW88055.1| ------------------------------------
----
gi|78070721|gb|AAI07743.1| ------------------------------------
----

gi|38014150|gb|AAH08760.3| ------------------------------------
----
gi|32140760|ref|NP_116277.2|
GPRGPPGPRGRPGPPGPPGGPIQLQQDDLGAAFQTWMDTSGALRPESYSY
gi|444738435|emb|CCQ43283.1| ------------------------------------
----
gi|959027357|gb|JAP01810.1| ------------------------------------
----
gi|119608461|gb|EAW88055.1| ------------------------------------
----
gi|78070721|gb|AAI07743.1| ------------------------------------
----

gi|38014150|gb|AAH08760.3| --------
MEEIFGSLNSLKLEIEQMKRPLGTQQNPARTCKDLQLCHPDF
gi|32140760|ref|NP_116277.2|
PDRLVLDQGGEIFKTLHYLSNLIQSIKTPLGTKENPARVCRDLMDCEQKM
gi|444738435|emb|CCQ43283.1| ------------------------------------
------MVLLQLEC
gi|959027357|gb|JAP01810.1| ----------
EATADGAAAANQNAPHKARPKPSAPRLFADSDDDEEGDEE
gi|119608461|gb|EAW88055.1| --------
DEPASKAQRRKEKRQRVKGNLTPLTGRNYRQLLERLQARQSR
gi|78070721|gb|AAI07743.1| -------
EKQATDDIIKNMSFENQVKYLEMKTTNEKLLQELDTLQQQLDS

.

gi|38014150|gb|AAH08760.3|
PDGEYWVDPNQGCSRDSFKVYCNFTAGGSTCVFPDKKSEGARITSWPKEN
gi|32140760|ref|NP_116277.2|
VDGTYWVDPNLGCSSDTIEVSCNFTHGGQTCLKPITASKVEFAISR----
gi|444738435|emb|CCQ43283.1|
TMETWWDSSSCVLMENPGDSWCKFRTEGSVCGTGLEES------------
gi|959027357|gb|JAP01810.1|
EHGSRFDIRPEFEGPAGQKLMELQSRFGTDERFRLDSR------------
gi|119608461|gb|EAW88055.1|
LDELRGQDEGKAQELEAKMKWTNLLYKAEGVKIRDDER------------
gi|78070721|gb|AAI07743.1|
QNMKKESLEAEIAHSQVKQEAVLLHEKLYELESHRDQMIAE---------

gi|38014150|gb|AAH08760.3|
PGSWFSEFKRGKLLSYVDAEGNPVGVVQMTFLRLLSASAHQNVTYHCYQS

```
gi|32140760|ref|NP_116277.2|          --------------------------
VQMNFLHLLSSEVTQHITIHCLNM
gi|444738435|emb|CCQ43283.1|        ----------------------------
VAFCSELS-RTLQSDPVTCFS-
gi|959027357|gb|JAP01810.1|       ------------------------------
FLEDEEDEEEEKREVQKSPM
gi|119608461|gb|EAW88055.1|            -------------------------
LLQEALKRKEKRRAQRQRRWEKRTA
gi|78070721|gb|AAI07743.1|           ---------------------------
DKSIGSPMEEREKLLKQIKDDNQ
                                                              :    .

gi|38014150|gb|AAH08760.3|                               VAWQDAATGS-
YDKALRFLGSNDEEMSYDNNPYIRALVDGCATKKG-YQK
gi|32140760|ref|NP_116277.2|
TVWQEGTGQTPAKQAVRFRAWNGQIFEAGGQFRPEVSMDGCKVQDGRWHQ
gi|444738435|emb|CCQ43283.1|      ------------------------------------
-----------GPK
gi|959027357|gb|JAP01810.1|       E-----------------------------------
-----------DEE
gi|119608461|gb|EAW88055.1|       G-----------------------------------
------VVEKMQQR
gi|78070721|gb|AAI07743.1|        E-----------------------------------
-----IASMERQLT


gi|38014150|gb|AAH08760.3|
TVLEIDTPKVEQVPIVDIMFNDFGEASQKFGFEVGPACFMG----
gi|32140760|ref|NP_116277.2|
TLFTFRTQDPQQLPIISVDNLPPASSGKQYRLEVGPACFL-----
gi|444738435|emb|CCQ43283.1|      SVFCLRTCNMYIKKQ---------------------
---------
gi|959027357|gb|JAP01810.1|
ALEEEKKKNLSILQSVLRTGQQASDKTAGKAKQFRDVSAL-----
gi|119608461|gb|EAW88055.1|
QDRRRQNLRRKKAARAERRLLRARKKGRILPQDLERAGLV-----
gi|78070721|gb|AAI07743.1|
DTKEKINQFIEEIRQLDMDLEEHQDPTNYGWKILEKNTGSFKKQV
                                        .
clustalw.dnd

(
(
gi|444738435|emb|CCQ43283.1|:0.41856,
(
gi|38014150|gb|AAH08760.3|:0.35056,
gi|32140760|ref|NP_116277.2|:0.33980)
:0.08746)
:0.00757,
(
gi|959027357|gb|JAP01810.1|:0.41728,
gi|119608461|gb|EAW88055.1|:0.42738)
```

```
:0.01073,
gi|78070721|gb|AAI07743.1|:0.43492);
 Exec
```

In nutshell, GO enrichment and analysis gave no significant terms. Barchart of gene function and their involvement in cell cycle regulation. Enrichment in term of disease, pathways, domain etc. lower P-value of Blast hit and maximum similarity with low penalty found in MSA show evolutionary conserved genes of DNA repair process. All combined studies definitely insight the role of vital gene involve in DNA repair process particularly the antigenic role of protein coded by those gene. It is proposed that antigenic sites have significant role in DNA repair genes and as there is involvement of various kind of mechanisms and their associated disease this kind of predictions will definitely help the biotechnologists, biochemists, and biomedical scientists to use this information for the design of new therapeutic targets.

REFERENCES

1. Leinonen,R., Akhtar,R., Birney,E. et al. (2010) Improvements to services at the European Nucleotide Archive. Nucleic Acids Res., 38,D39–D45.

2. Kaminuma,E., Mashima,J., Kodama,Y. et al. (2010) DDBJ launches anew archive database with analytical tools for next-generation sequence data. Nucleic Acids. Res., 38, D33–D38.

3. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J. et al. (2010) GenBank Nucleic Acids Res., 38, D46–D51.

4. SPIN: http://www.ebi.ac.uk/swissprot/Submissions/spin/ (17 February2011, date last acessed).

5. CiteXplore: http://www.ebi.ac.uk/citexplore/ (17 February 2011,date last acessed).

6. Dexheimer, Thomas S. "DNA repair pathways and mechanisms." DNA repair of cancer stem cells. Springer Netherlands, 2013. 19-32.

7. https://www.bioconductor.org/packages/3.3/bioc/vignettes/ReactomePA/inst/doc/ReactomePA.html#pathway-enrichment-analysis.

8. http://www.uniprot.org/uniprot/?query=yourlist:M201604279DAXYZURW8&sort=yourlist:M201604279DAXYZURW8.

9. https://en.wikipedia.org/wiki/Antigen.

10. Caldecott, Keith W. "XRCC1 and DNA strand break repair." *DNA repair* 2.9 (2003): 955-969.

11. Helleday, Thomas, et al. "DNA double-strand break repair: from mechanistic understanding to cancer treatment." DNA repair 6.7 (2007): 923-935.

12. Drabløs, Finn, et al. "Alkylation damage in DNA and RNA—repair mechanisms and medical significance." DNA repair 3.11 (2004): 1389-1407.

13. Singh, Narendra P., et al. "A simple technique for quantitation of low levels of DNA damage in individual cells." Experimental cell research 175.1 (1988): 184-191.

14. Farmer, Hannah, et al. "Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy." Nature 434.7035 (2005): 917-921.

15. Cleaver, J. E. "Defective repair replication of DNA in xeroderma pigmentosum." (1968): 652-656.

16. Friedberg, Errol C., et al., eds. DNA repair and mutagenesis. American Society for Microbiology Press, 2005.

17. Lindahl, Tomas, and Richard D. Wood. "Quality control by DNA repair." Science 286.5446 (1999): 1897-1905.

18. Prives, C. A. R. O. L., et al. "DNA-binding properties of simian virus 40 T-antigen mutants defective in viral DNA replication." Molecular and cellular biology 3.11 (1983): 1958-1966.

19. Schneider, J. E. N. S., and E. L. L. E. N. Fanning. "Mutations in the phosphorylation sites of simian virus 40 (SV40) T antigen alter its origin DNA-binding specificity for sites I or II and affect SV40 DNA replication activity." Journal of virology 62.5 (1988): 1598-1605.

20. Keller, Janis M., and James C. Alwine. "Activation of the SV40 late promoter: direct effects of T antigen in the absence of viral DNA replication." Cell 36.2 (1984): 381-389.

21. Tratschin, J. D., IRVING L. Miller, and BARRIE J. Carter. "Genetic analysis of adeno-associated virus: properties of deletion mutants constructed in vitro and evidence for an adeno-associated virus replication function." Journal of virology 51.3 (1984): 611-619.

22. Bang, Holger, et al. "Mutation and citrullination modifies vimentin to a novel autoantigen for rheumatoid arthritis." Arthritis & Rheumatism 56.8 (2007): 2503-2511.

23. Kuo, MY-P., M. Chao, and J. O. H. N. Taylor. "Initiation of replication of the human hepatitis delta virus genome from cloned DNA: role of delta antigen." Journal of virology 63.5 (1989): 1945-1950.

24. Clark, R., et al. "Biochemical activities of T-antigen proteins encoded by simian virus 40 A gene deletion mutants." Molecular and cellular biology 3.2 (1983): 220-228.

25. Radic, M. Z., and M. Weigert. "Genetic and structural evidence for antigen selection of anti-DNA antibodies." Annual review of immunology 12.1 (1994): 487-520.

26. Fink, Daniel, et al. "The role of DNA mismatch repair in platinum drug resistance." Cancer research 56.21 (1996): 4881-4886.

27. Modrich, Paul, and Robert Lahue. "Mismatch repair in replication fidelity, genetic recombination, and cancer biology." Annual review of biochemistry 65.1 (1996): 101-133.

28. Drummond, James T., et al. "Isolation of an hMSH1-p 160 heterodimer that restores DNA mismatch repair to tumor cells." Science 268.5219 (1995): 1909.

29. Rajagopalan, Harith, et al. "Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status." Nature 418.6901 (2002): 934-934.

30. Kramer, Wilfried, et al. "Cloning and nucleotide sequence of DNA mismatch repair gene PMS1 from Saccharomyces cerevisiae: homology of PMS1 to procaryotic MutL and HexB." Journal of Bacteriology 171.10 (1989): 5339-5346.

31. Kessler, Steven W. "Rapid isolation of antigens from cells with a staphylococcal protein A-antibody adsorbent: parameters of the interaction of antibody-antigen complexes with protein A." The Journal of Immunology 115.6 (1975): 1617-1624.

32. Roth, J. U. R. G. E. N., M. O. I. S. E. Bendayan, and L. E. L. I. O. Orci. "Ultrastructural localization of intracellular antigens by the use of protein A-gold complex." Journal of Histochemistry & Cytochemistry 26.12 (1978): 1074-1081.

33. Caliceti, Paolo, and Francesco M. Veronese. "Pharmacokinetic and biodistribution properties of poly (ethylene glycol)–protein conjugates." Advanced drug delivery reviews 55.10 (2003): 1261-1277.

34. Benjamin, David C., et al. "The antigenic structure of proteins: a reappraisal." Annual review of immunology 2.1 (1984): 67-101.

35. Avery, Oswald T., and Walther F. Goebel. "Chemo-immunological studies on conjugated carbohydrate-proteins: II. Immunological specificity of synthetic sugar-protein antigens." The Journal of experimental medicine 50.4 (1929): 533.