Studies on DNA Repair Mechanism and Non- Small Cell Lung Cancer

Dissertation submitted in partial fulfilment of the requirement for the degree of

MASTER OF SCIENCE

IN

MICROBIOLOGY

By

Namra Khan

217856

Under the guidance of

Dr.Tiratha Raj Singh

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

DEPARTMENT OF BIOTECHNOLOGY AND BIOINFORMATICS

WAKNAGHAT, SOLAN, 173234, H.P., INDIA

DECLARATION

I hereby declare that work reported in the M.Sc. thesis entitled "**Studies on DNA repair mechanism and Non-Small Cell Lung Cancer**" submitted at Jaypee University of Information Technology, Solan, India, is an authentic record of my work carried out under the supervision of Dr. Tiratha Raj Singh. I have not submitted this work elsewhere for any other degree or diploma. I am fully responsible for the contents of my M.Sc. thesis.

Namra Khan Date:

Enrollment Number-217856

Department of Biotechnology and Bioinformatics

Jaypee University of Information Technology Solan, India

SUPERVISOR'S CERTIFICATE

This is to certify that the work reported in the M.Sc. thesis entitled "**Studies on DNA repair mechanism and Non-Small Cell Lung Cancer**" submitted by Namra Khan at Jaypee University of Information Technology, Solan, India, is a bonafide record of his original work carried out under my supervision. This work has not been submitted elsewhere for any other degree or diploma.

Supervisor:

Dr.Tiratha Raj Singh Date -

Professor

Department of Biotechnology &Bioinformatics

Jaypee University of Information Technology

Solan, India

Acknowledgment

This project report has been kept on track and been seen through to completion with the support and encouragement of numerous people.

First and foremost, I would like to express my utmost gratitude and appreciation to my project supervisor, **Dr. Tiratha Raj Singh** for his guidance, supervision, and assistance throughout my project work. His expertise and ever-ready guidance contributed a major part in making this project a success.

My sincere thanks also goes to **Dr. Sudhir Syal**, Head of the Department, Department of Biotechnology and Bioinformatics, JUIT, Solan, for allowing me to work on such a diverse project.

Namra khan (217856)

M.Sc. Microbiology II year (4^{th} semester)

Jaypee University of Information Technology

Solan, HP

TABLE OF CONTENTS

CONTENT	Page Number
ABSTRACT	Vi
CHAPTER 1	1
INTRODUCTION	1
1.1 Introduction	1
CHAPTER 2	2
LITERATURE SURVEY	
DNA damage resources	2-4
2.1 Fixing base DNA damage	4
2.1.1 Rendering DNA damage	4
2.2 Base excision repair	5-6
2.3 Nucleotide excision repair	6-7
2.4 Mismatch repair	7-9
CHAPTER 3	
INTRODUCTION	
3.1 Chemotherapy	10-11
3.1.1 Risk of lung cancer, DNA repair ability	11-12
3.1.2 The impact of cisplatin and NER capacity	12
3.1.3 DNA repair polymorphism	12
3.2 Approaches to NSCLC management	13
3.3 Liquid biopsy	13-14
3.4 Approaches for treating NSCLC	14-15
3.5 XRCC1 DNA repair protein	15-16
CHAPTER 4	
INTRODUCTION	
4.1 Cytoscape	17-18
4.1.1 Result of cytoscape	18
4.1.2 Visualizing clusters of interacting proteins	18-19
4.1.3 Results of cytoscape	20-21
4.2 String tool	21-23
4.2.1 Result of string tool	23-24
4.3 KEGG pathway	25-26
4.4 Fanmod	26-28
4.4.1 Result of Fanmod	27-28
CHAPTER 5 CONCLUSION	29
REFERENCES	30-31

ABSTRACT

Human beings are exposed to a variety of DNA damaging compounds on a regular basis, which can have an impact on their health and affect disease states. On the other hand, DNA repair systems consistently protect DNA either by eliminating or by accepting the harm done to ensure a general survival. It is recognized that deviations from this fine-tuning can cause cellular metabolic equilibrium to become unstable. This is seen in a variety of tumours where disruption or dysregulation the DNA repair pathways leads to genomic instability. Testing for genotoxicity and controlling the use of commonly used physical, biological and chemical agents have become crucial as they have an influence on human health. In this work, we will outline DNA damage processes and the mechanisms that counteract them in order to shed light on the molecular underpinnings of genotoxicity within cell and establish the groundwork for the other articles in the issue. This study is associated with the DNA repair mechanism and associated non small cell lung cancer (NSCLC). Bioinformatics based analysis were performed to annotate biological entities and networks for their crucial regulatory role in NSCLC.

CHAPTER-1

INTRODUCTION

1.1 Introduction

Cellular DNA damage has an impact on mutagenesis and development of cancer. In human cell every day, both endogenous and external factors cause thousands to hundreds of thousands of DNA-damaging events. Alteration to the cellular genome may have an impact on the transcription of the gene and the following translating of such a DNA into proteins essential for signalling and for cellular functions. A mutation may also passed on to the daughters generations of cells if it is not fixed before mitosis. If cells losses the ability to effectively repair the damaged DNA, here may be reactions [1]. Senescent, or irreversibly dormant, cells can occur. Multiple laboratories reported in 2005 that senescence, which inhibits mitosis and prevents the cell from evolving further, can take place both in in-vitro & in-vivo in cancer cells. The cells might undergo apoptosis. If there is enough DNA damage, the cell may be forced to undergo programmed cell death by an apoptotic signaling cascade [1]. The cell might be develop malignant properties, for example immortality & unchecked divisions. Process through which cellular DNA is damaged and repaired to cause ageing, apoptosis, or cancer.

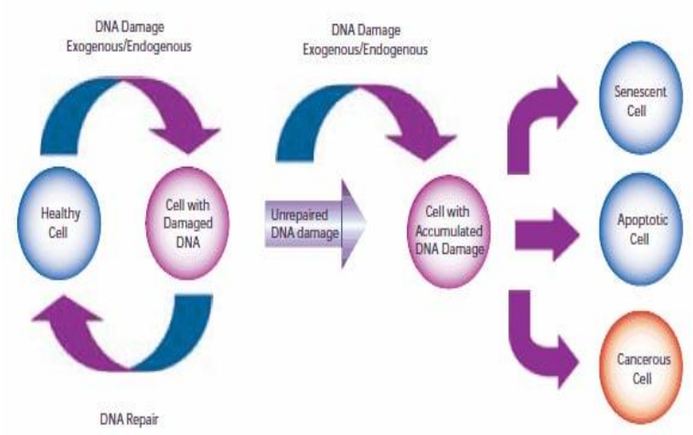


Fig.1.1 showing DNA repair pathway

Cells have evolved numerous DNA repair mechanisms, such as very little process redundancy, mismatched, nucleotide excision, and base excision repair mechanisms, to explain the types and severity of damaged DNA that takes place[2]. If cells sustain too much damage, they may have evolved to undergo apoptosis or senescence rather than using energy to efficiently repair the harm. A cell's capacity for repair depends on a variety of components, including cell type and age.

Chapter-2

DNA DAMAGE RESOURCES

The primary cause of DNA changes that result in cancer has long been assumed to be exogenous sources of damage. On the other hand, According to Loeb and Jackson, endogenous DNA damage causes a sizable portion of the cancer-causing mutations. five Similar types of DNA damage can come from cellular and environmental causes. Physical and chemical mutagens can damage DNA. Physical mutagens, such as Ultravoilet (200 to 300 nm wavelength) rays beyond the sun, are predominantly radiation of sources[2]. UV radiation creates covalent bonds which crosslink nearby bases of pyrimidine in a DNA strands. Ionising radiation changes DNA through producing free radicals inside of the cell, which produce oxygen like species that are reactive, double-strand and broken single-strands within the double helix. Alkyl groups can be covalently attached to DNA bases by chemical mutagens; examples of Nitrogen mustard like compounds, which could ethylate or methylate the DNA bases, are examples of DNA alkylate agents. Procarcinogen are metabolically transformed from chemically inertion precursors into high-reactive carcinogens. As a result of their interactions with DNA, these carcinogens can produce Adduct, that are chemical compounds bonded to DNA. Benzo[a]pyrene is polyaromatic heterocycle, although it's not cancer-causing in and of itself [2]. The consequence is benzo-[a]-pyrenediol epoxide (BPDE), the cancer causing metabolites that can create a covalent DNA adduct through two successive oxidation processes carried out by cytochrome P450 enzymes.

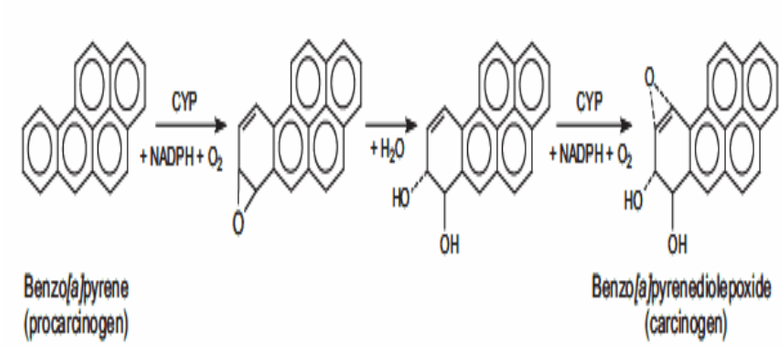


Fig.2.1Showing extremely carcinogenic benzo[a]pyrenediolepoxide is produced by the oxidation of benzo[a]pyrene by P450 enzymes.

Endogenous metabolic and biochemical processes, some of which are less understood, also lead to DNA damage. 6 Hydrolysis processes can remove the nucleotide base from the DNA strand partially or completely [3]. Depurination is the process when a purine base (such as guanine or adenine) spontaneously breaks the chemical connection that connects it to the deoxyribosyl phosphate chain. In a mammalian cell, there are thought to be 10,000 depurination events every day. 7 Depyrimidination, or the removal of a pyrimidine base from cytosine and thymine, also happens, with a frequency 20 to 100 times lower than depurination. Deamination takes place inside the cell when amine groups from the adenine, guanine, and cytosine rings are lost. As a result, hypoxanthine, uracil, and xanthine are produced, respectively. These abnormal nucleotides are detectable by DNA repair enzymes, can then fix them.C-T point mutation could, however, result from an uncorrected uracil nucleotide being read incorrectly as a thymine during later DNA replication. A reaction between S-adenosyl methionine and DNA methylation, a particular type of alkylation, takes place inside the cell (SAM). The methyl group of SAM, an intracellular metabolic intermediate, is extremely reactive. The spontaneous deamination of the methylated 5-methylcytosine product is a substantial source of mutation error. When amine group is lost, thymine base got produced, which DNA repair enzymes do not recognise as abnormal base. A C-T point mutation results as a result of the substitution being maintained during DNA replication.

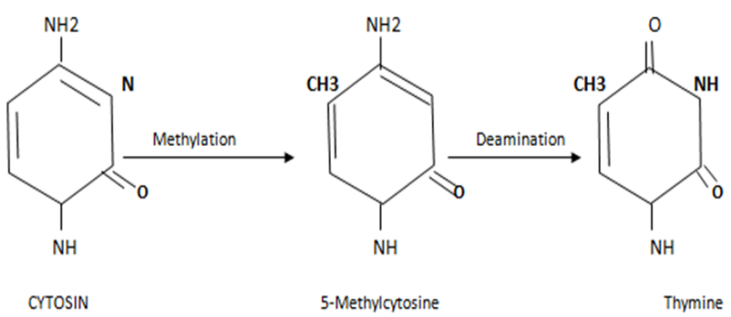


Fig.2.2 Showing how a cytosine to thymine phase 2 mutations results in C and T point mutations

Reactive oxygen species (ROS), produced by routine metabolic activities, oxidise bases to change their chemical composition. Oxidation can happen to both pyrimidine and purine bases. The nucleotide 8-oxo-deoxy guanine is produced most frequently when guanine is oxidised toguanine-8-oxo-7,8-dihydro (8-oxo-dG). Instead of base-pairing with deoxycytotidine as would be predicted, the 8-oxo-dG can pair with deoxyadenosine. The DNA that is reproduced after these errors are not found and fixes by mismatch repairing enzymes will have a C-A point mutation. DNA single-strand or double-strand breaks, depurination, and depyrimidination are further effects of ROS [3]. When the cell cycle is in the S phase, DNA replication may result in the introduction of additional genetic alterations. With a tiny but substantial error rate, template DNA polymerases may

integrate an erroneous a Watson-Crick-paired nucleotide vs template DNA. The polymerase may substitute chemically modified nucleotide precursors for regular bases while creating DNA. Additionally, When replicating DNA fragments with a lot of repetitive nucleotides or sequences, dna polymerase is prone to "stuttering" (microsatellite regions). Strand slippage, which happens when the template and copied DNA strands shift out of alignment, is what causes this enzymatic "stuttering. As a result, the daughter strand has either too few or too many nucleotides since the polymerase is unable to incorporate precise no. of nucleotides specified by template DNA. It is possible for the DNA to be cut into single or double strands [4].The deoxyribose molecule of the DNA deoxyribosylphosphate chain may sustain damage, leading to single strand breaks. Following the elimination of deoxyribose phosphate by AP, breaks also occur as a middle stage of the base excision repair process. "-endonuclease 1". The deoxyribose backbone and the nucleotide base are both removed from the DNA structure when a single strand break happens. Since the DNA may be more prone to damage while it is unravelling to be used as a replication template, double stranded cleavage most frequently happens when cells are going through the S-phase.

2.1 FIXING BASE DNA DAMAGE

2.1.1 RENDERING DNA DAMAGE

Alkylated bases and UV photolesions are two minor subcategories of Genetic alterations that are simple to fix. The following great literature references are recommended to readers for further information regarding UV photoreactivation through photolyase damages reported in marsupials and lower organisms: Yi and He; Errol C. Friedberg (2005); Kato et al. (1994); (2013). The repair of alkylated DNA damage will be briefly discussed here. Alkylated bases are reversed by two distinct types of enzymes in animals and people. The O6 alkylguanine DNA alkyl transferase enzymes first removes O alkylated DNA lesions, like, ethyl, di-chloroethyl, O6-methyl benzyl, groups of aliphatic, pyridy loxo butyl adduct of a guanine, or it can even fixes the O6-G-alkyl-O6 (2011). By moving the alkene from O2 from the Bases in dna to cysteine residues within the catalysis pocket, a single molecule eliminates the alkyl adducts inside single step reactions [Kaina et al., 2007] [4]. In the realm of cancer, AGT has a unique and complicated relevance. On the one hand, AGT is utilised to create pseudo substrates that can be combined with the therapeutic alkylating drugs to overcome resistant to chemotherapy of cancer [Tubbs et al, 2007]. This is done by taking use of the enzyme's ability to target a wide range of substrates. The absence of AGT expression, on the other hand, has been linked to a specific subset of malignancies Mokarram et al. (2013) and Lee et al. (2011). Additionally, the AGT enzyme is inhibited by the family of AGT homolog’s known as alkyltransferase-like proteins (ATLs), which divert reroute the NER pathway for the repair of large alkyl damage Margison et al.(2007);Tubs et al (2009) [4].

2.2 BASE EXCISION REPAIR

BER fixes single base damage from alkylation, oxidative, deamination, and abasic sources that do not appear to significantly disrupt the DNA helix. This repair activity inside the nucleus is mostly active during the cell cycle's G1 phase (Hubscher. and Dianov, 2013) [5]. In BER transactions, the DNA damage site undergoes chromatin remodelling, which is proceeded by such a DNA glycosylase that detects this damage [Odell et al., 2013]. According to research by (Huffman et al.(2005), Krokan & Bjoras(2013), at least 11 distinct DNA glycosylases are capable of identifying and excising damaged bases from both twisted and untwisted helices.) . On the other hand, some DNA glycosylases are bifunctional and have a glycosylase as well as an extra lyase activity. The second group includes, as examples, DNA glycosylase 1, Nei like DNA-glycosylase I, or DNA glycosylase II (NTHL1, NEIL1, NEIL2) (NEIL2) (2012) Jacobs and Schar. Both NEIL3 & 8-oxoguanine DNA glycosylase (OGG1) carry out both single- and the dual glycosylation tasks, it should be noted. 2011 [Svilar et al.] According to Dianov and Hubscher (2013), dual-purpose glycosylases begin the BER long patches repair route, whereas monofunctional glycosylases form an abasic site that is committed to the short patches repairing pathway. Short repair uses the abasic sites as a substrates for the AP endonucleases [5]. This enzyme cuts phosphodiester link 5′ to the abasic site, releasing a deoxyribose phosphate at 5′ end and producing hydroxy residue at 3 end.

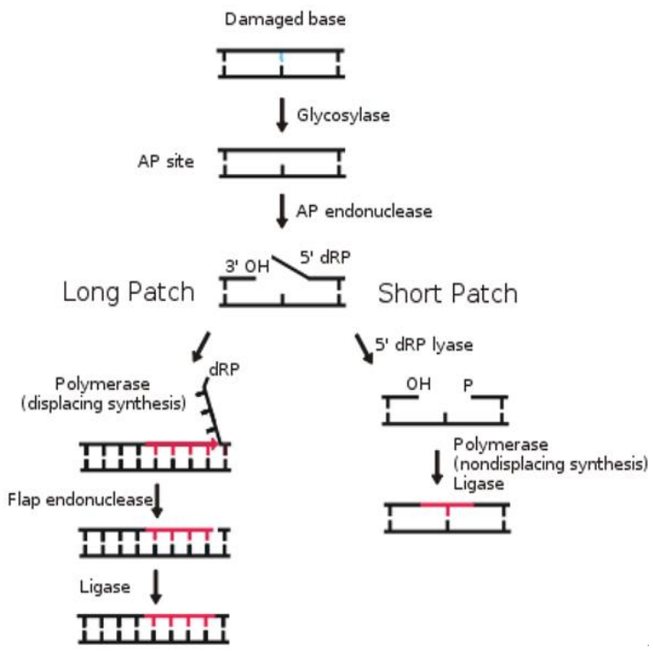


Fig. 2.3 Showing Base Excision Repair of damaged DNA

(Source-Hans E. Krokan and Magnar Bjørås. Base Excision Repair. *Cold Spring Harb Perspect Biol*. 2013, 5(4): a012583)

This 5′ DRP lyases activities of a POL filling this repair gaps, filled by POL DNA ligase 1 as well as by complexes of LIGIII, DNA ligase III & XRCC1, then ligated. The 3′ phosphodiester activity of APEI is used as a broad patch repair to fill the gap left by the

dual-purpose glycosylase. Later, POL (in non-proliferating cell) and POL synthesized in strand displacements manner, which followed by the removal of flap by both the flap nucleases & a LIG1-mediated ligase. Triplet repeat instability is thought to be caused by Base excision Repair of 8oxo-G mutations at CAGS repeats, while reduced expression of OGGI is associated with ageing process, neurological disorders, and the cancer. Kovtu et al. (2007), Tian et al. (2009), Curtin (2011), Mollersen et al. (2013), Chatterjee N. (2013), Krokan & Bjoras (2013). POL mutations are specifically seen in solid tumours and act as a prominent, sequence-specific mutator. BER's sub-pathway, which repairs broken single strands and damage purine bases, has also been demonstrated to require PARP1. Using this repair pathway importance in preservation of the overall genome integrity is further enhanced by the knowledge that mitochondria also carry out long and short patch BER, with POL performing the stage of synthesis.

MULTIPLE AND SIGNIFICANT BASE DAMAGE REPAIR

2.3 NER

NER is the method of a choice for eliminating large lesion brought along by Uv rays, benzo[a] pyrene adducts, as well as chemotherapeutics damages. Human syndromes caused by NER deficiency include Brain, oculo-facial, and skeletal syndrome (COFS), UV-Sensitive Syndrome (UVSS), Cockayne Syndrome (CS), and Xeroderma Pigmentosum (XP), which is linked to an increased risk of developing skin cancer [Vermeulen and Fousteri, 2013; Errol C. Friedberg, 2005]. To be more specific, the NER pathway, like the BER pathway, contributes to instability processes in illnesses with triplet repeats. Hubert et al. (2011), Dion (2014), Lin et al. (2006). Chromatin remodelling brought on by both the chromatin & NER components opens up the possibility for the NER mechanism to be utilised on the specific DNA lesions in order to initiate NER [Scharer, 2013]. NER can be classified into two basic categories: transcription-coupled NER and global genomic NER [6]. The major DNA damage sensor in GG-NER is the Xeroderma Pigmentosum, Completion Class C protein, which is coupled with CETN 2 and RAD23 B. The transcription initiation factor II H (TFIIH) complex, which is made up of 10 protein subunits and may switch between transcription initiation and NER activities, using XPC bound to lesion as its substrates. double excision & gap fillings are coordinate in the last stage to stop the creation of ssDNA gaps that would cause DDR. During the NER's stage of GG-incision, most of the protein are put together and committed. By using the structure-specific restriction enzymes XPF-ERCC1 & XPG, the damaged strands is cut away from the lesion's 5′ and 3′ ends, respectively [Fagbemi et al., 2011] [7]. The polymerase employed depends on how proliferative the cell is. For instance, in non-replicating cells, POL-dependent repair is predominant, whereas in the replicating cell, POL are the primary NER polymerases. cells that are dividing, LIG1 dependent ligation takes place. However, the XRCC1-LIG3 complex fills the gap because to the expression level of both dNTPs & LIG1 in non-proliferating cell. TC-NER pathways, which are second NER pathways, started by lesion stalled RNA

polymerase2. The first step in this pathway involves enlisting both TC-NER-specific proteins CSA and CSB , which are required for the subsequent formation of other TC-NER components. These comprise the core NER factors as well as TC-NER-specific proteins like UVSSA ,XAB2, and USP7, with the exception of the UV-DDB and XPC complexes, which are unique to Once the CSA-CSB complex has stabilized at the lesion site, RNA polymerase II is backtracked (or reverse translocated), revealing the lesion location. The lesion attracts the TFIIH. Once the lesions are removed from the transcribed strand, the remaining series of event projected to follow exact same order as in the GG-NER.

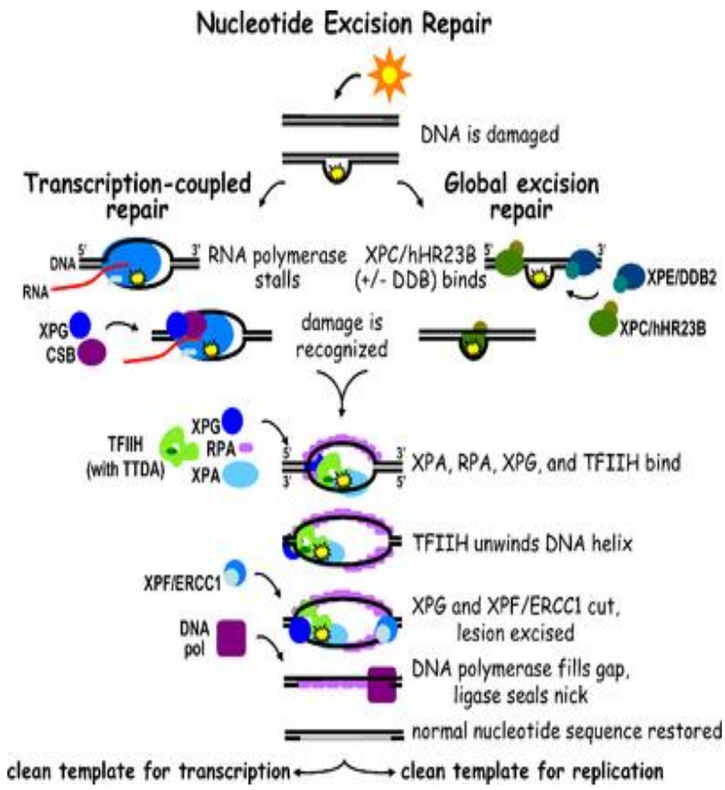


Fig. 2.4 Showing Nucleotide Excision Repair of Damaged DNA

(Source-Le May N, Egly JM, Coin F (2010). "True lies: the double life of the nucleotide excision repair factors in transcription and DNA repair")

2.4 MISMATCH REPAIR

To find and fix base mistakes that may happen in Replication of dna & recombination along with different kinds of DNA damage, a method referred to as MMR. Repairing a mismatch depends on the strand. Errors are frequently present in the freshly synthesised (daughter) strand during DNA synthesis. Before beginning repair, the mismatch repair device isolates the dna synthesis strand first from template (parental) [7]. Transient hemimethylation separates the strands in gram-negative bacteria. However, it is unclear what the precise mechanism is in other

prokaryotes and eukaryotes. It is considered that freshly synthesised lagging-strand DNA in eukaryotic includes nicks temporarily and serves as a signal to direct mismatch proofreading mechanisms to the proper strand. Recent discoveries support the notion that such nicks must also be evident in the leading strand. Recent studies have shown that nicks are sites for orientation-specific, RFC-dependent load of the proliferative cell nuclear antigen, a replication sliding clamp with one face positioned toward the 3'-OH end. In event of a mismatch and MutSalpha & MutSbeta, loades PCNA will then direct to the MutLalpha endonuclease's action to the daughter strand. The ability of a cell to retain genetic stability may be compromised by such mutational event that modifies the superhelical structure of DNA. The complexity of the damage detection and repair mechanisms is comparable to a replicating machinery itself, which emphasises a significance that changes has given to the DNA fidelities. Mismatch bases include Guanine\ thymine and Adenine\cytosine pairings, for instance (see DNA repair). The tautomerization of nucleotides during DNA replication frequently results in mismatches. The damage is repaired by locating the mismatch-induced deformity, recognizing both template & non-template strands, and switching the erroneous base for the proper nucleotide [7]. Only a small portion of a elimination process involves the mismatched nucleotide. The freshly generated DNA strand can have a less or may be thousands of deleted base pairs. From prokaryotes to eukaryotes, MMR is a highly conserved mechanism. S. pneumoniae, is the first proof of mismatch repair was discovered. Future E. coli research has identified several genes which, when mutationally inactivated, produce hypermutable strains.. The "Mut" proteins, which are main functional elements of the mismatch repair mechanism, are therefore the gene products. Three of these proteins—MutS, MutH, and MutL—are crucial for spotting the mismatch and guiding the repair machinery toward it. By recognising a mismatch nucleotide on a daughter's strand & forming a dimer (MutS2), MutS binds the altered DNA. At hemi ethylated locations throughout the daughter's DNA, MutH attaches, but its action are dormant. It is only turned on when it comes into touch with a MutL2 that binds the MutS-DNA complexes and serves as a link between MutS2 and MutH, turning the latter on. By looping out the DNA, the d (GATC) methylation site closest to the mismatch, that can upto one kb away, is sought after. The daughter strand is cut by MutH close to the hemi ethylated region after MutS-DNA complex activation. When the two strands are separated with a certain 3' to 5' polarity, MutL enlists UvrD helicase [8, 9]. After that, a complete (MutSHL) complex moves with DNA in direction of a mismatches, freeing the strand to be excised as moves. Ss-DNA tail is digested by an exonuclease that follows the complex. Whether MutH cuts the strand on the 3' or 5' side of the mismatch determines which exonuclease is called upon. ExoVII or RecJ (both 5' to 3' exonucleases) are utilised if the MutH-generated nick is located just at 5' end of a mismatch. However, if indeed the nick is at the 3' end of a mismatch, a 3' to 5' enzymes ExoI has been used. The mismatch site as well as the nucleotides surrounding it are entirely gone once it has been eliminated. The broken strand can then be repaired by DNA Polymerase III employing the other original strands as templates as well as the single-strand break that the exonuclease produced. Furthermore, DNA Ligase can fill the gap that the exonuclease left behind. The daughter strand is then swiftly methylated by DNA methylase [7, 10].

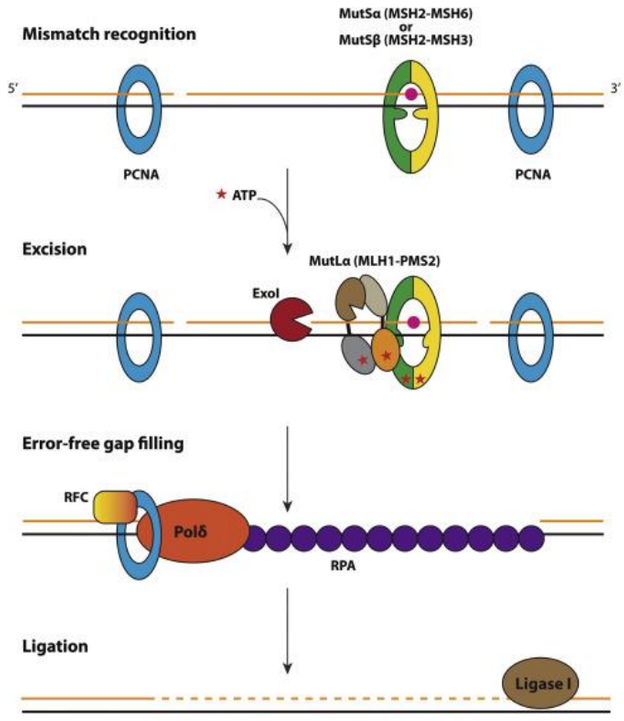


Fig. 2.5 Showing mechanism of Mismatch repair of damaged DNA

CHAPTER-3

3.1 Chemotherapy

The effectiveness of chemotherapy based on cisplatin appears to have plateaued, and empirical strategies are aiming to includel biological agents with various mode of action, however it's unclear how they might help in healing. The following two queries were raised when creating this overview on cisplatin resistance: Why we are writing this? For whom? Given the growing number of genetic cancer and chemoresistance reasons, we think medical oncologists ought to be involved. We can only get around these systems by becoming familiar with them. We shed a little understanding on the DNA repair deficits that lead to NSCLC as well as the influence of cisplatin in this work [10]. Some DNA mismatch repairing genes, like as ERCC1, have been demonstrated to be important in establishing mechanisms of resistance and can be utilised to modify cisplatin-containing chemotherapy regimens. The straight sequence of roughly 3x109 copy of all the four nucleotides which are C,T,G and A that are align in DNA is where cell repair capacity is stored. Numerous studies demonstrate that DNA damage activates DNA damage checkpoints, which regulate DNA repair transcription of genes, the progression of a cell cycle, and the triggering of apoptosis. The methods used to repair polymorphic DNA differ between people and with ageing. These traits affect chemosensitivity of cancer cells to DNA-reactive toxic medicines. [1]. DNA repair works against the development of tumours and helps patients resist cancer treatments [1]. Major DNA repair pathways come in a variety. The relationship between excision repair and cisplatin resistance is very strong, especially with regard to nucleotide excision repair (NER) [10].Additionally crucial to chemotherapy resistance is base excision repair (BER). In conclude, double-strand lesions caused by cytotoxic medicines, uv, & reactive oxygen species being repaired via homologous recombination & non-homologous DNA ends joining. Additional techniques include one-step repair mismatch repair that directly reverse DNA damage. As the amount of the alkyl group increases, MGMT's proportionate contribution to the same restoration of O6-alkylguanines inside DNA diminishes, whereas excision repairs become more important. Diagnosis using chloroethyl nitrosoureas shows a correlation between MGMT activity and OSR. To create deadly interstrand cross-links in the DNA of the target cell. It's interesting to note that MGMT levels in tumours differ significantly; this information has been employed in pharmacogenomic interpretation. Hypermethylation of MGMT, which was discovered in 40percent of brain cancer treated with BCNU, was associated with markedly improved survival. Curiously, tumour MGMT has been connected to temozolomide action [11]. Whenever temozlomide and CPT-11 were coupled, the resistance mechanism was circumvented in cancerous cells that were MGMT competent as well as MMR deficient. This possibility that altering DNA repair pathways is becoming increasingly significant inside the search for better chemotherapeutic results. Despite being a cutting-edge scientific technology, genes profiling using cDNA array is not presently applied in clinical settings. Throughout carcinogenesis and cancer progression, cancer cells develop a variety of errors in signaling pathways caused by genetics. Defects in DNA repair are one of tumour cells' Achilles' heels since they can lead to lung oncogenesis and benefit treatment, especially when combined with the most extensively

researched agent, cisplatin. Like so many DNA alkylators, cisplatin functions as a cross-linker to inhibit Replication of dna, which is a fundamental goal of chemotherapy for cancer. It is the model of cytotoxic medications used to treat non-small-cell lung cancer (NSCLC) [12]. Oxaliplatin, carboplatin, and cisplatin cause the cross-linking of guanine bases. Oxaliplatin's cross-link is structurally considerably different from those of cisplatin and carboplatin because of the large, adduct contains a 1,2-diaminocyclohexane groups. We have achieved a therapeutic peak after some more than 20 years of clinical trials, which is commonly perceived pessimistically among non-experts in the field of cancer diagnosis and therapy, who come to the conclusion that nothing is better than the outdated cisplatin/etoposide combination and who place their belief in new oral drugs. There have been some studies where the median survival, response rate, and time to progression were comparable among various chemotherapy protocols [12]. A 10% increase in 1-year survival was seen when cyclophosphamide, doxorubicin, and vindesine were combined with cisplatin at doses of 40 or 120 mg/m2. The median survival time increased by 1-1/2 months, and the absolute improvement at 1 year was 10%, based on combined information from 8 cisplatin studies with 776 patients. A nearly identical amount of patients (797) were assigned at random for this reason a British author presents the research approach MIC (mitomycin/ifosfamide/ cisplatin), radiation only, and hospice care, and the extent of benefit matched to the results of the meta analysis. New cisplatin formulations, including vinorelbine, have been studied recently. According to the Le Chevalier et al. trial, vinorelbine/cisplatin had a longer median survival time than vindesine/cisplatin or vinorelbine. Further recent times, another South west Oncology Group study revealed that the median survival of vinorelbine / cisplatin were similarly to this with the paclitaxel/ carboplatin [13]. When considering for clinical features as prognosis in multivariate, it is demonstrated in a study of SWOG experiments with just a sum of 2531 patients that cisplatin & female gender present as independent predictors , addition to effective status. With in research conducted by the European Melanoma Working Party, women's genders was also acknowledged as a distinct predictive factor. Cisplatin activity is increased when DNA repair capacity is inadequate, and females has less DNA repairing ability than that of males. (ECOG) trial, the standard measure in chemotherapy, included 1155 eligible patients and reported a median survival time of 7.9 months, a response rate overall of 19%, and survival rates of 33% after one year and 11% after two years. The three experimental arms and paclitaxel/cisplatin did not differ from one another. The cisplatin/ gemcitabine group, however, had a considerably shorter median time to progression. These findings seem to point to a correlation among chemoresistance and relative levels of recovery and reconstruction cross-complementing mRNA production. This is likely that the degree of ERCC1 expression which results in cisplatin failures rises if cisplatin is combined with gemcitabine since ERCC1 is among the main genes there in NER pathway and is required enable cisplatin & gemcitabine to incentivised [12, 13].

3.1.1 Risk of lung cancer, DNA repair ability, and chemoresistance

The therapeutic cytotoxic characteristics of several cancer chemotherapy drugs, including cisplatin, are caused by interstrand cross-links. Similar to this, a lot of carcinogens have

dual functions, forming DNA monoadducts as well as intra as well as interstrand cross links. Genetic factors regulate DNA repairing capacity, which influences lung cancer risk and therapy effectiveness. The ability to lymphocytes from peripheral blood to recover DNA has indeed been assessed using the host-cell reactivated test [13].

3.1.2 The impact of cisplatin and NER capacity

The disruption of both the DNA macromolecule with cisplatin is assumed to be primarily caused either by formation of interstrand cross-links and double-stranded adducts, which are repaired either by NER pathway. Furthermore, it is anticipated that compared to respective MMR-proficient equivalents, tumours with MMR deficiencies exhibit greater levels of cisplatin resistance. Damaged detection, double incision/excision, fix synthesis, and ligation are the steps which make it up the NER pathway. 30 or more protein.

3.1.3 DNA repair polymorphisms

Chromosome breakage, insertions and deletions, are prevented by XRCC3, a part of the homology double-strand repair pathway. Rad51, a crucial element of recombination repair, and XRCC3 are structurally similar. Exon 7T/ C polymorphism at location 18O67 of XRCC3 gene has been linked to an increased risk of melanoma development.

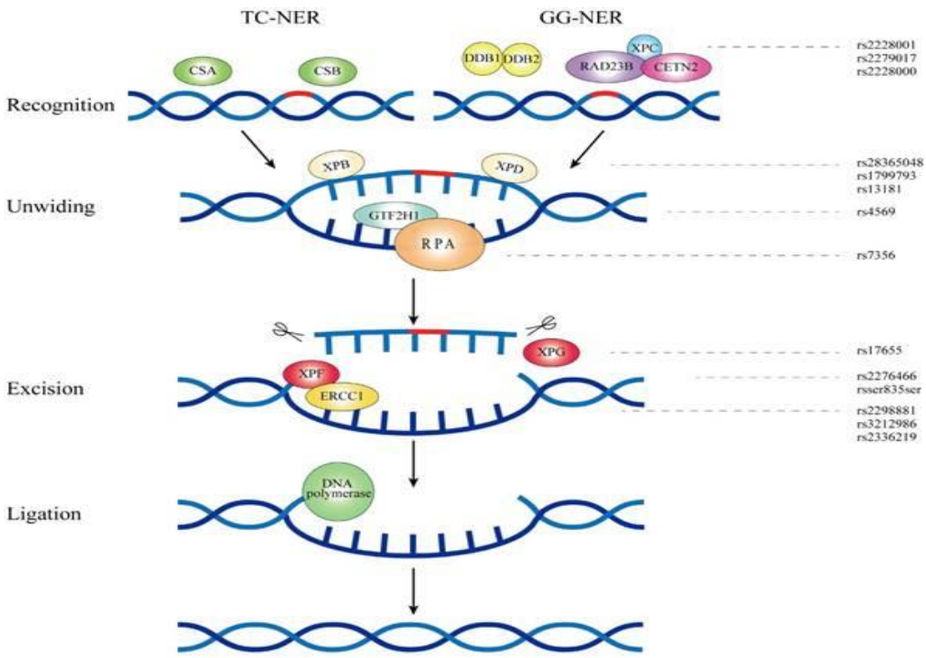


Fig. 3.1 Showing DNA repair polymorphism

3.2 Approach to NSCLC Management

Detecting lung cancer

For individuals who are at risk for lung cancer, three low-dose tomography (ct) scans are recommended annually, the National Lung Screening Trial (NLST) found a 20% reduction in lung cancer cases aftera 6.5-year median follow-up. United States recommends annual monitoring for persons between the ages between 50 & 80 who are currently smokers or have recently quit smoking and have smoking histories comprising 30, even more years based on such findings. Recent research from the NELSON experiment demonstrated that high-risk individuals who underwent screening had a considerably lower 10-year lung cancer mortality rate than those who did not (risk of death decreased by 24% in men and 33% in women) [14]. These studies show the undeniable efficacy of reduced CT testing in saving lives in a highly populated, with four testing cycles more than a five-year period preventing 60 deaths from lung cancer out of 6583 tested people. Unfortunately, this promising research has not been supported by actual practise. In the National Health Interview Survey, which was undertaken during 2010 to 2015, only 3.9% of a 7 million people who smoke who've been certified for screening for lung cancer in 2015 actually received the necessary care. This show in order of making well-informing decisions, smokers and medical professionalist alike need to be better educated on the advantages of lung cancerous screening. Patients having SARS-CoV-2-infected lung cancer were recently, monitored as part of the TERAVOLT (Thoracic cancers international covid 19 collaboration) registry project. The development of COVID-19 problems was found to be significantly predicted by a history of smoking [14].

3.3 Liquid biopsy

A liquid biopsy process involves examining several cancer indicators, such as circulation cancer DNA (ct DNA), micro RNA, & circulating cancerous cells (CTCs). Finding identifiable genetic variations that might one day guide medicine and help with response evaluation is much less intrusive and may be done on blood, serum, urination, CSF, as well as other sources. The best method for determining an initial diagnosis of NSCLC in modern clinical practise is still tissue diagnosis. A tissue-based investigation should be performed when possible because a negative fluid biopsy result does;nt exclude the existence of a carcinogenic change. In comparison to tissue analysis individually, 282 patients having untreated NSCLC participated with in non-invasive versus surgical lungs assessment (NILE) research, which revealed revealed ctDNA screening used to have a 48% greater rate of biomarker identification. It also had a quicker turning time [15]. The Fda approves the Guardant 360 CDx test, which uses liquid biopsies to detect genetic changes associated with any tumor cells.Another complete liquid biopsy test for pan-tumors that has been given Approval from the fda is indeed the Founding One Liquid CDx. However, the range of sensitivities is from 60% to 85%. The specificity (80%–

95%) of CtDNA testing for EGFR genetic alterations is good. The AURA3 study's findings demonstrated that the outcomes were predicted by the early clearance of ctDNA mutations [16]. Plasma samples taken at the beginning of the study, at 4 and 6 weeks after 2nd probably resulting therapy, compared to those with detected mutations (5.7 month, 95% CI 4.1-7.7), individuals with plasma EGFR clearing at 6 weeks had a median PFS of higher (11.1 month, 95% CI, 8.3-12.6). Patients plasma from the AURA2 study was also collected for real-time PCR testing to determine whether there were any of EGFR T790M-resistant alterations in additional to tissue samples. It was more likely for patients having additional disease to have a positive ctDNA test. The FLAURA3 study has identified several molecular alterations as mechanism of 1-line osimertinib resistance, including METs multiplication, HERII modifiers, PIK3CA, & EGFR C797S mutations. RET fusions, ALK /ROS1 complexes, BRAF V6OOE alterations, and MET exons fourteen skipped genetically changes have all been successfully identified using ctDNA. The ALK complexes was similar with tissue testing in the BFAST research, which examined the plasma of 2200 individuals. The response rate for individuals who got alectinib based on the findings of plasma tests was 92%, and the number of responses over a year was 78%. Recently, a portion of a POPLAR and OAK samples underwent a plasma assay to determine the frequency of tumour mutations (TMB). Patients receiving atezolizumab for NSCLC were found to have blood TMB as a predictive biomarker for PFS [16].

3.4 Approaches for Treating NSCLC

Depending on disease stages, morphology, genetic anomalies, and the patient's health, various therapeutic options for NSCLC include operation, radiotherapy, chemotherapeutic, immunotherapy, & targeted therapies therapy. Beginning NSCLC patients [Phase I, Phase 2, and Phase IIIA, which is frequently the case when Nitrogen gas lymph node malignancy is discovered during surgery] surgical intervention with a curative intent is indicated for those who are medically stable. Although additional platinum-based chemotherapy is recommended to patients with stages II-IIIA sickness which has an actual lower mortality risk of 5.4percentage points at five years, the risk of recurrence is material as well as the incidence of toxicity is extremely high. Broad discussions is suggested prior to therapy, especially with stage IIB & stage IIIA illnesses. Genomically targeted drugs haven't yet demonstrated an improvement in survival rates in initial patient. About 30% of NSCLC patients will have locally progressed illness. Overall majority of time Of diagnosis NSCLC patients have non-surgical candidates, and concomitant chemoradiotherapy is indeed the total gold standard of care [16, 17]. Immunotherapy is now being used as a last option. Owing to targeted therapy, a considerable fraction of NSCLC patients who have advanced illness have experienced better clinical results. Therefore, it is advisable to conduct molecular testing, preferably utilising a broad panel-based method, to detect these beneficial genetic variations. Tyrosine kinase inhibitors targeting genetic mutations such as, BRAF V600E, MET Exon

14, and NTRK are now able to treat a number of subtypes of NSCLC patients. For both cancerous and non-cancerous NSCLC, If no explicit targeting alterations are made, PD-L1 expression may be used to determine the form of therapy.

3.5 XRCC1 DNA repair protein

The most common DNA lesions are DNA single-strand breaks (SSBs), This may be caused directly by harm to such deoxyribose moiety or inadvertently by DNA repair of base excision. By increasing the frequency of chromosomal defects and speeding up the rate of mutation, unfixed SSBs constitute a significant threat to genetic integrity and cell viability [17]. The mechanisms for SSBR in mammalian cells are composed of a series of coordinating, sequential processes responsible for damage identification, terminal process, gaps filling, and joining. For example, even in short-patch process of BER, a glycosylase which recognises damages identifies and removes a damaged base, leaving an abasic region whose 5′ phosphate groups link is destroyed by apurinic/apyridinic restriction enzymes 1. DNA ligase III seals this ligatable nick at the end. In SSBR and BER, X-ray repair cross complementing 1 (XRCC1) is thought to play a significant role, according to a large body of evidence. A protein, which is presumed to have no discernible enzyme activities, is thought to operate as a scaffold for additional repair agents. Eight-oxoguanine DNA glycosylase, polynucleotide kinase/phosphatase, polymerases 1 and 2 (PARP-1 and 2), and XRCC1 all exhibit physical interactions. SSB repairing is also known to entail the enzymes Enzyme dna III, DNA polymerase, and polynucleotides [17]. Ataxia-oculomotor apraxia I is caused by a gene mutation called ataxin, has also been linked to XRCC1 in recent reports. Mice that lack XRCC1 are fatal during the embryonic stage. Alkylating chemicals, reactive oxygen species, and ionising radiation can all cause DNA damage, but mutant mouse as well as CHO cells that lacks functional protein XRCC1 proteins are extremely sensitive in it. In rodents cells with XRCC1 deficiency, SSB rejoining is highly hindered, pointing to a problem with DNA repair. Additionally, the frequencies for spontaneous sister-chromatid exchange or chromosomal aberrations are higher in these cells. Intriguingly, XRCC1 transgenic complemented XRCC1 mice, which produce XRCC1 at surprisingly low amounts, appear to grow properly, as well as the fibroblasts from such animals have virtually normal sensitivities to alkylating agents. These experiments revealed the XRCC1 isn't the rate-limiting element in mouse cell SSBR. Important findings from CHO cells point to DNA replication disruptions as the cause of the XRCC1 mutant cells' increased susceptibility to genotoxins [18]. This might be as a result of the replication fork encountering more unrepaired SSBs, them persisting longer, or not enough replicating origin firing. The relevance of XRCC1 at the this cell phase of the cycle is demonstrated by the fact that XRCC1 foci expand throughout S-phase & founder with PCNA around replication sites. Since there's no human cell lacking the XRCC1 protein, few such functional studies looked at how the XRCC1 protein functions in cells. SSB sites produced via light

irradiation within HeLa cells are where XRCC1 is drawn to. XRCC1 Numerous molecular epidemiological studies have focused on the relationship between snps and an increased likelihood of developing cancer over the past few yearsDepending on the type of cancer and the degree of environmental interaction with DNA-damaging substances, the existence of specific snps seems to be linked to a larger or reduced vulnerability to malignancy. These findings imply that the factor alleles may now modify the function of XRCC1. In this work, the first step of current research in with this phenomena, we show that XRRC1 is essential for DNA damage-induced cell survival, efficient DNA fixing, & genetic stability in human cells. We altered the levels of XRCC1 for human breast cancer cells using an RNAi [19, 20].

CHAPTER-4

4.1 CYTOSCAPE

Cytoscape tool offers 100s of applications that users can download to add additional features, such as clusterMaker24, which uses a variety of cluster techniques, & PTMOracle5, which enable PTMs analysis in the context with proteins network. As a wide-ranging network tool rather than a network database, Cytoscape needs ingest networks from other sources [21]. This makes a suitable pairing for analysing proteomics data between STRING and Cytoscape. We did this by developing stringApp, a Cytoscape application that makes it easier to import the content of STRING systems to Cytoscape, as well as additional user-provided data. The app also offers many of the features and the design of Cytoscape's STRING web interface. The programme enables a number of searches to find the networks, beginning with a set other proteins, a specific illness from the Diseases database,6, a PubMed search, or an array of diseases. Additionally, it gives users access to other data from related resources, such as information on a therapeutic target from the Pharos, smallest molecule interaction vn from the STITCHS, subcellular localization from the COMPARTMENTS, tissues expression from TISSUE, and short molecule interactions from STITCH. Through the use of Cytoscape's graphical user interface, these functionalities allow users to perform difficult network analysis and visualisation tasks with ease. In a particular use scenario, we show how the stringApp & Cytoscape may be used to analyse and visualise a proteomics data collection [21].

Function	Operation	Input	Output
STRING: protein query	Create a network for one or multiple proteins	Protein name(s) or identifier(s)	Network
STRING: disease query	Create a network of proteins associated with a disease	Disease name or identifier	
STRING: PubMed query	Create a network by entering a PubMed query	PubMed query	
STITCH: protein / compound query	Create a network for one or multiple proteins and compounds	Protein/compound name(s) or identifier(s)	
Expand network	Expand network by more interactors	STRING network Number and type of interactors to expand by	Network with nodes and edges added
Query for additional nodes	Add more proteins or compounds to an existing network	STRING network Protein/compound name(s) or identifier(s)	
Change confidence	Change confidence of the network	STRING network New confidence cutoff	Network with edges added or removed
Retrieve functional enrichment	Retrieve functional enrichment for several categories	STRING network	Enrichment table
Filter enrichment	Filter terms in the enrichment table by category or redundancy	STRING enrichment table Filter options	Filtered enrichment table
Draw charts	Show selected enrichment terms as charts on the nodes	STRING network Enrichment table	Network with enrichment visualization

Fig. 4.1 Showing main string app features with corresponding inputs and outputs

(Source- D. Szklarczyk *et al.*, "The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets," *Nucleic Acids Res.*, vol. 49, no. D1, pp. D605–D612, 2021.)

The stringApp sends a requests over to the STRING enrichment (API) to receive the function enrichmenting analysis findings the entire STRING networks or a specific subsets of it. Cytoscape tables named STRING Enrichments, that shows enriched words together with gene counts, related values of FDR, & gene sets, stores and displays data. The software gives the user the option to remove duplicate words, which reflect comparable sets of genes, and filter the enrichment results to reveal term from any combination of the six term categories, which are very useful for vast networks.

4.1.1 RESULTS

The two sources of information to which stringApp was designed to connect were the STRING database enabling high-quality protein-protein networks and a Cytoscape software package for integration with the network, analysis, and visualisation. Because of this, the stringApp's main job is to retrieve the network data from STRING, import it into Cytoscape, and preserve the majority of the STRING database's functionality and appearance. It also gives users the option to analyse the system using the full set of Cytoscape features and combine the results with their own data. However, stringApp imports data on protein-chemical interactions and protein-protein interactions from STITCHs for diseases or related PubMed queries.

4.1.2 VISUALIZING CLUSTERS OF INTERACTING PROTEINS

Proteome-scale Thousand of protein & 10 of thousands of connections make up PPI networks, which are difficult to meaningfully exhibit using standard graph presentation tools. GenePro enables this by use nodes, which represents sets of a genes or protein, into which a collection of a desired attributes may be mapped, visualised, or interactively queried. Additionally, the nodes in the graphs are linked to one another based on derived measure that provides pertinent details on the strength of connection between the nodes. Specific examples shown in a Figure 1a depicts graphs where singlel nodes represent yeast multi complexes of protein (clusters) that were obtained using the clustering techniques from a PPIs network that was recently described in a thorough, high-throughput & pull-down investigation (Krogan et al., 2006). When at least two protein interactions between clusters have been identified, the clusters/nodes are connected to one another; the number of contacts is inversely correlated with the edge thickness. This visualisation places each of the protein cluster in context as part of the overall network of interactions, highlighting those clusters that are intricately linked to one another and indicating the degree to which the anticipated complexes may interact with one another in vivo. Some

clustering techniques result in overlapping clusters, which share components. User can provide links between the nodes to express a quantity of a common protein in such situations.

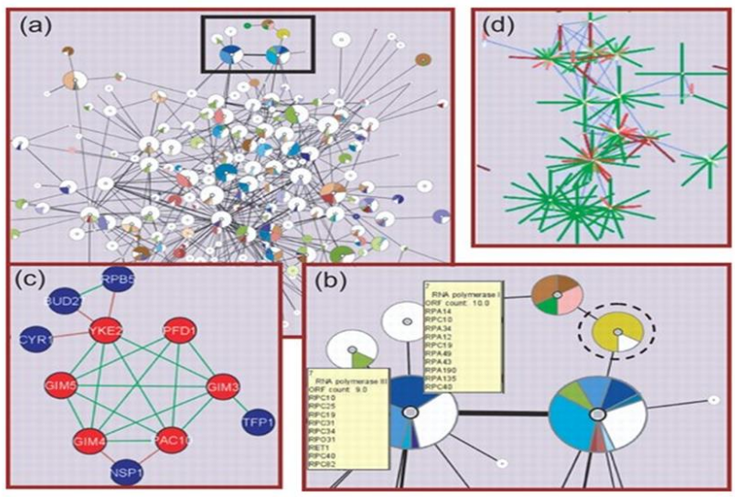


Illustration of features in GenePro, a Cytoscape plugin for advanced visualization and analysis of protein complex networks. (**a**) Protein cluster network of *S.cerevisiae*, with each node representing one cluster. (**b**) Illustrating nodes in (a) as pie charts, with each wedge of the pie representing the fraction of genes with a common feature, here the known CYGD complex to which they belong; the actual genes included in this fraction are listed. (**c**) A new graph representing interactions between individual genes belonging to the node/clusters highlighted by a dotted circle in (b). (**d**) mRNA levels for individual genes in some complexes are shown as red spikes (up-regulation) and green spikes (down-regulation), the length and color shade of the spikes illustrate the magnitude of the expression levels.

Fig.4.2Showing illustration of Cytoscape features

GenePro gives an extra set of capabilities to make it easier to validate and analyse the calculated clusters. A close view of the section of a web cluster seen in (Figure 1a) is shown in Figure 1b. every node in these view, which represents group of a proteins, displays as pie charts, with the size & colour of wedges indicating proportion of group's protein that have exact same feature. This property shown in Fig 1a & b to show how a calculated clusters & annotated complexes overlap. The number and name of all proteins associated with specific CYGDs complex (such as the RNA pol. I complex) are displayed when the mouse is placed over a specific wedge. When mouse is clicked over the same wedge, proteins associated with that complex are highlighted in all other nodes throughout the network. This last characteristic (data not given) gives a description of the distribution of protein from an identical CYGD complex among various clusters.

4.1.3. RESULTS OF CYTOSCAPE

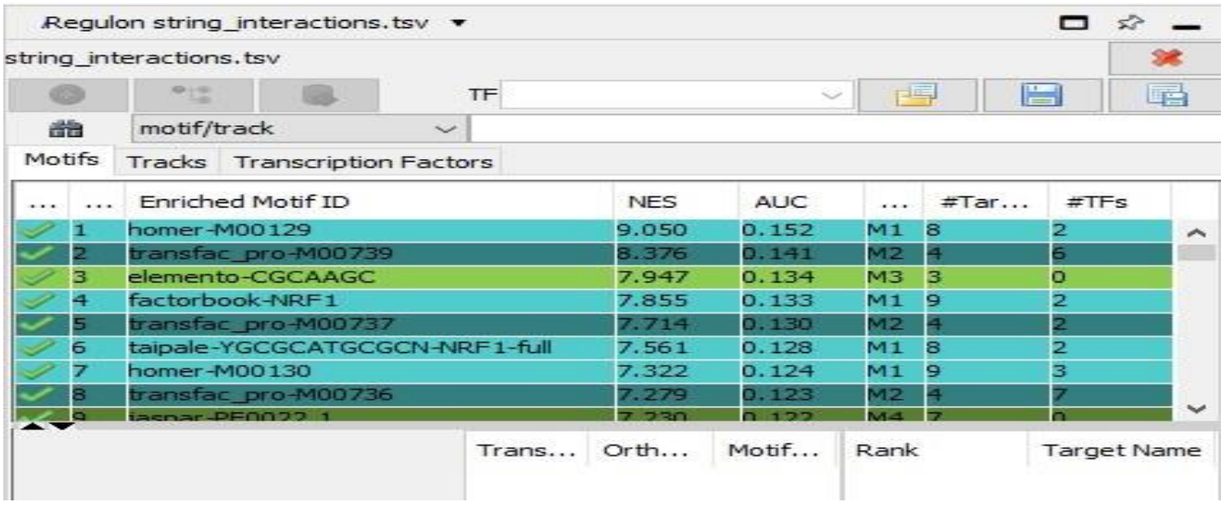


Fig. 4.3(showing enriched motif ID)

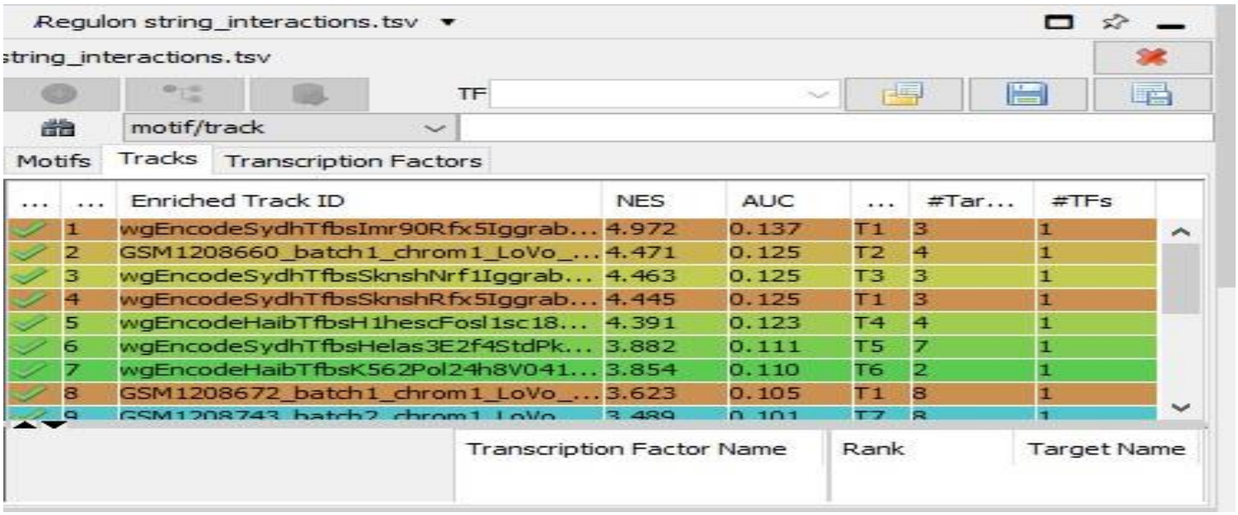


Fig. 4.4(showing the enriched track ID)

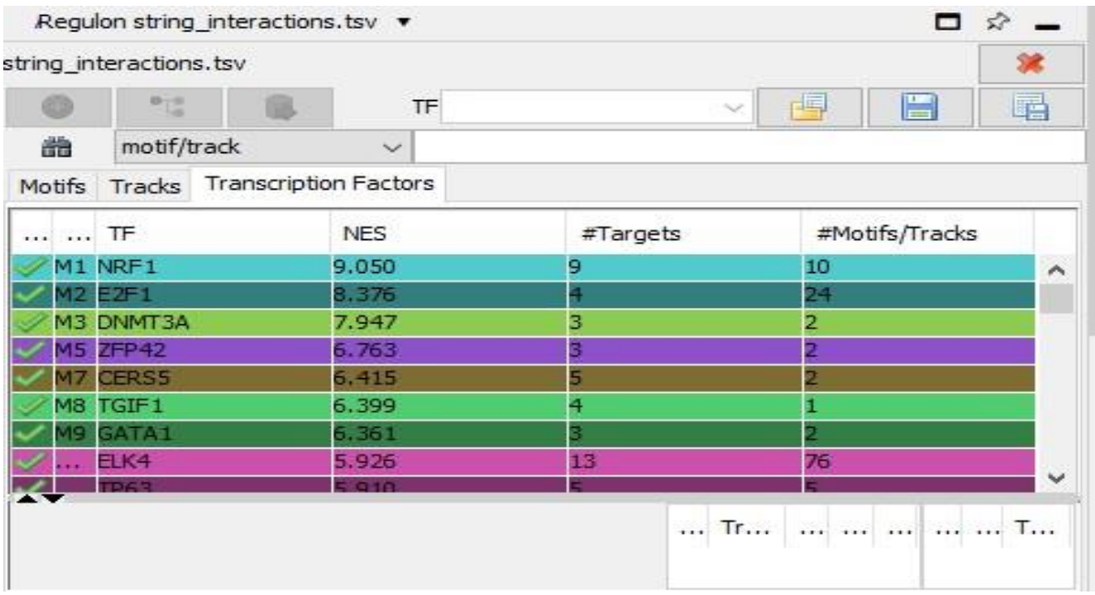


Fig.4.5(showing the transcription factors)

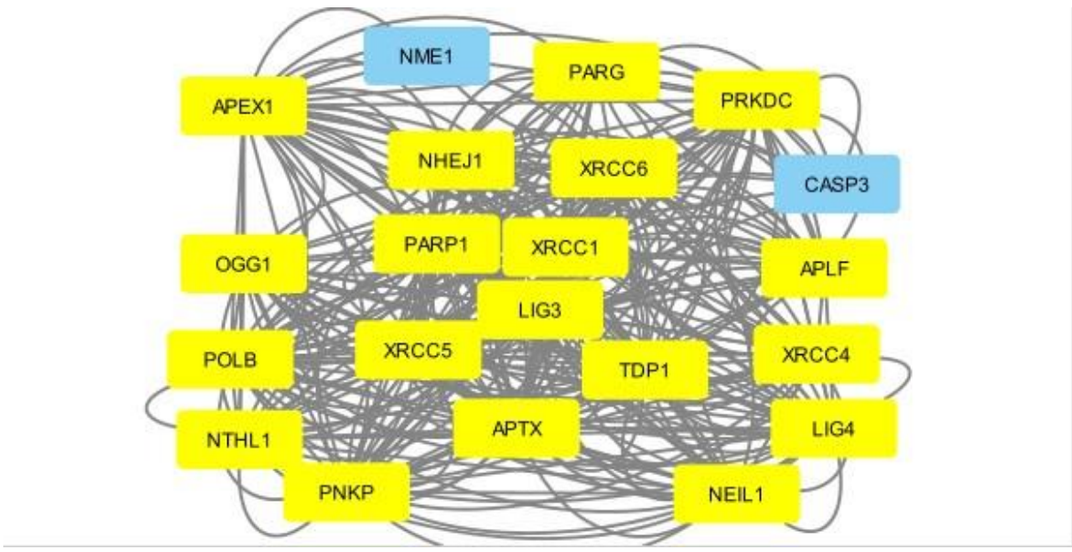


Fig.4.6 (interaction network obtain consisting of different regulatory genes)

4.2 STRING TOOL

New techniques for predicting protein function have been developed as a result of the accessibility of full genome sequences (1-6). These techniques, in contrast conventional, homology based functional assignment, anticipate the functional relationship between proteins rather than the function of the individual proteins based on genomics linkage of their own genes. Strategy is based in finding a genes that consistently co-occurs close to one another on chromosomes (in putatives,operon) which tends to encode functional interacting of proteins, such as those that are a member of same protein complexes & metabolic pathway (1-3,7-9) [23, 24]. In this article, presence of a web server which gets all the genes that consistently appear

within prospective operons for a particular query gene. The STRING server stands for "search tool for recurring instances of neighbouring genes." It also retrieves the genes which is indirectly connected to query genes through other genes using an iterative method. One seed gene serves as the tool's foundation. It finds the presents genes which is consistently occur with this present gene in clusters on the genome in various, phylogenetic distinct species, in the zero iteration. The term "run" refers to a group of genes on the same strand that are not broken up by more than 300 base pairs. It was first used by Overbeek et al. to designate gene clusters. Furthermore, we regard two genes that have truly fused together as one gene when they are part of the same run [24]. The tool continues this procedure in following iterations utilising all the newly discovered genes found in the preceding iteration as seeds, revealing set of genes that is indirectly connected to seeds gene. Iterations go on indefinitely (convergence) or until the user-specified maximum number of iterations is achieved, whichever comes first. Usually, query genes serves as seed. If query gene itself is'nt a member of conserve gene cluster, the tools use query genes orthologs which are members of conserved genes cluster as the seed. The tool does blast search against proteins from published genomes when a proteins sequence is provided as a query. If an ideal match discovered, the gene is employed as a seed. In alternative, the users may chooses seed from a list of blast hits. Programme also displays genes from the species from which other genes have already been obtained but which were not retrieved using conserve genes order based on a results of previous iteration. The existence or the absence of genes, which are'nt in conserved cluster, adds information to cluster. One of distinguished characteristics of the server that sets It's distinct from whats currently offered, servers like KEGG is its specific focus on a (iteratively) seeking & showing integral conserve genomics organisations for given gene. 1 image with all data is shown, along with additional details on the genes & their environment. The additional data comprises surrounding genes that are not conserved, gene order, a relative positions of the gene clusters within entire genome, & relatively directions of the gene transcription. Even for very basic, single-celled model organisms, an entirely complete understanding of all functionally significant interactions between proteins is still lacking [24]. However, knowledge is a necessary for the systems level comprehension of a cellular behaviour & required in order to contextualise the molecular roles of individual proteins within the setting of the cell.Several small-scale as well as high-throughput research have been conducted to identify direct physical binding among proteins, and the majority of these reported interactions can be obtained from specialised interactions databases (1-4) and also multipurpose databases based on a particular model organisms (5- 7). However, because expansion of a interactions data is significantly trailing behind the rate of the genome sequencing, there are currently no interaction data available for the majority of known proteins and genomes. Indirect links between proteins, such as genetic relationships or common pathway memberships, are also crucial for a thorough knowledge of cellular function, although they are typically not recorded in interaction databases [25]. Also, proteins do not simply interact physically. The scientific literature and a number of route databases (8, 9) are where you may find them instead. The STRING database, which stands for "Search Tool for the Retrieval of Associated Genes/Proteins," intends to gather, forecast, and assemble the majority of protein-protein relationships, including both direct and indirect connections. Using a collection of predictive

algorithms (10), STRING translates known interactions among model animals into different species based on projected orthology of the relevant proteins (11), covering creatures that have not yet been studied experimentally. Across a purely predictive resource that mostly covered prokaryotes, STRING has developed into a thorough tool that incorporates protein association data from all domains of life. The interaction network may be filtered at any desired stringency using the benchmarked numerical confidence scores that are provided with each interaction in the database. Relational database tables house all of the data in STRING. Although the interaction data is free to download, the whole database must be downloaded with a licence agreement to prohibit redistribution [26].

4.2.1 RESULT OF STRING TOOL

- number of nodes:21
- number of edges:159
- average node degree:15.1
- avg. local clustering coefficient:0.902
- expected number of edges:26
- PPI enrichment p-value:< 1.0e-16

Fig 4.7(showing the nodes and edges)

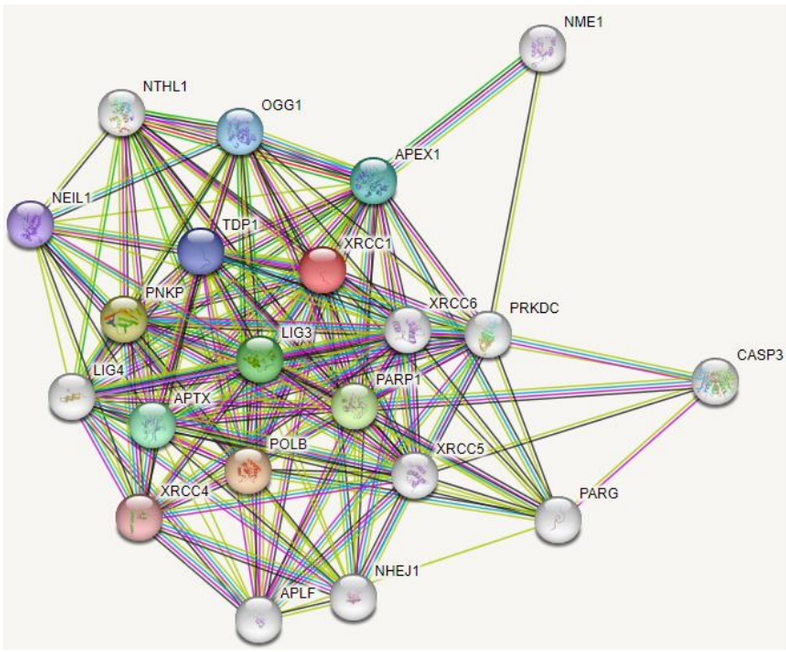


Fig.4.8(Interacting network of XRCC1 protein obtaining through STRING database)

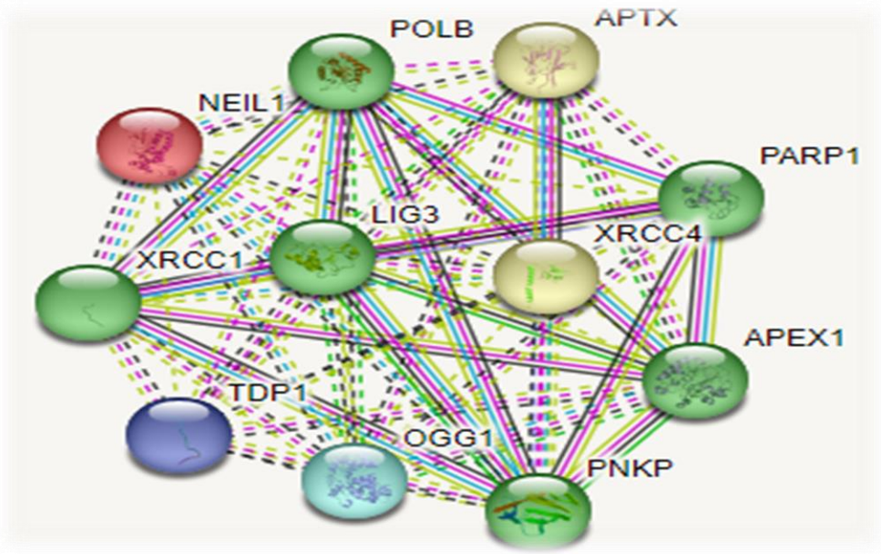


Fig.4.9(Interacting network of XRCC1 gene)

bubble	cluster Id	gene count	protein names
	Cluster 1	1	NEIL1
	Cluster 2	2	APTX,XRCC4
	Cluster 3	6	APEX1,LIG3,PARP1,PNKP,POLB,XRCC1
	Cluster 4	1	OGG1
	Cluster 5	1	TDP1

Fig.4.10(showing the clusters)

number of nodes:11
number of edges:55
average node degree:10
avg. local clustering coefficient:1
expected number of edges:11
PPI enrichment p-value:< 1.0e-16

Fig.4.11(showing the nodes and edges)

4.3 KEGG PATHWAY

KEGG path mapping, is the process of mapping of a genes in a genome to manually crafted route maps, was the initial development of KEGG (Kyoto Encyclopaedia of Genes & Genomes), which formed in 1995 as a iintegrated database source for a biological analysis of full sequence genomes. Only 4 databases that is made up KEGG at time were PATHWAY, GENES, COMPOUND, & ENZYME. Because ENZYME only included metabolic pathway maps, the KEGG network mapping was done through ENZYME. Later, KEGG was extensively expanded, adding BRITE and MODULE to PATHWAY, GENES to GENOME, COMPOUND to GLYCAN & REACTION, and ENZYME to KO for purpose of the KEGG pathway mapping [27]. Additionally, KEGG is becoming increasingly often used to analyse high-throughput data from transcriptomics, proteomics metabolomics, metagenomics, and other fields in addition to genomics. In order to allow for sequence similarity based KO assignment, KOs defined as groupings with both sequence similarity and functional orthologs. Although the word "ortholog" is used, there is no predetermined threshold for similar scores, and a KO can be made up of a single gene or solely genes from closely associated species [27].

Three categories—KEGG organisms, viruses, and addendum—now make up the GENES database. The majority of GENES is made up of the KEGG organisms category, which includes entirely or almost entirely sequenced genomes pulled from the GenBank and RefSeq databases. The 3 and 4 letter species code, such as "hsa" for Homo-sapiens, is used to identify each organism. Particular viral genomes aren't distinguishable in the viruses category, which is indicated as the two-letter code 'vg' in the bimonthly release of RefSeq. The NCBI taxonomy identification, which frequently corresponds to many sequenced genomes, is used to differentiate viruses when appropriate.A manually compiled sequencing dataset for functionally characterised proteins is the supplement category [28-30]. The collection of genomes is based on publications, and the bulk among the author-submitted sequences contain protein accession numbers from GenBank/ENA/DDBJ (ProteinIDs). This categories is identify by 2-letter code "ag". In accordance with concepts, each GENE item identified by syntax "org:gene" (Table I), wherein "org" denotes organism's coding and "gene" denotes GeneID, Locus_tag, the ProteinID depending on the data source.

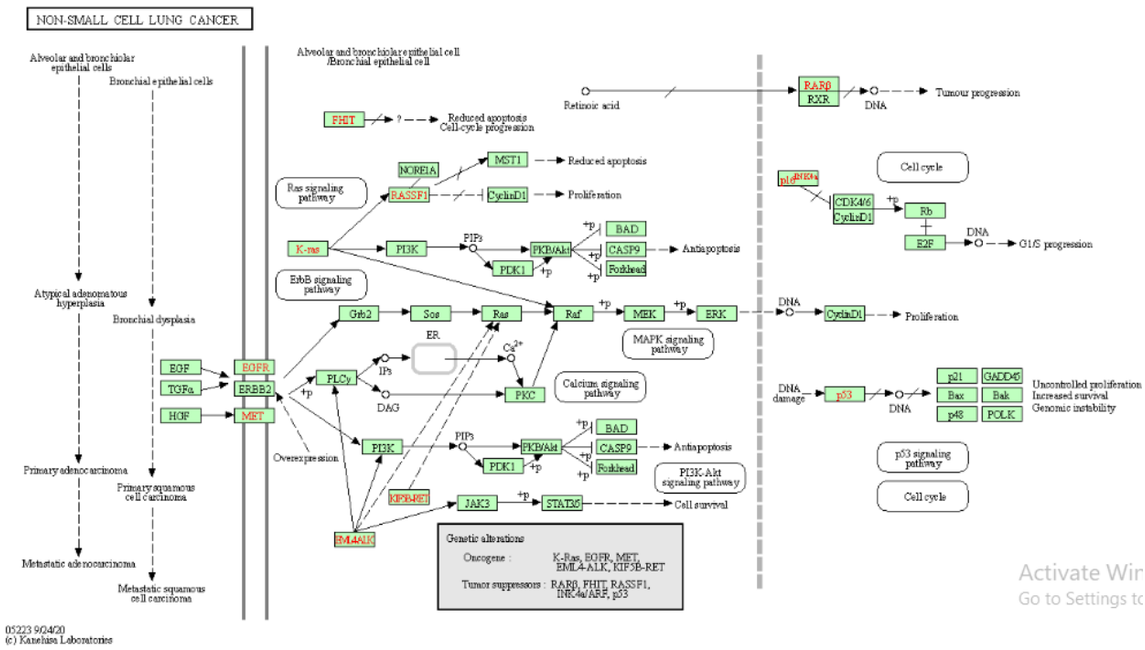


Fig.4.12 Showing KEGG pathway

4.4 FANMOD

Compared to random networks, many biological networks include specific tiny subnetworks far more frequently. The term "network motifs" was created by Milo et al. (2002, 2004) who recommended using such an abundance of "topological modules" (Vespignanii, 2003) to reveal a structural design concepts of a biological networks. Three computationally intensive subtasks make up the task of finding network motifs: The second subtask has received some attention, while the other two have received far less up until lately. Kashtan et al. (2004) suggested a technique for sampling subgraphs to accelerate the first subtask. However, this technique has a number of shortcomings, including the fact that it only offers non-uniform sampling and performs badly as motif size grows; Wernicke (2005) gives a more thorough study of these issues [33].

1. Determining the number and kind of subgraphs present in the input network.

2. choosing the subsections that are isomorphic, or topologically identical, and dividing the resulting subsections in accordance.

3. figuring out whether subgraph classes are more common than random graphs in the context of a specific random graph model.

FANMOD is the network motif recognition tool which use the innovative RAND-ESU method (Wernicke, 2006) to sample & count subgraphs. The approach to orders magnitudes

quicker than other one that has ever been used to perform this task, making it possible to find larger themes in larger networks than was previously feasible. In addition, FANMOD makes it feasible to recognise motifs in coloured networks, which is not achievable with other available tools. There are around 700 lines of non-library code in the C++ programme that creates the FANMOD utility. The wx WIDGETS framework (Smart et al., 2006) [33], and it is available for a wide range of platforms including Linux, Mac OS, and Windows, is used to create the GUI along with system dependent functionality. Up to eight vertices in size, FANMOD can recognise network patterns. Using the technique outlined by Wernicke (2005), all subgraphs of specified size can be either enumerated or evenly sampling in the input network. The subgraphs are categorised into isotropic subgraph classes using a rendition of the common graph-labeling technique NAUTY (McKay, 1981). FANMOD determines the frequency billion subgraph classes in a certain amount of random graphs.The user may select from a variety of switching techniques in order to retain specific graph attributes while generating random graphs by shifting edges between vertices in the original network. Depending on the amount of colours used for edges and vertices, motifs of up to seven vertices in size can be discovered in coloured networks. Colours have no impact on how quickly the tool runs; in fact, because canonical graph labelling is made easier, it runs a tiny bit quicker overall. The amount of edges between vertices with various colours can be preserved by random networks at an optional step. You may export the determined relevance of each subgraph in the network in a number of different forms. A rapid review and sharing of results is made possible by an HTML export function with a number of filters [34].

4.4.1 RESULTS OF FANMOD

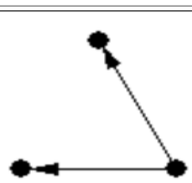
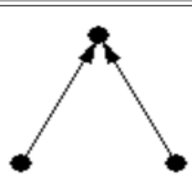
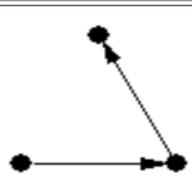
ID	Adj	Frequency [Original]	Mean-Freq [Random]	Standard-Dev [Random]	Z-Score	p-Value
6		31.933%	32.415%	0.0044633	-1.0805	0.789
36		16.807%	17.422%	0.0057237	-1.0743	0.768
12		49.58%	48.936%	0.010292	0.62581	0.464

Fig.4.13 Showing results of FANMOD tool

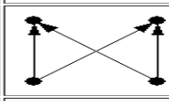
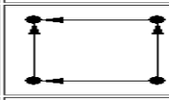
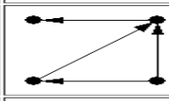
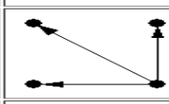
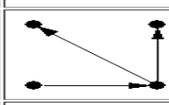
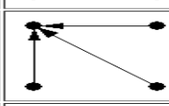
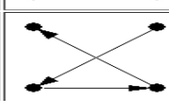
ID	Adj	Frequency [Original]	Mean-Freq [Random]	Standard-Dev [Random]	Z-Score	p-Value
204		2.439%	0.058383%	0.0013998	17.007	0
2182		2.8455%	0.1647%	0.0023097	11.607	0
2118		2.439%	0.48131%	0.0052171	3.7525	0.002
14		10.569%	7.7693%	0.007777	3.6001	0
28		11.382%	8.8461%	0.0075632	3.353	0
2184		2.8455%	1.9444%	0.0035636	2.5288	0
536		21.951%	17.457%	0.022211	2.0232	0.023

Fig. 4.14 Showing results of FANMOD tool

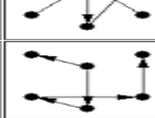
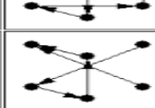
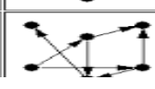
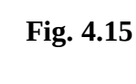
ID	Adj	Frequency [Original]	Mean-Freq [Random]	Standard-Dev [Random]	Z-Score	p-Value
4200488		2.2533%	1.0089%	0.0046279	2.6889	0.006
541130832		1.1655%	0.54203%	0.0023456	2.658	0.008
32884		1.2432%	0.66313%	0.0022192	2.6139	0.006
4325452		1.6317%	0.73836%	0.0035165	2.5404	0.018
4204556		3.4188%	2.0099%	0.0057187	2.4636	0.017
4198504		1.2432%	0.68039%	0.0023521	2.3928	0.009
1057832		3.4188%	2.0115%	0.0062201	2.2624	0.015
67405856		1.8648%	0.92803%	0.0042103	2.2249	0.028
8404016		0.6216%	0%	0	undefined	0
8402456		1.2432%	0%	0	undefined	0

Fig. 4.15 Showing results of FANMOD tool

CHAPTER-5

Conclusion

Deoxyribose damage may directly cause DNA single-strand breaks, or they may occur indirectly as anticipated side effects of DNA base excision repair. These lesions continually threaten the integrity of the genetic code since single-strand breaks are brought about by endogenous reactive chemicals such reactive oxygen species. Through its capacity to interact with a variety of enzymatic elements involved in repair activities, the XRCC1 protein significantly contributes to the facilitation of the repairing for single-strand breaks in cells of mammals. Here, XRCC1-mediated protein-protein interactions and the repair pathways in which they function are discussed. During the repair of base excision and at direct breaks, models for repairing of single-strand breaks are described. Therefore we studied the interaction partners of XRCC1 gene, then the data is collected and ruled over to the cytoscape tool through which we studied the molecular interaction of XRCC1 gene with other genes. Therefore we got the list of genes and proteins that helps in repairing the DNA damage and NSCLC.

References

[1] M. G. Marinus, "DNA mismatch repair," *EcoSal Plus*, vol. 5, no. 1, 2012.

[2] N. Pećina-Šlaus, A. Kafka, I. Salamon, and A. Bukovac, "Mismatch repair pathway, genome stability and cancer," *Front. Mol. Biosci.*, vol. 7, p. 122, 2020.

[3] G.-M. Li, "Mechanisms and functions of DNA mismatch repair," *Cell Res*, vol. 18, no. 1, pp. 85–98, 2008.

[4] M. Fousteri and L. H. F. Mullenders, "Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects," *Cell Res*, vol. 18, no. 1, pp. 73–84, 2008.

[5] W. Li, K. Jones, T. J. Burke, M. A. Hossain, and L. Lariscy, "Epigenetic regulation of nucleotide excision repair," *Front. Cell Dev. Biol*, vol. 10, 2022.

[6] M. Kusakabe, "Mechanism and regulation of DNA damage recognition in nucleotide excision repair," *Genes Environ*, vol. 41, no. 2, 2019.

[7] C. Kisker, J. Kuper, and B. Van Houten, "Prokaryotic nucleotide excision repair, " Cold Spring Harb," *Cold Spring Harb. Perspect. Biol*, vol. 5, no. 3, 2013.

[8] A. Karimian, Y. Ahmadi, and B. Yousefi, "Multiple functions of p21 in cell cycle, apoptosis and transcriptional regulation after DNA damage," *DNA Repair (Amst.)*, vol. 42, pp. 63–71, 2016.

[9] E. C. Friedberg, "A brief history of the DNA repair field," *Cell Res.*, vol. 18, no. 1, pp. 3–7, 2008.

[10] A. Torgovnick and B. Schumacher, "DNA repair mechanisms in cancer development and therapy," *Front. Genet.*, vol. 6, p. 157, 2015.

[11] N. Chatterjee and G. C. Walker, "Mechanisms of DNA damage, repair, and mutagenesis, " Environ," *Environ. Mol. Mutagen*, vol. 58, no. 5, pp. 235–263, 2017.

[12] S. Gurubhagavatula *et al.*, "XPD and XRCC1 genetic polymorphisms are prognostic factors in advanced non-small-cell lung cancer patients treated with platinum chemotherapy," *J. Clin. Oncol.*, vol. 22, no. 13, pp. 2594–2601, 2004.

[13] J. Chen, Q.-W. Zhao, G.-M. Shi, and L.-R. Wang, "XRCC1 Arg399Gln and clinical outcome of platinum-based treatment for advanced non-small cell lung cancer: a meta-analysis in 17 studies," *J. Zhejiang Univ. Sci. B*, vol. 13, no. 11, pp. 875–883, 2012.

[14] R. Chen *et al.*, "Emerging therapeutic agents for advanced non-small cell lung cancer," *J. Hematol. Oncol.*, vol. 13, no. 1, p. 58, 2020.

[15] C. Zappa and S. A. Mousa, "Non-small cell lung cancer: current treatment and future advances," *Transl. Lung Cancer Res.*, vol. 5, no. 3, pp. 288–300, 2016.

[16] J. Chen, H. Wang, and Z. Li, "Association between polymorphisms of X-ray repair cross complementing group 1 gene and pancreatic cancer risk: A systematic review with meta-analysis," *Pathol. Oncol. Res.*, vol. 25, no. 3, pp. 897–904, 2019.

[17] R. Ali, *Supplementary materials and Methods from targeting PARP1 in XRCC1-deficient sporadic Invasive breast cancer or preinvasive ductal carcinoma in situ induces synthetic lethality and chemoprevention*. 2023.

[18] A. P. Das, S. Saini, S. Tyagi, N. Chaudhary, and S. M. Agarwal, "Elucidation of increased cervical cancer risk due to polymorphisms in XRCC1 (R399Q and R194W), ERCC5 (D1104H), and NQO1 (P187S)," *Reprod. Sci.*, vol. 30, no. 4, pp. 1118–1132, 2023.

[19] J. Pan *et al.*, "Circular RNA FLNA acts as a sponge of miR-486-3p in promoting lung cancer progression via regulating XRCC1 and CYP1A1," *Cancer Gene Ther.*, vol. 29, no. 1, pp. 101–121, 2022.

[20] B. Liu *et al.*, "Association of APEX1 and XRCC1 gene polymorphisms with HIV-1 infection susceptibility and AIDS progression in a northern Chinese MSM population," *Front. Genet.*, vol. 13, p. 861355, 2022.

[21] R. K. Mandal and R. D. Mittal, "Genetic variant Arg399Gln G>A of XRCC1 DNA repair gene enhanced cancer risk among Indian population: Evidence from meta-analysis and trial sequence analyses," *Indian J. Clin. Biochem.*, vol. 33, no. 3, pp. 262–272, 2018.

[22] X. Kong, Y. Shi, S. Yu, J. Liu, and F. Xia, "Academic social networks: Modeling, analysis, mining and applications," *J. Netw. Comput. Appl.*, vol. 132, pp. 86–103, 2019.

[23] S. Yu, Y. Feng, D. Zhang, H. D. Bedru, B. Xu, and F. Xia, "Motif discovery in networks: A survey," *Comput. Sci. Rev.*, vol. 37, no. 100267, p. 100267, 2020.

[24] D. Szklarczyk *et al.*, "STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets," *Nucleic*

Acids Res., vol. 47, no. D1, pp. D607–D613, 2019.

[25] D. Szklarczyk *et al.*, "Correction to 'The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets,'" *Nucleic Acids Res.*, vol. 49, no. 18, p. 10800, 2021.

[26] D. Szklarczyk *et al.*, "The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets," *Nucleic Acids Res.*, vol. 49, no. D1, pp. D605–D612, 2021.

[27] X. Gu, B. Jin, Z. Qi, and X. Yin, "Identification of potential microRNAs and KEGG pathways in denervation muscle atrophy based on meta-analysis," *Sci. Rep.*, vol. 11, no. 1, p. 13560, 2021.

[28] Y.-Q. Qiu, "KEGG Pathway Database," in *Encyclopedia of Systems Biology*, New York, NY: Springer New York, 2013, pp. 1068–1069.

[29] M. Kanehisa and Y. Sato, "KEGG Mapper for inferring cellular functions from protein sequences," *Protein Sci.*, vol. 29, no. 1, pp. 28–35, 2020.

[30] M. Kanehisa and S. Goto, "KEGG: kyoto encyclopedia of genes and genomes," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 27–30, 2000.

[31] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa, "KEGG: Kyoto Encyclopedia of genes and genomes," *Nucleic Acids Res.*, vol. 27, no. 1, pp. 29–34, 1999.

[32] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, "KEGG: new perspectives on genomes, pathways, diseases and drugs," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D353–D361, 2017.

[33] S. Wernicke and F. Rasche, "FANMOD: a tool for fast network motif detection," *Bioinformatics*, vol. 22, no. 9, pp. 1152–1153, 2006.

[34] M. Megraw, S. Mukherjee, and U. Ohler, "Sustained-input switches for transcription factors and microRNAs are central building blocks of eukaryotic gene circuits," *Genome Biol.*, vol. 14, no. 8, p. R85, 2013.

[35]Sadhika Behl, Arushi Sharma, Prashanth Suravajhala, Tiratha Raj Singh (2021). Computational studies to explore the role of MSI associated DNA mismatch repair mechanisms in HNPCC through expression and interaction data. *International Journal of Bioinformatics Research and Applications, 16* (4), pp. 408-416, DOI: https://doi.org/10.1504/IJBRA.2020.113023

[36] Ankita Shukla, Ahmed Moussa, Tiratha Raj Singh (2016). DREMECELS: A Curated Database for Base Excision and Mismatch Repair Mechanisms Associated Human Malignancies. *PLoS ONE, 11* (6), pp. 1-8, DOI: https://doi.org/10.1371/journal.pone.0157031

[37] Manika Sehgal, Rajinder Gupta, Ahmed Moussa, Tiratha Raj Singh (2016). An Integrative Approach for Mapping Differentially Expressed Genes and Network Components Using Novel Parameters to Elucidate Key Regulatory Genes in Colorectal Cancer. *PLoS ONE, 10* (7), pp. 1-18, DOI: https://doi.org/10.1371/journal.pone.0133901

[38] Ankita Shukla, Manika Sehgal, Tiratha Raj Singh (2015). Hydroxymethylation and its potential implication in DNA repair system: A review and future perspectives. *Gene , 564* (2), pp. 109-118, DOI: https://doi.org/10.1016/j.gene.2015.03.075

[39] Manika Sehgal, Tiratha Raj Singh (2014). DR-GAS: A database of functional genetic variants and their phosphorylation states in human DNA repair systems. *DNA Repair, 14* (2014), pp. 97-103, DOI: http://dx.doi.org/10.1016/j.dnarep.2014.01.004