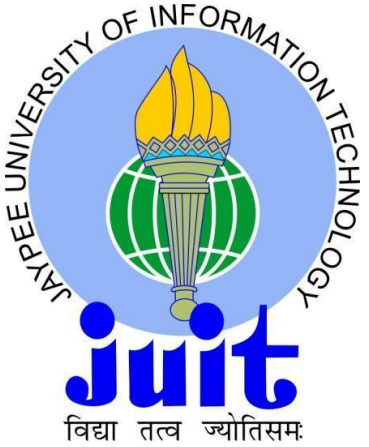
COMPUTATIONAL CHARACTERIZATION OF NUCLEOTIDE EXCISION REPAIR (NER) PATHWAY IN SKIN CANCER

Enrollment Number: 131504, 131519

Name of Students: Tanya Singh, Varsha Mehta

Name of Supervisor: Dr. Tiratha Raj Singh



Submitted in partial fulfillment of the Degree of

Bachelor of Technology

DEPARTMENT OF BIOTECHNOLOGY AND BIOINFORMATICS

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

WAKNAGHAT, SOLAN

CONTENTS

CERTIFICATE

This is to certify that the project report entitled as "***Computational Characterization of Nucleotide Excision Repair (NER) Pathway in Skin Cancer***", carried out for the partial fulfillment of the requirement for award of the Degree of Bachelor of Technology in Bioinformatics by Ms. Tanya Singh and Ms. Varsha Mehta, Jaypee University of Information Technology, Waknaghat, Solan (H. P.), has been carried out under my supervision.

This work has not been submitted partially or fully to any other University or Institute for the award of this or any other degree or diploma.

Signature: _____

Supervisor: Dr. Tiratha Raj Singh

Designation: Assistant Professor (Senior Grade)

Date: _____

Acknowledgement

We would like to express our sincere gratitude and thanks to our project guide, Dr. Tiratha Raj Singh, whose patience, guidance, encouragement, and dedication had motivated us to do this project under him.

We would also like to express sincere thanks to Dr. Rajinder S. Chauhan (Head, Department of Biotechnology and Bioinformatics) for his mandate and support to work on this project.

We would also like to thank Mr. Ankush Bansal, for his dedication and guidance throughout our project work, along with his help and collaborative efforts on us and our work.

Furthermore, we would like to thank the Administration of the Department of Biotechnology and Bioinformatics for all of their technical expertise and support.

We also appreciate all of the support our parents, friends and teachers have provided us throughout the curriculum, we could not have done this without the help and support of all these people.

Thank you

Summary

The thesis of our project is divided into four major parts. The first part, or Chapter 1 is related to background study of the skin cancers, how and what kind of data was collected with respect to the rare forms of skin cancers, namely Xeroderma pigmentosum, Cockayne Syndrome and Trichothiodystrophy.

The second part or Chapter 2, describes the design, development and implementation perspectives for making a database, i.e., we developed a Skin Cancer Knowledgebase which aims to be a sole information repository for the above mentioned diseases.

Chapter 3, or the third part, is associated with biological/bio-computational analysis of the data we have collected and stored, and represented in our database. Lastly, Chapter 4 serves as conclusion of the complete project.

The main aim of this project was to explore the problems underlying in the study of rarest forms of skin cancer, why these diseases have been considered as the hotspot for the research in skin cancers, and how our database and computational analysis is helpful in resolving this issue.

_____ _____	_____
(Signature of students)	(Signature of Supervisor)
Tanya Singh, Varsha Mehta	Dr. Tiratha Raj Singh
Date: _____	Date: _____

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
ATP	Adenosine Tri-phosphate
NER	Nucleotide Excision Repair
XP	Xeroderma pigmentosum
CS	Cockayne Syndrome
TTD	Trichothiodystrophy
DB	Database
NCBI	National Center for Biotechnology Information
KEGG	Kyoto Encyclopedia of Genes and Genomes
HGNC	HUGO Gene Nomenclature Committee
COSMIC	Catalogue of Somatic Mutations in Cancer
SNP	Single Nucleotide Polymorphism
NCBI-SRA	NCBI-Sequence Read Archive
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
FDA	U.S. Food and Drug Administration
GWAS	Genome-Wide Association Study
NGS	Next-Generation Sequencing
GO (database)	Gene Ontology Database
PPI	Protein-Protein Interactions
TCGA	The Cancer Genome Atlas
CCLE	Cancer Cell Line Encyclopedia
SBML	Systems Biology Markup Language

PonyORM	Pony Object-Relational Mapper (Python ORM)
KB	Knowledge Base
GUI	Graphical User Interface
HTML	Hyper Text Markup Language
CSS	Cascading Style Sheets
PHP	Hypertext Preprocessor
XAMPP	Cross-Platform (X), Apache (A), MariaDB (M), PHP (P) and Perl (P)
DGIdb	Drug Gene Interaction Database
GEO	Gene Expression Omnibus
WB-DEGS	Within and Between Group Comparisons for Differentially Expressed Gene Selection
RMA	Robust Multi-array Average
ANOVA	Analysis of Variance
SAM	Significance Analysis of Microarray
FDR	False Discovery Rate
SLR	Simple Linear Regression

Introduction

DNA damage has emerged as a major culprit in cancer and many age related diseases. This damage to the cells owes to the endogenous stress like translation, or to the exogenous stresses like exposure to ultraviolet radiations, environmental mutagens, chemotherapeutic abducts etc. Hence, to conserve the genomic integrity and stability, certain repair labyrinthine mechanism is required that can perceive the different types of damage and the checkpoints that counteracts this DNA damage. One of the most important and the prime pathway to remove this bulky DNA lesion is the Nucleotide Excision Repair (NER) pathway, which has a sequential workflow ranging from damage recognition to damage removal, in addition with DNA synthesis. Deficiencies in NER repair proteins are also associated with the skin cancer-prone inherited disorder: Xeroderma pigmentosum (XP) and other neurodegenerative abnormalities like Cockayne Syndrome (CS) and Trichothiodystrophy (TTD).

In this project, we discuss this DNA damage w.r.t. cancer repair mechanism in detail as well its biological relevance by description of the biochemistry of the NER process, clinical features of the NER disorders, therapeutic treatments or drugs, followed by genetic and molecular investigation via designing a web-accessible bio-computational knowledge base, namely, SkinCancerDB, hence speculating the molecular basis underlying these pleiotropic syndrome.

1.1 Problem Statement

DNA damage is the major cause of various skin cancers, but the focus of the SkinCancerDB will primarily be: Xeroderma pigmentosum (XP), Cockayne Syndrome (CS) and Trichothiodystrophy (TTD). SkinCancerDB is a unique, exclusive and curated database of these above mentioned skin diseases and disorders. The main focus of our database is the NER pathway which eliminates these bulky DNA lesions, thereby conserving the genomic integrity.

To facilitate the development and discovery of new diagnostic and prognostic therapies, and for the characterization of these cancers, it is important to make complete use of scattered data on the above cancers available through publications, experiments, technical reports, clinical reports, databases etc. However, a comprehensive resource for cataloguing the computational characterization of NER pathway of the above mentioned skin cancers has not been developed.

Keeping in mind all the existing gaps in knowledge we have designed SkinCancerDB so as to concentrate and eliminate these gaps for a information intensive enriched database.

1.2 Objectives

Following are the objectives that we aim to achieve on the successful completion of the project -

- To provide one-stop information repository related to NER specific skin cancers, namely, XP, CS, TTD.
- To develop one of its kind NER Specific Skin Cancer database covering not only the Literature, Gene Expression, Molecular and Interaction data but also the Clinical Information like: Manifestations and Epidemiology, Drugs and Therapeutics, Clinical Trials etc., along with the in-depth analysis of the collected information.
- Creating a database that will continue to grow, add the advancements and up-gradation in biological, medical and technological terms of the aforesaid skin cancers.
- Computational add-ons through gene expression data analysis for these diseases to comprehend the information for scientific usage.

1.3 Expected Outcomes

- The foremost outcome of SkinCancerDB is easy storage, retrieval, mining and analysis of skin cancer genomic data in a resourceful and structured manner that will be applicable in various biomedical investigations.
- Further this collected data can be used to exploit many biological insights, which can serve as basis for discovery of novel therapeutic strategies or drugs, derived via understanding the functional patterns of such genomic modifications, molecular structural variations etc.
- Meaningful biological and molecular analysis of the collected data along with data visualization techniques.
- All this ever mounting data requires improved and up-to date information technology infrastructure, not only for data storage but also for the development of new computational tools, so as to transform the data suitable for meaningful analysis which further can be employed in large-scale genome characterization projects. This data

classification, characterization and modern computing technologies –are all expected to be encompassed in SkinCancerDB.

1.4 Development Strategy

The project follows a sequential approach from gathering to understanding the respective cancers via tackling the existing current gaps in the research area - mainly the unavailability of NER specific skin disorder database, website or experiments. This includes general background research of the disorders, data collection of bio-computationally relevant data, filtering the collected data, characterization and summarization of the data, followed by designing a web accessible knowledgebase SkinCancerDB and finally the biological/bio-computational analysis.

All the necessary tasks required for the completion as well as execution of the project have been categorized into three major phases:

- Phase I-Data Collection and Integration
- Phase II-Database Proposition and Plan
- Phase III- Biological/Bio-computational Data Analysis

This thesis is divided into four major parts. Chapter 1 is related to background study of the skin cancers followed by design and development perspectives in Chapter 2, whereas Chapter 3 is associated with biological/bio-computational analysis. Lastly, Chapter 4 serves as conclusion of the complete project. The detailed explanation of the above mentioned phases is discussed in later chapters separately, both descriptively and diagrammatically.

2. Chapter-1

Phase-I: Data Collection and Integration

2.1 Overview

This chapter encompasses the Phase I of the project, namely: Phase I - Data Collection and Integration.

During this phase, we collected all the relevant literature, linked them to our project problem and determined the best course of action, i.e., to go for expression analysis and to design a database. We carefully extracted and curated the data on the aforesaid cancers respectively. Finally, after the initial completion of this phase we devised a methodology, a Work Flow Structure, that explains the sequential procedure of the development of SkinCancerDB. Both the steps, data collection and methodology, are explained in detail below.

2.1.1 Data Collection

For the extensive understanding of the mechanism of NER and its relation with carcinogenesis, we applied a wide variety of *in-silico* approaches via most cited computational tools, software and databases on different forms of raw data. A myriad of sequential steps were employed for characterization and analysis on the collected data for respective skin cancers. To understand the methodology of characterization, first, let's discuss the diverse categories of data collected by us (Fig.1).

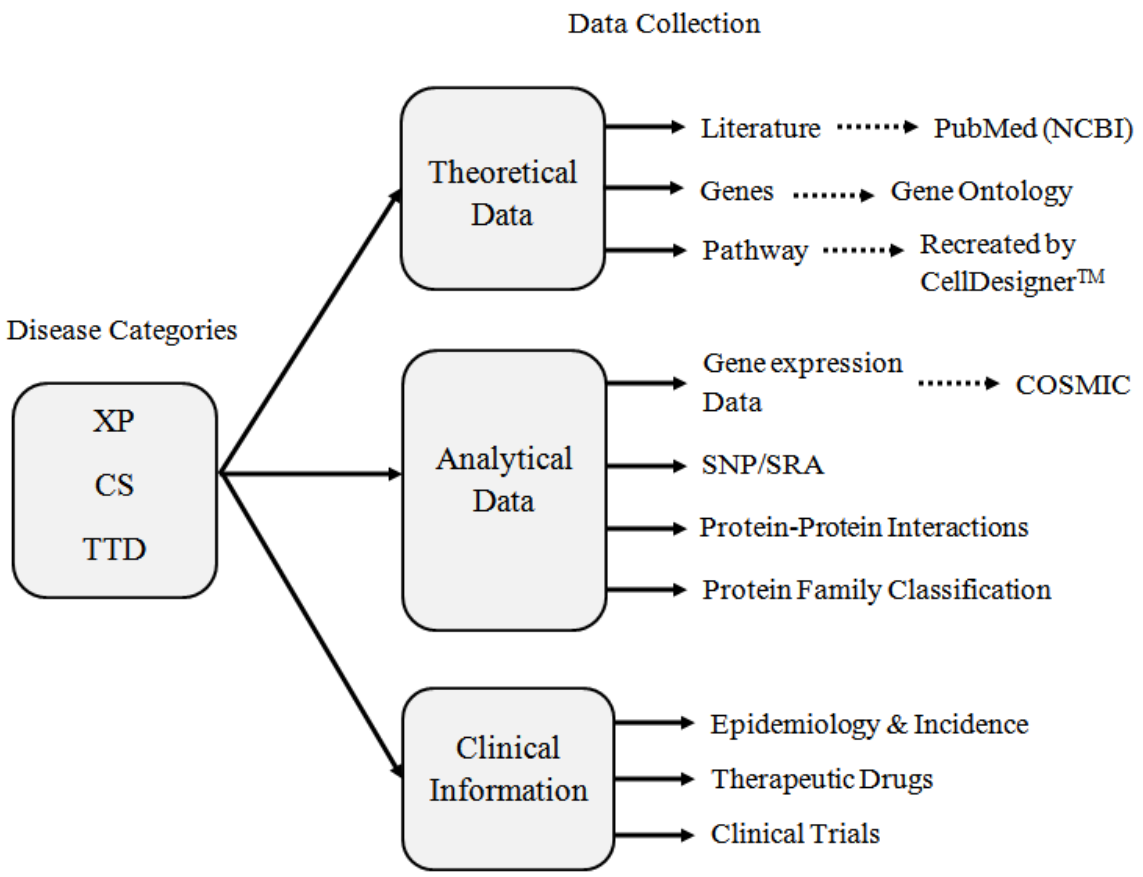


Fig.1: Figure showing the diagrammatic representation of data collected in detail, along with their verified resources.

2.1.2 Biological Data

Primarily, most of the data was collected from literature, regarding the background science of XP, CS and TTD skin cancers respectively, including the analysis of the case studies for these three diseases, and role of the DNA damage and repair mechanism or the Nucleotide Excision Repair (NER). The network motif or pathway of NER mechanism was studied and analyzed through KEGG and REACTOME pathway database, and considering that as the basis, the primary or crucial genes responsible in regulating the pathway were identified, listed and were functionally verified though gene ontology. The information about the potential genes was

collected through databases like GeneCards, HGNC (HUGO Gene Nomenclature Committee), and NCBI.

2.1.3 Evaluating Gene Expression

The main purpose for studying gene expression at an early stage was to identify biomarkers for early detection of disease which consecutively can further be aptly managed.

- To understand the expression level of genes in skin cancer, we analyzed COSMIC (Catalogue of Somatic Mutations in Cancer) database, where information about the mutations, samples and gene expression of each gene for respective skin cancers was given, along with different over-expression and under-expression levels of genes in different tissues of the body.
- Also we studied sequence variations for the respective cancer genes via NCBI dbSNP for in depth understanding of human variations and molecular genetics so as to further employ these principles in gene mapping, definition of population structure and evaluation of performance of functional studies associated with the respective skin cancers.
- Lastly, we searched from the NCBI SRA database. This database aims to promote the reproducibility and new discoveries in the bioinformatics domain by making the biological sequences available to the research community for comparisons of the data set. Raw sequence data information, sequence alignments along with high throughput sequencing platforms (Roche 454, Illumina Genome Analyzer®) of the associated cancer genes were searched and stored.

2.1.4 Protein Family Classification and Functional Annotation

Protein classification provides significant evidences about the structural properties, functional activities and metabolic tasks of the molecule, thereby serving an important basis for large scale genomic annotation. Hence, for such classification of the genes of the respective skin cancers, we used InterPro, a protein database for protein sequence analysis and classification; where while searching for the genes, we came to know about the family-based propagation of the genes

or proteins, domain familiarity, the related analogous proteins, the respective domain architecture and design, pathways & interactions, genus taxonomy and cross-references. This family classification and analysis formulates the basis of the essentials for the comparative genomics and phylogenetics.

2.1.5 Gene-Protein Interactions and Association

Understanding how proteins interact with DNA, determining what proteins are present in these protein-DNA complexes and identifying the nucleic acid sequence (and possible structure) required to assemble these complexes, are vital in understanding the role of these complexes in regulating various cellular processes. To analyze the functional association of these genes involved in skin cancers and their protein-protein interactions, STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database was used, which displayed the interaction network of the proteins known through experimental evidences, curated databases, or predicted through gene neighborhood, fusions, co-occurrence, co-expression or protein homology; along with the properties of each interacting protein, for each respective gene.

2.1.6 Clinical Information

- **Manifestation and Epidemiology**

  Clinical manifestation refers to the disease presentation phenomenon in patients in the form of signs and symptoms that are recognizable through a variety of means - visual and auditory examination, or medical testing. Whereas epidemiology is the study and analysis of disease pattern, its causative agents, its effects on health in one/many populations. The combination of these manifestation and epidemiology studies not only helps in diagnosis of these cancers but also help in managing their prevention.

- **Drugs and Therapeutics for Skin Cancer**

  Drugs or other chemical substances are the widely discovered, manufactured and used targeted therapeutics designed to wedge the development, intensification and amplification of cancer. They prevent the division of cancer cells or directly destroy them. To understand this drug targeted therapy for the prognosis as well as diagnosis of the respective skin cancers, we searched through FDA (U.S. Food and Drug

Administration), Medscape repositories and DrugBank database, from where 11 drugs approved by FDA and 1 drug undergoing phase-II trials from DrugBank were discovered, along with how each drug is helpful in treating the cancer at metastatic or advanced stage.

- **Clinical Trials**

  We searched through clinicaltrials.gov database for XP, CS and TTD clinical studies that are conducted around the world of human participants. These studies enrich the scientific research of these cancers by finding better therapeutic strategies, more potent drugs, better screening and diagnostic treatments.

So as to summarize our results in a singular table, following amount of data has been collected:

Table 1: Table showing total amount of data collected so far.

Data Type	***Results (no.)***
Literature	80 citations
Target Genes	12 genes (XP,CS,TTD)
Gene Expression Data	12 (for all genes)
Protein Family Classification	12 (for all genes)
STRING Interactions	10 Interactions (x 12, for each gene)
Therapeutic Drugs	12 (XP,CS,TTD)
SNP Analysis Data	43 (x 12, for each gene)
Clinical Trials	3 trials each for XP and CS, 1 trial combined for XP/CS/TTD

2.2 Methodology

This section of phase-I focuses on and takes into account the systematic approach required for the characterization, analysis and summarization of the collected data. The procedural work flow describes the sequential series of the projected steps along with the theoretical analysis of the different methods applied on to the diversified collection of data. Typically, it encompasses

concepts such as paradigm, theoretical model, phases and quantitative or qualitative techniques used to understand the principles associated with our branch of knowledge i.e. Skin Cancer.

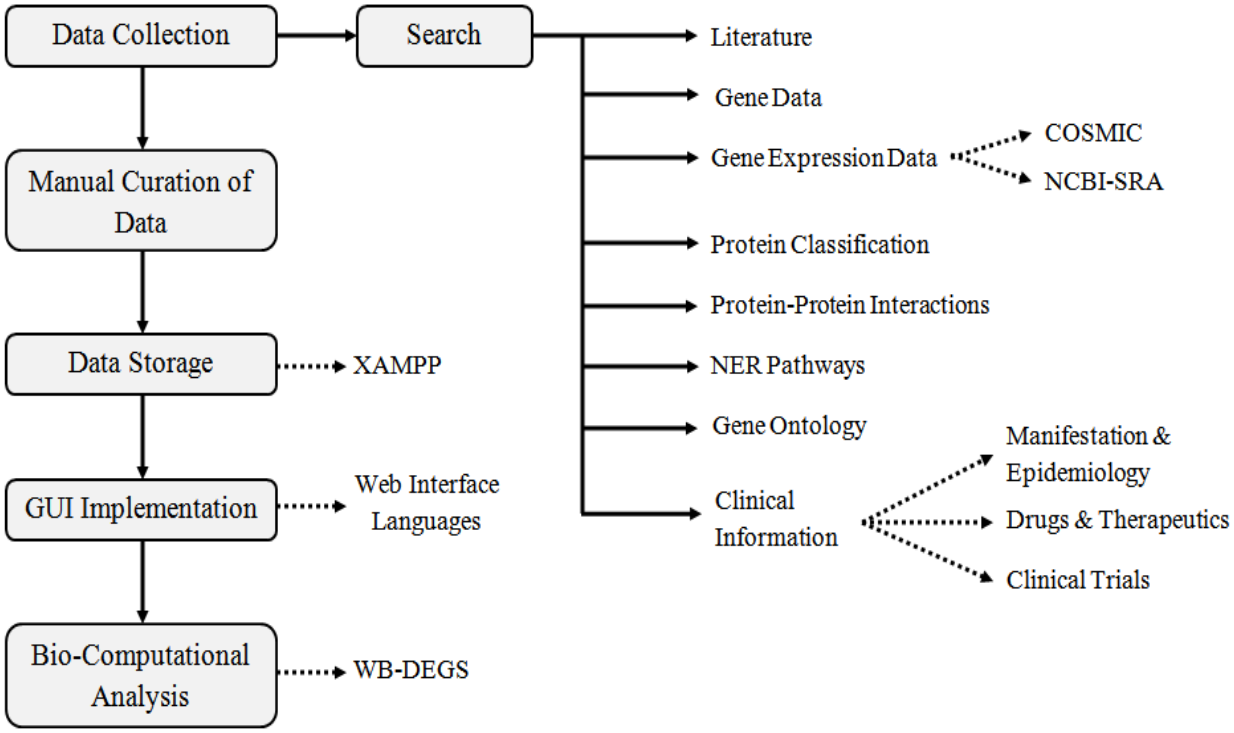


Fig.2: Procedural work flow for the different methods applied on the diversified collection of data.

2.3 Results

This section of the chapter primarily represents the results that were obtained after the successful completion of data collection, integration and methodology devising steps. The results for each category of data collected have been explained in detail and presented in tabular format below.

2.3.1 Literature

So far we have cited and collected 80 articles from NCBI related to Xeroderma pigmentosum, Cockayne Syndrome and TTD, and the literature references along with their PubMed ID, Authors and Description and have stored the records in a tabular format (Table 2).

Table 2: Sample table for literature articles related to DNA repair and skin disorders.

DNA Repair and Skin Disorders			
PMID	**Title**	**Author(s)**	**Description**
2408009	Complementation of the xeroderma pigmentosum DNA repair synthesis defect with Escherichia coli UvrABC proteins in a cell-free system.	Hansson J, Grossman L, Lindahl T, Wood RD	A newly developed cell-free system was used to study DNA repair synthesis carried out by extracts from human cell lines in vitro.
25207368	Ultraviolet damage, DNA repair and vitamin D in nonmelanoma skin cancer and in malignant melanoma: an update.	Reichrath J, Rass K.	In animal models, it has been demonstrated that UVB is more effective to induce skin cancer than UVA.
. . . . . .	. . . . . .	. . . . . .	. . . . . .
24918982	Growth and nutrition in children with trichothiodystrophy.	Atkinson EC, Thiara D, Tamura D, DiGiovanna JJ, Kraemer KH, Hadigan C.	Here, we show that TTD primary dermal fibroblasts contain low amounts of collagen type VI alpha1 subunit (COL6A1), a fundamental component of soft connective tissues.

2.3.2 Gene Data

By analyzing NER pathway from GeneCards, KEGG, HGNC and NCBI, we found around 12 genes (XPA, XPB, XPC, XPD, XPE, XPF, XPG, CSA, CSB, ERCC1, RPA, TTDA) involved in the pathway and their information is stored in tabular format (Table 3). Discovering the core genes that have high disease predisposition in humans has always been a major chore in biomedical research area. Hence, our prime focus has been in understanding the functionality of these genes. Through our Gene Ontology analysis we validated the contribution of these disease genes in the development of the respective skin diseases. Collective and comprehensive understanding of the genetic disposition of complex diseases can serve as basis for development of novel drug targets and may also help in discovery of new treatment strategies. Innovative large-scale genome analytic

technologies like genome-wide association study (GWAS) and next generation sequencing (NGS), allows for intensive hereditary and bio-medical study of human diseases.

Table 3: Sample table showing target genes that are crucial for NER-DNA repair mechanism.

Target Genes Involved in DNA Repair NER Pathway								
Symbol	**Name**	**Function**	**Location**	**Ensembl ID**	**OMIM ID**	**UniProtKB ID**	**HGNC ID**	**MGI ID**
XPA	Xeroderma Pigmentosum Group A-Complementing Protein	Involved in DNA excision repair. Initiates repair by binding to damaged sites with various affinities.	9q22.33	ENSG00000136936	611153	P23025	12814	99135
XPB	Xeroderma Pigmentosum, Complementation Group B	ATP-dependent 3-5 DNA helicase, involved in nucleotide excision repair (NER).	2q14.3	ENSG00000163161	133510	P19447	3435	95414
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
TTDA	Trichothiodystrophy Group A	This gene encodes a subunit of transcription/repair factor TFIIH, which functions in gene transcription and DNA repair.	6q25.3	ENSG00000272047	608780	Q6ZYL4	21157	107227

2.3.3 Gene Ontology

The Gene Ontologies of the cancer associated target genes are important for the functional analysis of the genes and their gene product. We searched the GO database for each gene, modified the search manually by extracting only those functions that were associated with NER pathway. Following table represents the extracted ontological information:

Table 4: Sample table showing Gene Ontologies for all the 12 genes.

Gene Ontologies for Target Genes			
Gene Name	**GO Id(s)**	**Function Name**	**Definition**
XPA	GO:0000715	NER, DNA damage recognition	The identification of lesions in DNA, such as pyrimidine-dimers, intrastrand cross-links, and bulky adducts. The wide range of substrate specificity suggests the repair complex recognizes distortions in the DNA helix.
	GO:0000717	NER, DNA duplex unwinding	The unwinding, or local denaturation, of the DNA duplex to create a bubble around the site of the DNA damage
	GO:0006281	DNA repair	The process of restoring DNA after damage. Genomes are subject to damage by chemical and physical agents in the environment (e.g. UV and ionizing radiations, chemical mutagens, fungal and bacterial toxins, etc.) and by free radicals or alkylating agents endogenously generated in metabolism
. . . . .	. . . . .	. . . . .	. . . . .
TTDA	GO:0070911	GG-NER	The nucleotide-excision repair process in which DNA lesions are removed from non transcribed strands and from transcriptionally silent regions over the entire genome.

2.3.4 Re-Construction of the NER pathway

After studying about the NER pathway and its role as a DNA repair mechanism in Xeroderma pigmentosum, and identifying the two types of NER as GG-NER (Global Genome-Nucleotide Excision Repair) and TC-NER (Transcription Coupled-Nucleotide Excision Repair), which only differ during the initial DNA damage recognition, the mechanism was simplified by reconstructing the combined pathway through the use of CellDesigner^{TM}, which is a software of structured diagram editor for drawing gene-regulatory and biochemical networks (Fig.3).

The networks are drawn based on the process diagram, with graphical notation system proposed by Kitano, and are stored using the Systems Biology Mark-up Language (SBML). Further, the diagram helps in identifying the important genes involved in DNA damage recognition and repair system. Purpose of this reconstruction was to provide the latest NER pathway model in a standard form so it can be reused by the scientific community. It will help in the reusability of this model in various available tools, softwares, and web based applications.

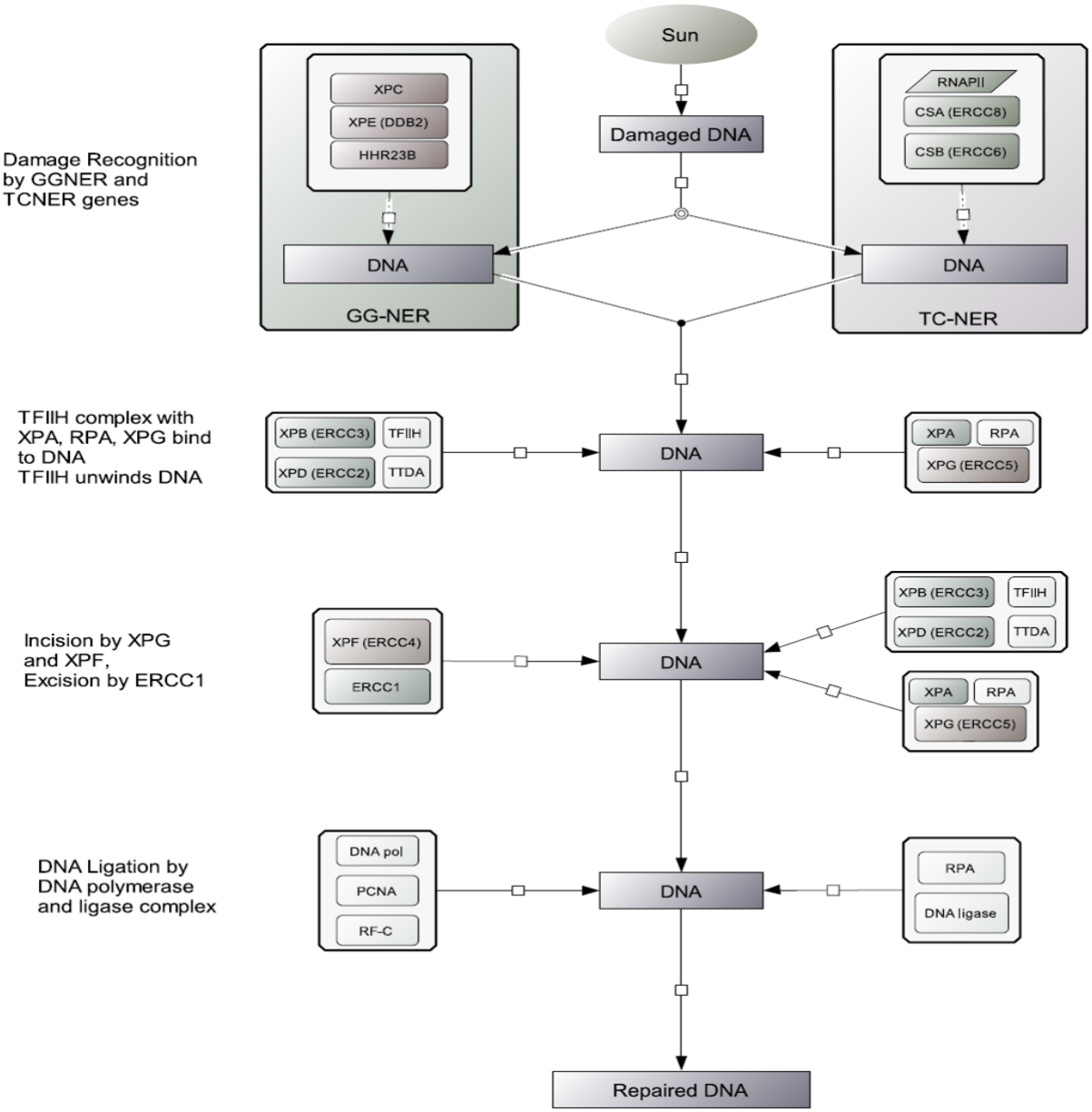


Fig.3: The CellDesigner^{TM} Reconstructed NER Pathway.

2.3.5 Gene Expression Data Analysis

The gene expression data is the product of the raw and unprocessed microarray data images which are converted into gene expression matrices-tables, where rows signify genes, columns imply various samples (tissues or experimental conditions) and values in each cell corresponds to the expression intensity of that particular gene in the given sample. For the extraction and understanding of the biological processes these matrices have to be analyzed. Technology is still struggling with the analysis and management of such data. But nevertheless, there are wide continuums of applications of analysis of gene expression data.

- On our collected COSMIC and NCBI SRA data, we can apply wide variety of supervised and unsupervised data analysis techniques for predicting gene function classes, cancer classification etc.

Table 5: Sample table showing gene expression analysis for NER pathway genes from COSMIC.

COSMIC Data for Gene Expression in Skin Cancer						
Gene_name	**Total samples**	**Overexp_ samples**	**Overexp_ samples (%)**	**Underexp_ samples**	**Underexp_ samples (%)**	**COSMIC gene ID**
XPA	473	9	1.9	12	2.54	COSG904
XPB	473	36	7.61	15	3.17	COSG687
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
TTDA	473	8	1.69	1	0.21	COSG58853

Table 6: Sample table showing SRA data for all 12 genes from NCBI-SRA.

NCBI-SRA Data for Target Genes							
Target gene	**Title**	**Primary ID**	**Study**	**Bioproject _ID**	**Sample species**	**Biosample ID**	**Instrument**
XPA	Whole-exome sequence analysis of XPA-deficient cells and their iPS cells	SRR2037122	Homo sapiens exome	PRJNA284651	Human XPA-deficient fibroblast cells and their iPS cells	SAMN03704960	Illumina HiSeq 1000
XPB (ERCC3)	Illumina HiSeq 2500 paired end sequencing	ERR1408866	Transcriptome_profiling_of_zebrafish_mutants_infected_with_pathogens	PRJEB9480	Danio rerio; ZMP_phenotype_189_1_48	SAMEA3951507	Illumina HiSeq 2500
. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
TTDA	Genome-wide mapping of ssDNA in human trichothiodystrophy (TTD) patient fibroblasts reconstituted with wt XPD at 41C (+ supercoiled pFLIP plasmid)	SRR823530	Genome-wide mapping of ssDNA in human	PRJNA193635	General sample from Homo sapiens	SAMN01998288	Illumina HiSeq 2000
	Genome-wide mapping of ssDNA in human trichothiodystrophy (TTD) patient fibroblasts at 37C (+ relaxed pFLIP plasmid)	SRR823523	Genome-wide mapping of ssDNA in human	PRJNA193635	General sample from Homo sapiens	SAMN01998281	Illumina HiSeq 2000

2.3.6 Protein Family Classification and Annotation

Protein family classification of our genes of respective skin diseases will provide us with several advantages for large-scale genomic annotation. This characterization and analysis of gene families will facilitate: recognition of proteins that are difficult to characterize based on pair-wise alignments ,support database maintenance by encouraging family-based annotation and making annotation errors apparent, provide an efficient model to retrieve significant biological information from gigantic amounts of data and lastly will reveal the associated gene families, which later can be utilized in comparative genomics and phylogenetics.

For the protein families and domains, the results for all 12 genes obtained from InterPro were recorded in Excel worksheet in table format as (Table 7):

Table 7: Sample table showing results for protein families and domains obtained from InterPro.

Families and Domains for Genes in NER Pathway				
Gene Name	**Protein**	**Family/Domain**	**Domain Architectures**	**Protein IDs (Homo sapiens)**
XPA	XPAC	XPA C-terminal	22	P23025
				F2Z2T2
				W0FSR8
XPB	ERCC3	ERCC3/RAD25/XPB helicase, C-terminal domain	98	P19447
				A8K359
				B3KRG2
				B3KTH1
				G3V1S1
				Q53HW5
. . . .	. . . .	. . .	. . .	. . . .
RPA	RPA1	Replication Factor A protein 1/RPA1 N-terminal domain	80	P27694
				I3L2M5
				I3L524
TTDA	GTF2H5	TFIIH subunit TTDA	22	Q6ZYL4

Further we can combine the classification information along with sequence patterns or profiles, to predict functional sequences like active sites, binding sites, and motifs etc using existing biological/bioinformatics approaches. This type characterization as well as prediction can provide major assistance in advancement of novel therapeutics and diagnosis.

2.3.7 Gene-Protein Interactions and Association

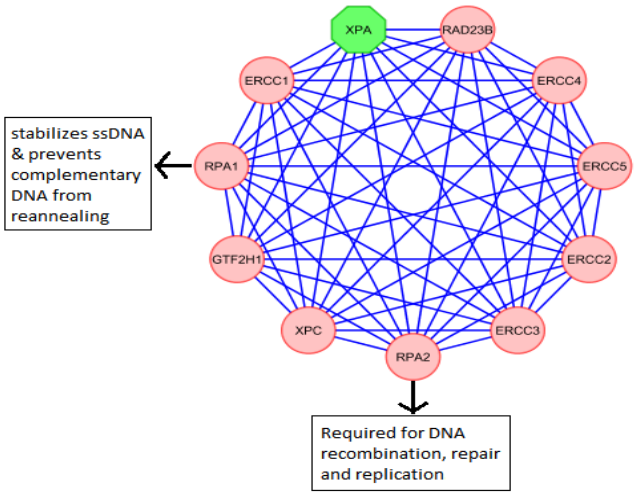
For understanding the roles played by various genes in complex diseases like skin cancer itself, it is important to scrutinize and examine the genes and their associated concerned network. From our study of these interactions of individual genes with other components and previously completed analysis of gene expression data- both of their integration, the resulting network data offers a strong potential approach that can help in prioritizing the disease-associated genes. No significant or accurate methods have been developed in this field and the problem still continues to persist. Hence, we can use this integrated approach for Skin Cancer Specific Prioritization of Genes and further which will not only bridge the differences in the existing current gaps in skin cancer but will also provide targeted, specific and validated results in this area of research.

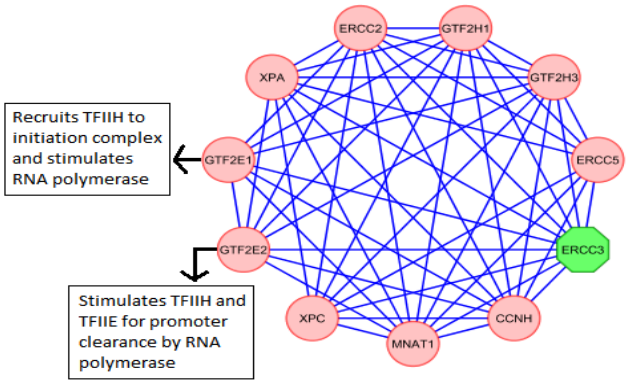
From STRING database, connectivity for the proteins of all 12 genes were discovered and were stored in a tabular format (Table 8):

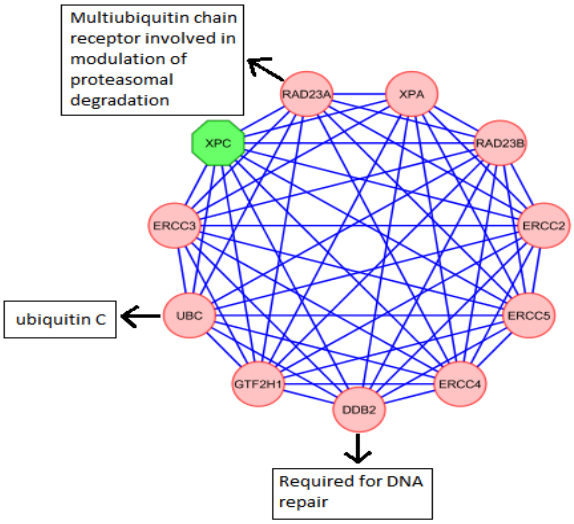
Table 8: Sample table showing genes associated with different proteins and their properties.

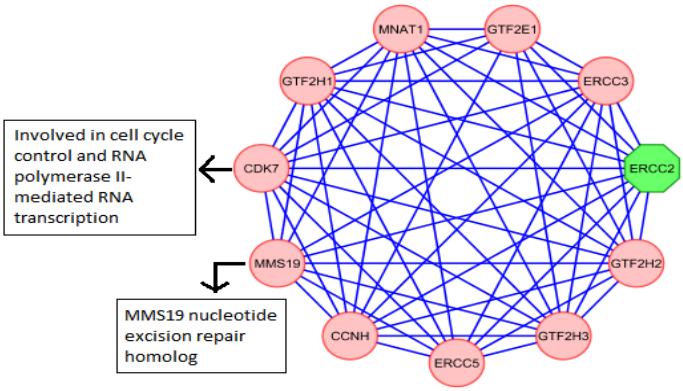
Protein-Protein Interactions for Target Genes from STRING Database				
Gene_name	**Predicted_functional_partners**	**Basis_of_interaction**	**Properties**	**Score**
XPA	RPA1	Experiments, databases, text mining	It plays an important role in DNA repair, cell metabolism etc.	0.999
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

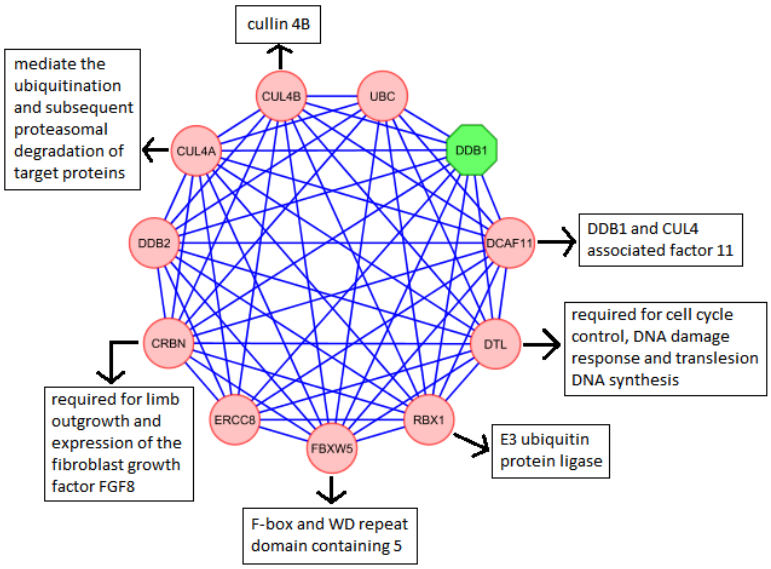
TTDA	GTF2E2	Coexpression, experiments, text mining	It leads TFIIH to initiation process of DNA transcription and regulates function of RNA polymerase.	0.953
	ERCC3	Experiments, text mining	It helps in regulating function of TFIIH complex during DNA repair and other metabolisms of DNA.	0.977

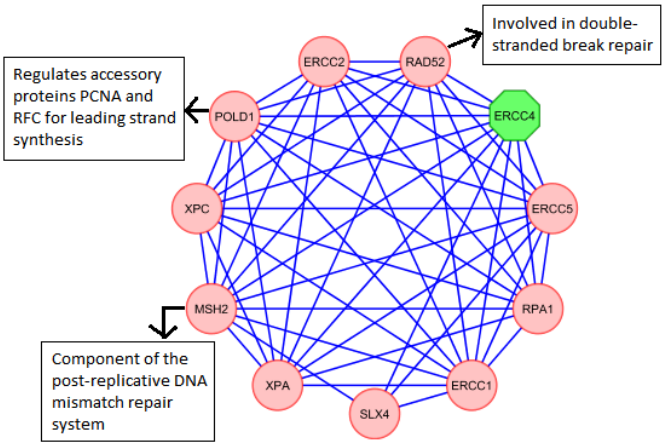


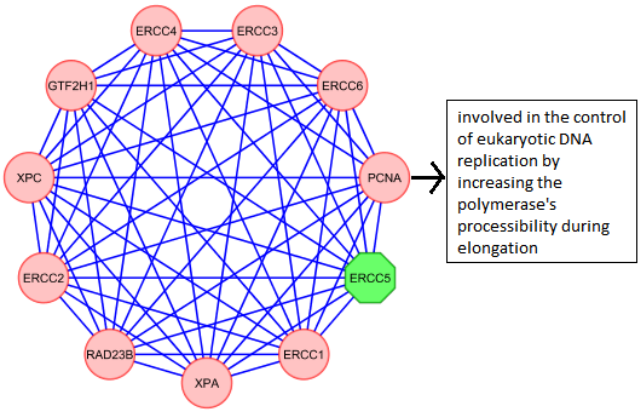


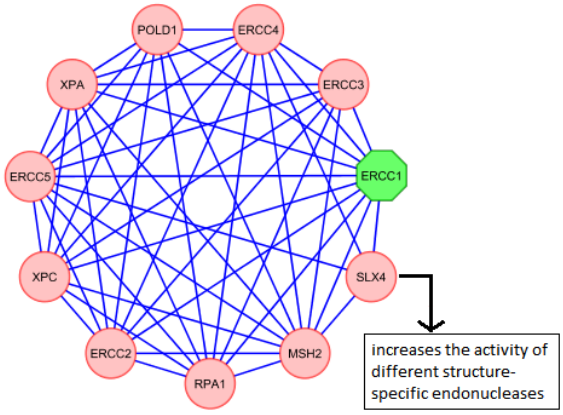





cullin 4B
mediate the ubiquitination and subsequent proteasomal degradation of target proteins
CUL4B
UBC
CUL4A
DDB1
DDB2
DCAF11
DDB1 and CUL4 associated factor 11
CRBN
DTL
required for cell cycle control, DNA damage response and translesion DNA synthesis
required for limb outgrowth and expression of the fibroblast growth factor FGF8
ERCC8
RBX1
FBXW5
E3 ubiquitin protein ligase
F-box and WD repeat domain containing 5


Involved in double-stranded break repair
Regulates accessory proteins PCNA and RFC for leading strand synthesis
ERCC2
RAD52
POLD1
ERCC4
XPC
ERCC5
MSH2
RPA1
Component of the post-replicative DNA mismatch repair system
XPA
ERCC1
SLX4


ERCC4
ERCC3
GTF2H1
ERCC6
XPC
PCNA
involved in the control of eukaryotic DNA replication by increasing the polymerase's processibility during elongation
ERCC2
ERCC5
RAD23B
ERCC1
XPA


POLD1
ERCC4
XPA
ERCC3
ERCC5
ERCC1
XPC
SLX4
increases the activity of different structure-specific endonucleases
ERCC2
MSH2
RPA1

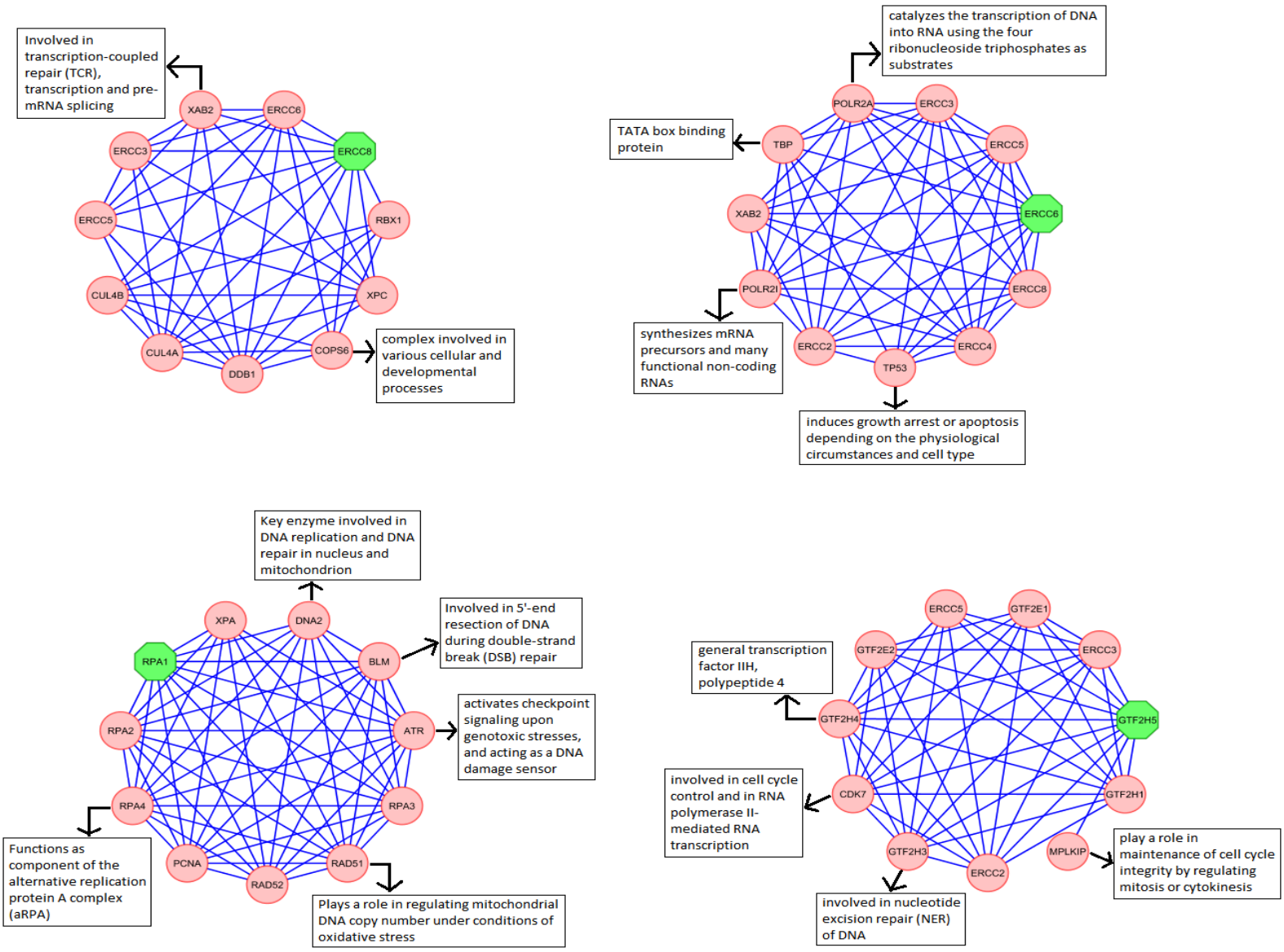


Fig.4: Figures showing common proteins involved in PPIs of all 12 genes by STRING database.

Table 9: Table showing common proteins involved in PPIs of all 12 genes by STRING database.

Common proteins	**Properties**
XPA	It regulates DNA repair mechanism.
XPC	It helps in sensing damage due to UV irradiation and binds as a XPC-binding factor to DNA during GG-NER.
ERCC1	It helps in 5-prime incision of damaged DNA during DNA repair process.
ERCC2	It regulates Vit-D receptor activity during sun exposure of skin.
ERCC3	It is involved in NER mechanism of DNA repair.
ERCC4	Like ERCC1, it also helps in 5-prime incision during DNA repair process.
ERCC5	It helps in making 3-prime incision on DNA during DNA repair.
ERCC6	This protein is involved primarily in TC-NER process of NER in DNA repair.
ERCC8	It is involved in TC-NER process of DNA repair mechanism.
GTF2H1	It is involved in Nucleotide Excision Repair (NER) of damaged DNA.
GTF2H2	It is generally involved in NER process.
GTF2H5	It is involved in Nucleotide Excision Repair (NER) of damaged DNA.
GTF2E1	It leads TFIIH to initiation process and regulates function of RNA polymerase complex during DNA transcription.
GTF2E2	It is involved in DNA transcription process.
MNAT1	It helps in stabilizing the H-CDK7 complex of cyclin, forming a functional CDK - activating kinase complex.
CCNH	It regulates CDK7 which is the subunit of the CDK - activating kinase complex.
DDB1	It binds to DNA damaged-binding protein 2 to form UV-DDB complex.
DDB2	It binds to DDB1 forming UV-DDB complex, essential in NER process.
RPA1	It has a major role in metabolism of DNA like recombination, DNA repair etc.
RPA2	Like RPA1, it is required for repair, replication & recombination of DNA.
RAD23B	This protein helps in modulating degradation of proteosome in multi-ubiquitin.
RBX1	The variability of substrate recognizing components help this protein in its functional specificity.
RAD52	This protein is involved in DNA double-strand break repair.

2.3.8 Drugs and Therapeutics

From our investigation and study in the treatment, therapeutics and drugs for the respective skin cancers it has come to our notice that no noteworthy progress till date has occurred in the drug development field. Majority of drugs undergoing phase-II clinical trial have failed due to their lower rates of efficacy, specificity and potency. Not only this but from the available drugs of the skin cancers no drugs have been specifically reported to cure any one of – *Xeroderma pigmentosum*, *Cockayne Syndrome* or *TTD*. This concentrates on a very severe subject of the lack of knowledge and experimentation in above mentioned skin cancers.

The information about drugs used for treatment of skin cancer was obtained from FDA website, DrugBank and Medscape, giving us around 12 drugs and medications, and the results were stored in tabular format (Table 10):

Table 10: Sample table showing list of drugs currently used for treating skin cancer.

List of Drugs for Treatment of Skin Cancer Malignancies						
Drug Name	**Active Ingredient**	**Disease Treatment**	**Current Status**	**Properties**	**Side Effects**	**Manufacturers**
Dacarbazine	Decarbazine	malignant melanoma	Approved / 15-06-2001	It has significant activity against melanomas.	Birth defects to children conceived or carried during treatment; sterility, possibly permanent, or immune suppression	Bayer Healthcare Pharmaceuticals, Inc., Whippany, NJ, USA
Imiquimod	Imiquimod	BCC, Xeroderma pigmentosa - undergoing investigation	Approved / 24-06-2014	treatment of biopsy confirmed primary superficial basal cell carcinoma.	erythema, edema, scabbing or crusting, flaking or scaling, erosion, weeping	Taro Pharmaceuticals, Inc., Hawthorne, NY, USA

. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
Imexon	cyanoaziridine	melanoma, multiple myeloma, ovarian cancer.	Undergoing Investigation (phase-II study)	Imexon is a thiol-binding small molecule which induces mitochondrial oxidation, leading to apoptosis.	NA	NA

2.3.9 Clinical Trials

We searched through clinicaltrials.gov database for XP, CS and TTD clinical studies that are conducted around the world of human participants. Three clinical trials were conducted for XP, three for CS and one for TTD. These studies enrich the scientific research of these cancers by finding better therapeutic strategies, more potent drugs, better screening and diagnostic treatments.

Following table categorizes the trials on the basis of cancer type and further includes the additional information associated with the trial.

Table 11: Sample table showing list of clinical trials carried out for XP, CS, TTD.

Clinical Trials information for XP, CS, TTD							
Disease Type	**Trial Name**	**Clinicaltrials .gov identifier**	**Start Date**	**Last Update Date**	**Description**	**Study Type**	**Current Status**
Xeroderma pigmentosum	Cancer Risk in Carriers of the Gene for Xeroderma Pigmentosum	NCT00046189	18-Sep-02	24-Jan-17	This study will help in determining whether family members of patients with XP have various abnormalities or not .	Observational	Recruiting
Cockayne Syndrome	Pharmacokinetics and Safety Study of Single and Multiple Oral Doses Prodarsan™ in Patients With Cockayne Syndrome	NCT01142154	Jun-10	22-Jun-11	This study is to compare the exposure of Prodarsan™ to the intravenous dosed Osmitrol in Cockayne Syndrome (CS) patients.	Interventional	Completed
. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
XP/CS/TTD	Examination of Clinical and Laboratory Abnormalities in Patients With Defective DNA Repair: XP, CS, or TTD	NCT00001813	26-Apr-99	24-Jan-17	This study aims to perform vigilant clinical examination of selected patients with XP, XP/CS, CS, or TTD and follow their clinical course.	Observational	Recruiting

2.4 Conclusion

There is an exponential increase of cancer related data from widely dispersed resources, like scientific publications, Genome Wide Association Studies, Gene Expression Experiments, or Protein-Protein Interaction Data, Epigenomics, Immunogenomics, stored in relevant repositories like Array Express, Cancer Cell Line Encyclopedia—CCLE, The Cancer Genome Atlas - TCGA etc. All this ever mounting data requires improved and up-to date information technology infrastructure, not only for data storage but also the development of new computational tools so to transform the data suitable for meaningful analysis, which further can be employed in large-scale genome characterization projects.

Hence due to limited amount of availability of analytical data (mainly expression data) our initial plan for execution of the expression analysis could not be carried out. Further we decided to build a one-stop information repository of the collected data by careful selection, screening, cleansing and transformation of this data.

3. Chapter-2

This chapter explains the Phase II of the development strategy, namely - **Phase-II: Database Design and Construction.**

3.1 Database Design and Construction

To facilitate the computational characterization and summarization of these cancers, it is essential to completely understand the available data through publications, biological reports, databases etc. This data may contrast largely in granularity, quality and complexity due to the varying data collection standards. Hence to understand the database construction process in detail we use KB-workflow scaffold as the reference image. The KB-workflow scaffold consists of series of sequential steps, each explaining the functional modules of the construction process. Data scattered across the primary databases, scientific literatures and clinical experiments acts as the input to the KB-workflow scaffold. From Fig.6, the first step allows for the automated data collection and integration, followed by data cleaning process, assigning annotations, enabling data storage and retrieval followed by bio computational analysis of data and finally the definitions of the keywords to facilitate biological search in order to understand the biochemistry of NER pathway and related skin cancers.

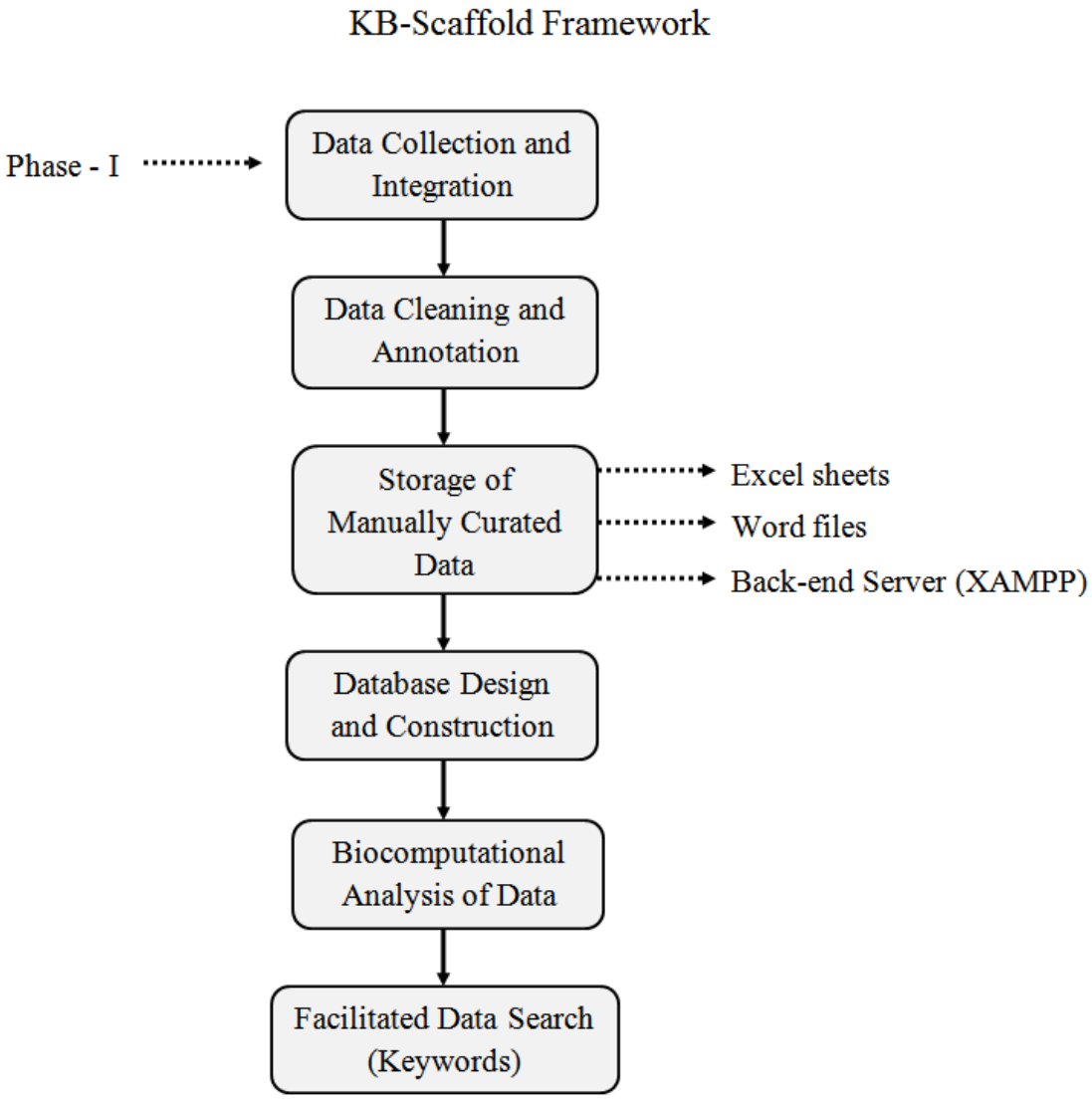


Fig.6: Figure showing the KB-scaffold framework for database construction.

3.2 Technical Specifications

Now taking phase II in account, the steps for Database Design and Construction are represented in Fig.7 and explained below.

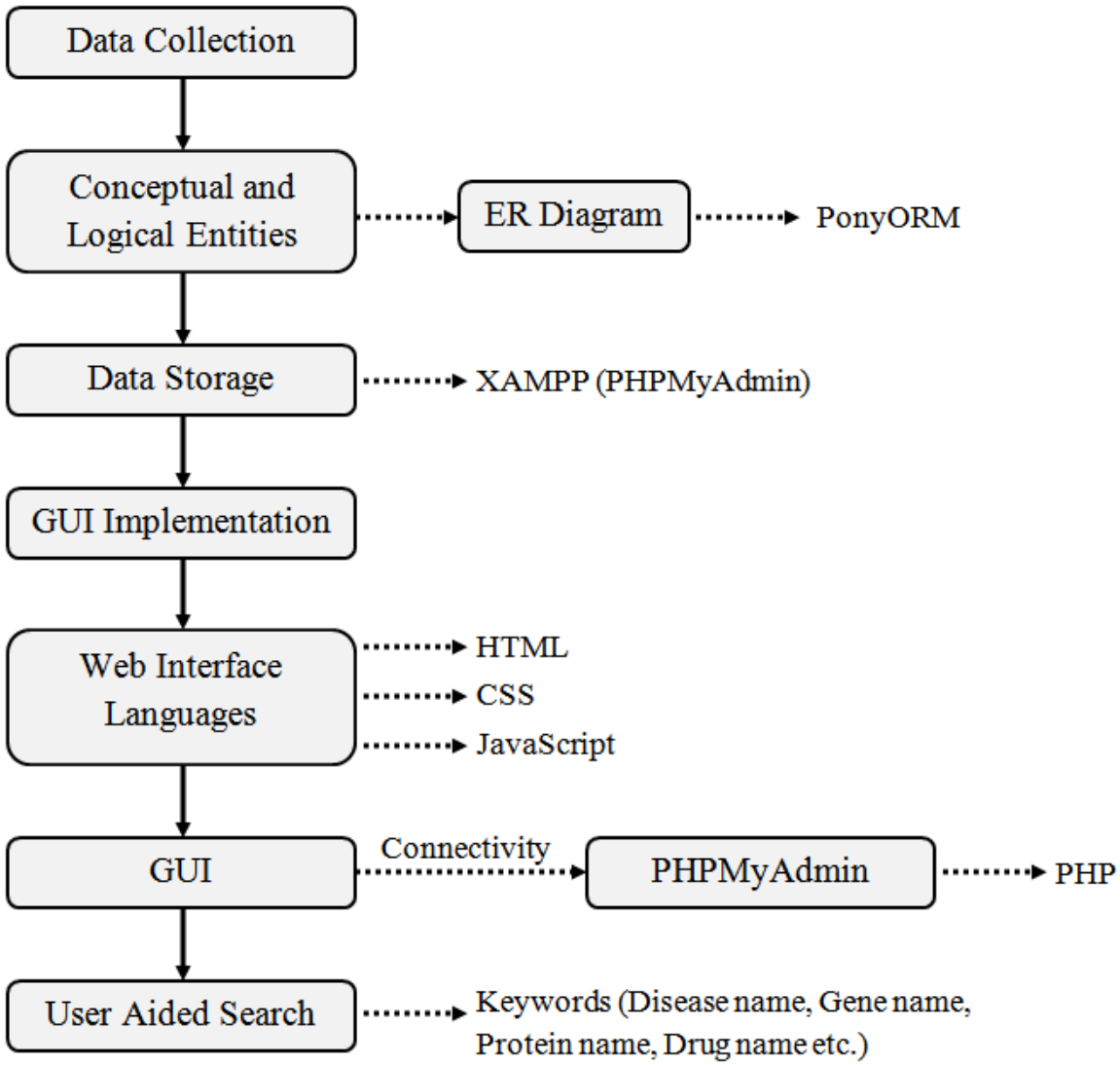


Fig.7: Figure showing steps and procedure for database design and construction.

This figure clearly explains all the steps and tools used for construction of database. The first step is the identification of major entities from the curated data required for the conceptual and logical schema of database, creating E-R diagram, storing the data of these entities in tabular format in a back-end server (XAMPP), then a GUI template is selected and modified using combinations of scripting languages - HTML5, CSS 3.0 and JavaScript, finally we connect the back-end data with our GUI using PHP. Now this web-accessible bio-computational knowledge base is ready for biological analysis of the cancers.

3.3 Results

This section of the chapter focuses on the various user centered search driven features offered by the constructed database – SkinCancerDB.

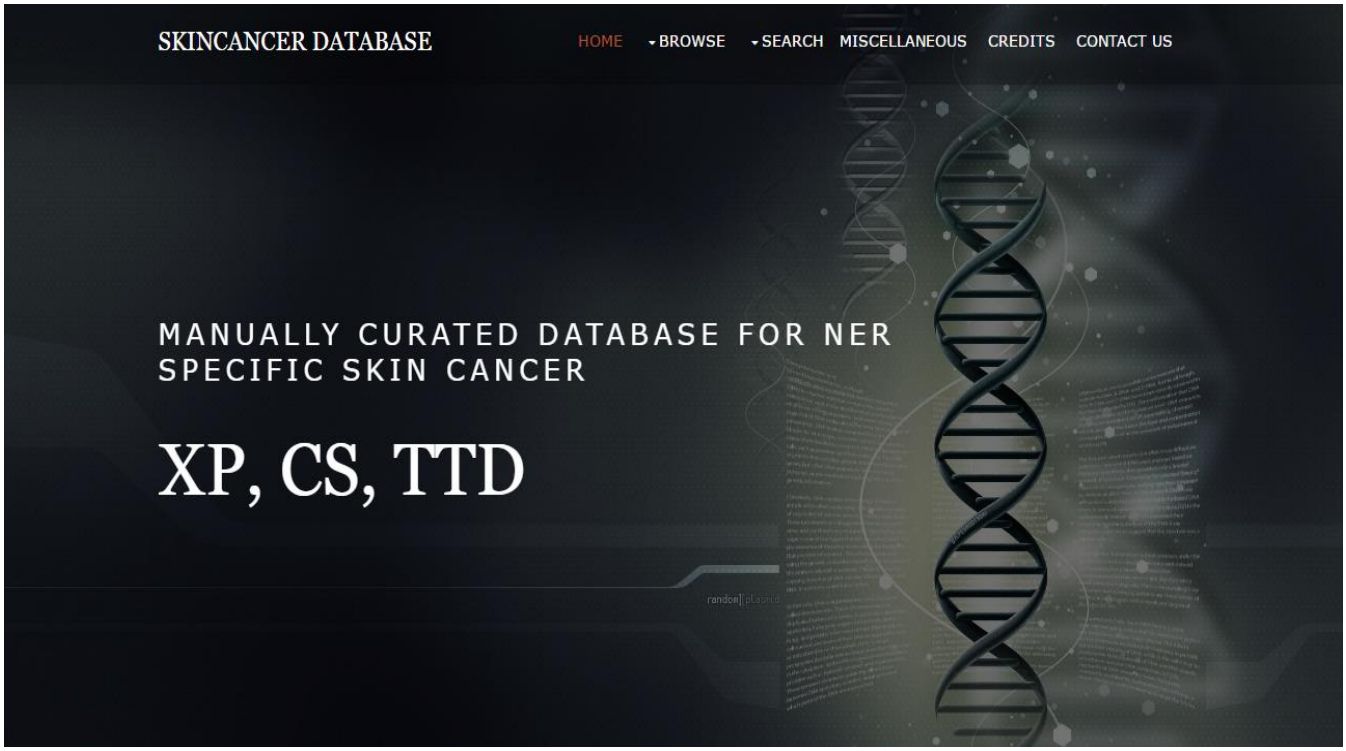


Fig.8: Homepage of SkinCancer Database.

Here, we present SkinCancerDB which is a unified and exclusive resource to understand the biochemistry of NER process and related skin cancers via bio-computational analysis of data collected and further offer characterization and summarization of data via database implementation in the form of web-accessible knowledge base.

3.3.1 Data classification

In SkinCancer DB, the classification was done primarily via cancer types. Genes were classified into three groups, depending on the diseases they were associated with; expression level was categorized based on high and low clinical risk for cancer, drugs for specific cancers were curated, clinical trials were combined and further search features were modified for the user.

The features of SkinCancerDB along with the functional informational content held by them is represented by the image below:

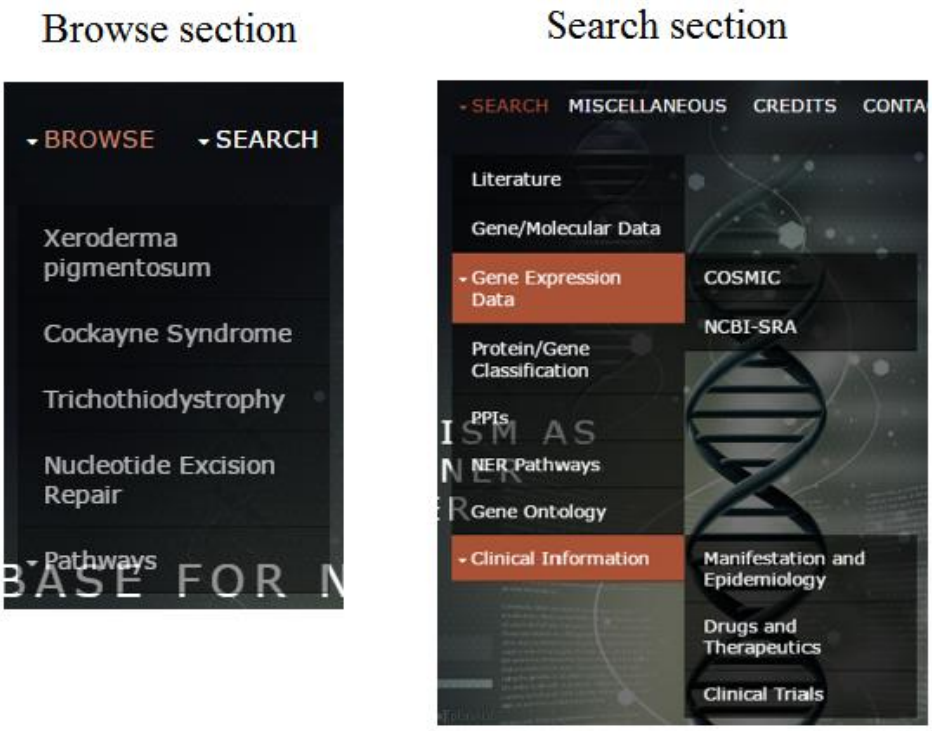


Fig.9: Image showing the expansion of Browse and Search sections in SkinCancerDB.

The database holds - disease causing genes extracted from different experimental studies and journals along with their annotation, tissue wide and protein expression levels, evolutionary affiliations, protein–protein interactions, clinical information (available drugs, therapeutic strategies, clinical trials), NER pathway mechanism and its components with the respective functions, biochemistry and genetics of the related skin cancers.

Detailed working mechanism of these features is explained in Chapter-3.

4. Chapter-3

This chapter consists of the functional working of every feature of SkinCancerDB , hence focusing on the phase III development strategy namely, Biological/Bio-Computational Data Analysis.

Phase-III: Biological/Bio-Computational Data Analysis

4.1 Data search

Firstly, by going on Browse Page, the user gets comprehensive biological description of XP,CS and TTD. Next in the line is Search page. SkinCancerDB supports multi keyword searching. The user can simply query the database using - Cancer names/types, Gene names/types, Drug names etc.

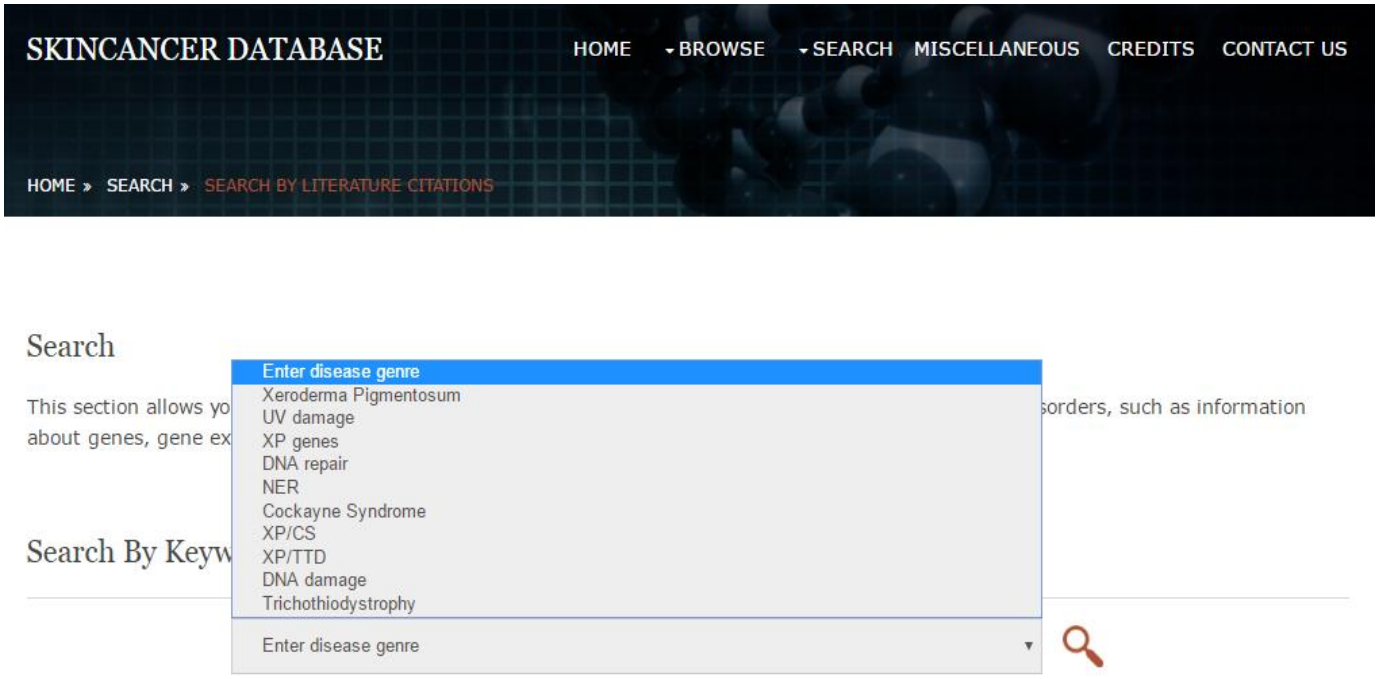


Fig.10: Example of a facilitated keyword-driven search option for Literature Citations in SkinCancerDB.

In addition to this, Miscellaneous page also provides the user with updated information related to these three cancers in the form of-country centered reports, popular journals and recent publication in the field with the hyperlinks for detailed evidence and annotation.

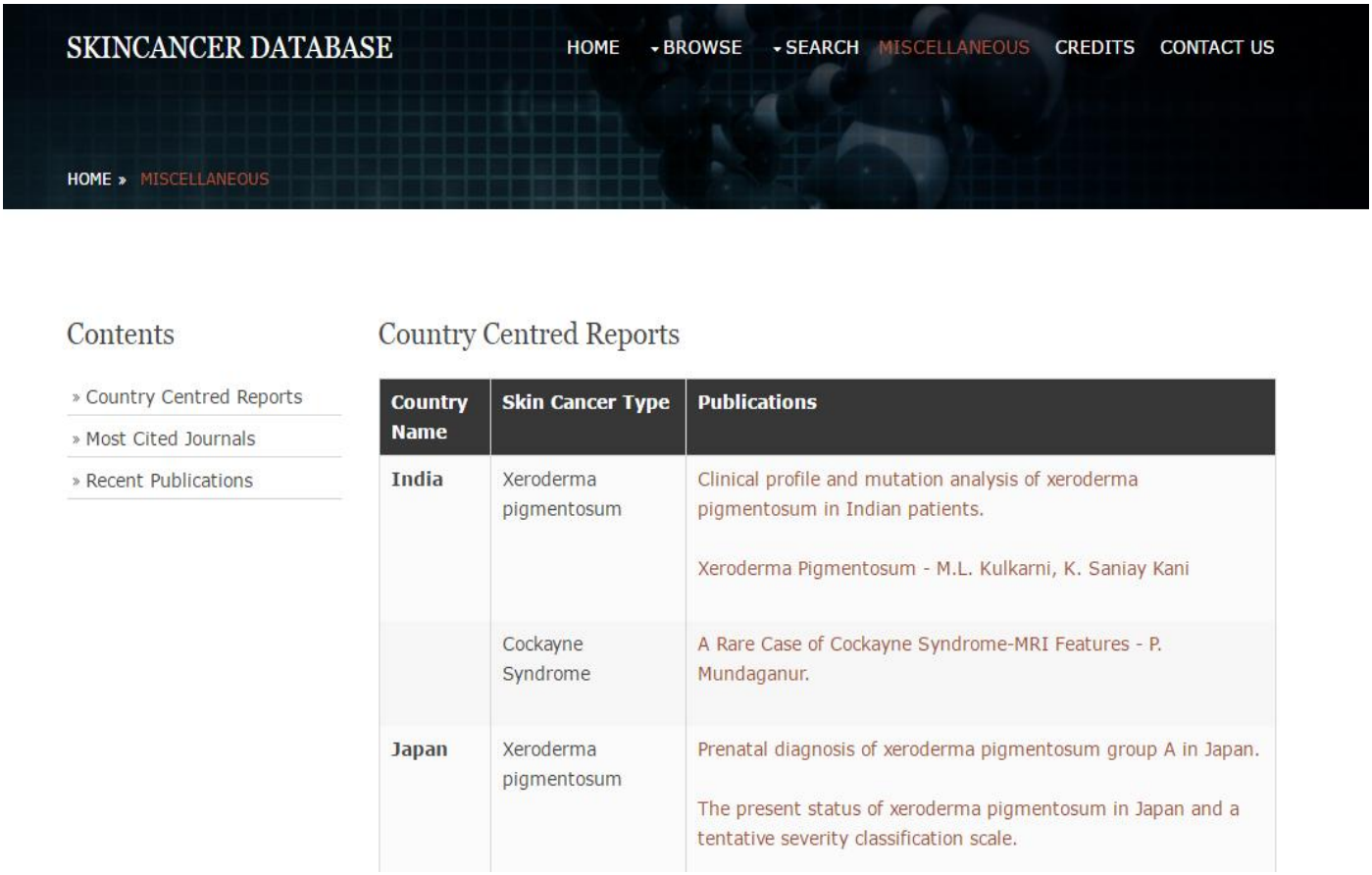


Fig.11: Image showing Miscellaneous page in SkinCancerDB.

4.2 Representative Entry in SkinCancerDB

The figure below (Fig.12) displays the sample output page created by querying SkinCancerDB using the keyword 'Xeroderma Pigmentosum'. Associated annotations of the cancer can be obtained by clicking various subcategorized options in the search page - Literature, Gene/Molecular Data, Expression Data, Protein Interactions, NER Pathway, Clinical Information.

A Literature Citations

Genre	PMID	Title	Author(s)	Description
Xeroderma Pigmentosum	2408009	Complementation of the xeroderma pigmentosum DNA repair synthesis defect with Escherichia coli UvrABC proteins in a cell-free system.	Hansson J, Grossman L, Lindahl T, Wood RD	A newly developed cell-free system was used to study DNA repair synthesis carried out by extracts from human cell lines in vitro.
Xeroderma Pigmentosum	26722489	A meta-analysis of xeroderma pigmentosum gene D Ls751Gln polymorphism and susceptibility to hepatocellular carcinoma	Wang Y, Zhao Y, Zhang A, Ma J, Wang Z, Zhang X	This study used a meta-analysis approach to comprehensively investigate the correlation between XPD polymorphism and HCC susceptibility in Chinese population

B Genes Information

Gene Name	Full Name	Function	Location	Ensembl ID	OMIM ID	UniProtKB ID	HGNC ID	MGI ID
XPA	Xeroderma Pigmentosum Group A-Complementing Protein	Involved in DNA excision repair. Initiates repair by binding to damaged sites with various affinities, depending on the photoproduct and the transcriptional state of the region.	9q22.33	ENSG00000136936	611153	P23025	12814	99135

C COSMIC Gene Expression Data

Gene Name	Total Samples	Overexpr. Samples	Overexpr. %	Underexpr. Samples	Underexpr. %	COSMIC ID
XPA	473	9	1.9	12	2.54	COSG904

D NCBI-SRA Data

Target Gene	Title	Primary ID	Study	Bioproject ID	Sample Species	Biosample ID	Instrument
XPA	Whole exome sequence analysis of XPA deficient cells and their iPS cells	SRR2037122	Homo sapiens exome	PRJNA284651	Human XPA deficient fibroblast and iPS cells	SAMN03704960	Illumina HiSeq 1000

E Protein Family Information

Protein Name	Gene Name	Family/Domain	Domain Architectures	UniProtKB ID(s)
XPAC	XPA	XPA C-terminal	22	P23025, F2Z2T2, W0FSR8

F Protein-Protein Interactions

Gene Name	Functional Partners	Interaction Basis	Properties	Score
XPA	RPA1	Experiments, databases, text mining	Involved in DNA replication, recombination, repair	0.999
XPA	ERCC3	Databases, text mining	Involved in NER mechanism of DNA	0.999

G Gene Ontology Data

Gene_Name	Gene_Ontology ID(s)	Function_Name	Definition
XPA	GO:0006289	Nucleotide-excision repair	A DNA repair process in which a small region of the strand surrounding the damage is removed from the DNA helix as an oligonucleotide.

H Drugs And Therapeutics Information

Drug Name	Active Ingredient	Disease Treatment	Current Status	Properties	Side Effects	Manufacturers
Imiquimod	Imiquimod	Actinic keratosis, Basal cell carcinoma, Kaposi's sarcoma, (Xeroderma pigmentosum - undergoing investigation)	Approved / 24-06-2014	Treatment of biopsy confirmed primary superficial basal cell carcinoma in immunocompetent adults	Erythema, edema, scabbing or crusting, flaking or scaling, erosion, weeping	Taro Pharmaceuticals, Inc., Hawthorne, NY, USA

I Clinical Trials Data

Disease Type	Trial Name	Clinicaltrials.gov identifier	Start_Date	Update_Date	Description	Study Type	Current Status
Xeroderma pigmentosum	Cancer Risk in Carriers of the Gene for Xeroderma Pigmentosum	NCT00046189	18-Sep-02	24-Jan-17	This study will determine if family members of patients with xeroderma pigmentosum (XP) have various abnormalities, including: skin abnormalities;	Observational	Recruiting

Fig.12: This image shows the sample output page for the keyword ‘Xeroderma Pigmentosum’ in sub-diverse options of the Search page in SkinCancerDB.

A. 'Literature' section includes all the journals and literature evidences related to the skin cancers with their hyperlinked PubMed IDs.
B. In the 'Gene/Molecular Data' page, gene ID, gene official symbol, gene name, gene chromosomal location and its function are displayed.

C. The 'Gene Expression Data' page shows the collected COSMIC and SNP/SRA data for the respective cancer genes.

D. The 'Protein Family' classification page shows the protein families or domains associated to the target genes, along with their UniProtKB IDs, also hyperlinked.

E. In the 'Protein-Protein Interactions' page, the interactions between the gene and its functional partners is depicted, where the basis of interaction between the former and the latter is based on experimental data, text mining data, information through databases etc. and the score indicates the confidence of an interaction of protein networks, i.e., on the basis of evidence (interaction basis), the database evaluates whether the PPI for that gene is true or not.

F. In the 'NER Pathways' page, the target genes involved with repair mechanism pathways are described. Through the entry, we can link to the corresponding pathway in KEGG database.

G. In the 'Gene Ontology' page, particular functions such as DNA repair mechanism, GG-NER, TC-NER etc. are described for each gene, and through their GO IDs we can link to GO Consortium where these functions are explained in detail.

H. In the 'Drugs and Therapeutics' page, drugs for respective cancers are presented with their FDA approval date, chemical name and functionality.

I. In the 'Clinical Trials', we have displayed information about the clinical trials carried out for each disease and ClinicalTrials.gov identifier is hyperlinked for further references.

(Note:- The similar search can be conducted for other cancers as well (CS, TTD) by using search keywords like - cancer names, literature ID's, gene names, drug names and many others.)

4.3 Bio-Computational Analysis

For the biological analysis of the collected data, we searched through GEO repository to assemble the microarray data for all the twelve cancer genes. But we only got results for four out of twelve genes. This is primarily due to due to rarity of XP, CS and TTD cancers and hence, limited scientific and biological research conducted in the area.

The figure below explains the procedural steps followed for implementing the Bio-computational analysis of microarray data through the use of WB-DEGS application (Fig.13):

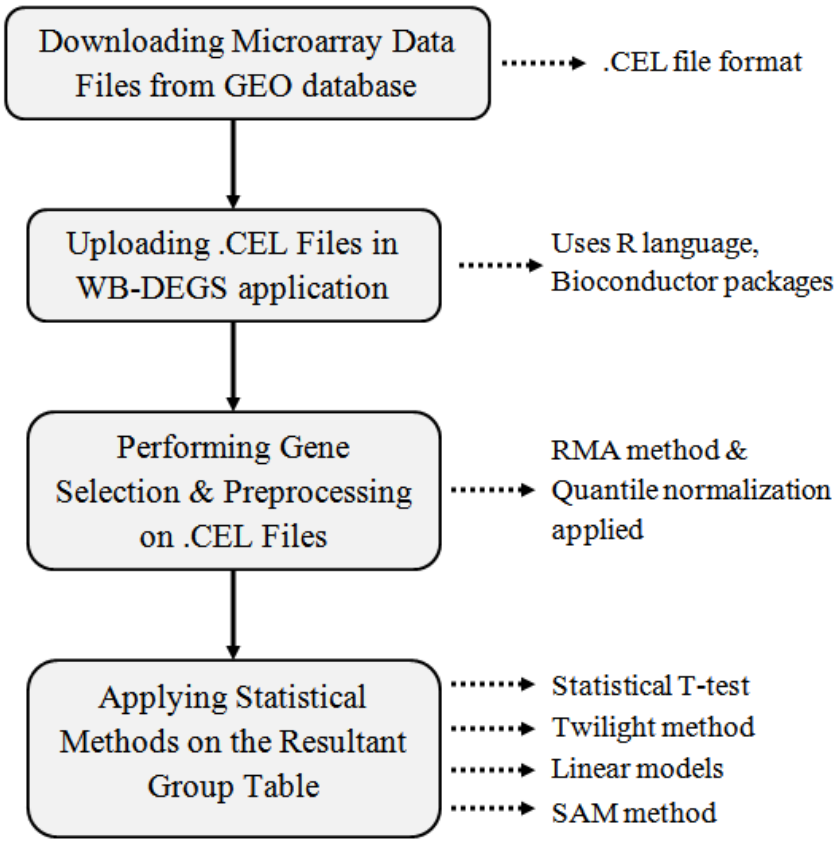


Fig.13: Flowchart describing the workflow of microarray data analysis in WB-DEGS.

The results were obtained for within and between group comparisons of the microarray data for CSB, where the replicate data was divided in two groups (up-regulated and down-regulated genes), and the comparison was made on the basis of Background Correction (RMA method) and Normalization methods (Quantile normalization); and further the gene selection and statistical methods were applied to the data.

The statistical methods used along with the results obtained for CSB gene are as follows:

- **Simple Statistical Test (T-test)** - By this method, genes are assigned to one of the two groups depending on the cut-off values set for t-test parameters. We obtained the results in T-test as volcano plots where the red region shows the selected genes being expressed amongst the within group comparison and between group comparison.

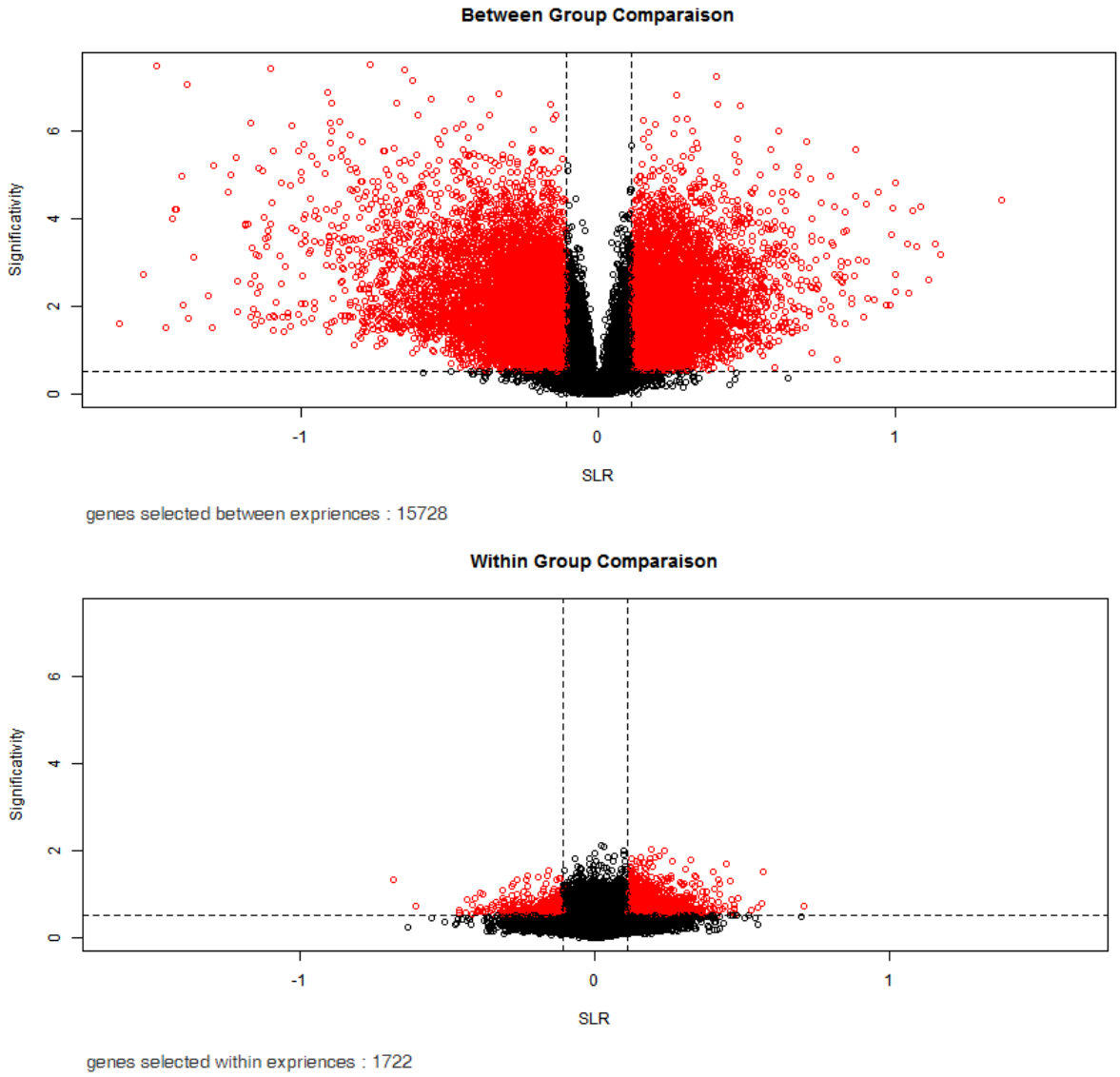


Fig.14: Volcano Plots showing T-test analysis for CSB gene.

- **Twilight method** - This function uses two-condition test and a stochastic downhill search to compute permutation-based P-values and obtain size distribution plots. Based on the cut-off in P-value set for the twilight method, we obtained the volcano plots which represent the selected genes and non-selected genes as red and black, respectively.

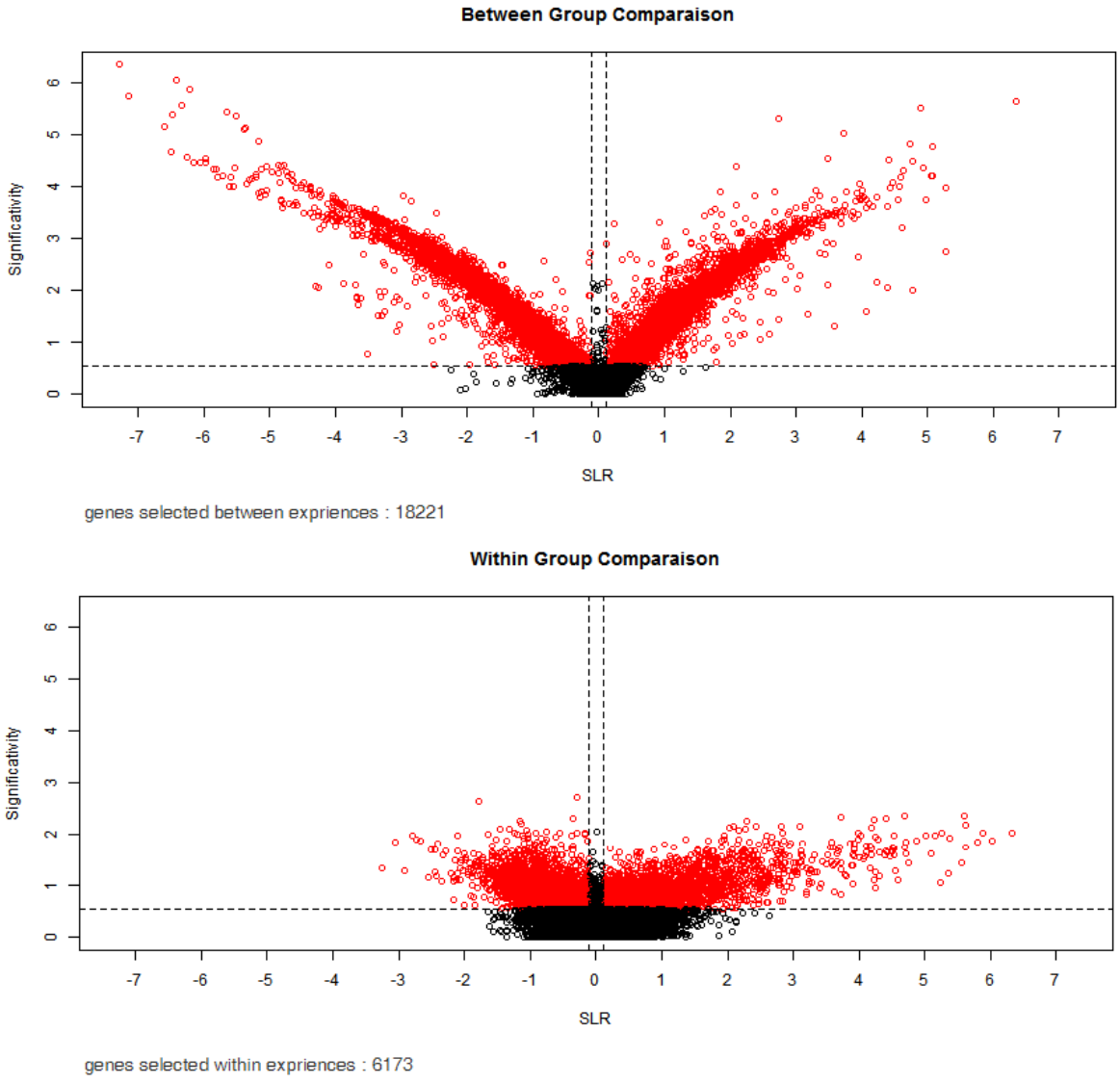


Fig.15: Volcano plots showing Twilight method analysis for CSB gene.

- **Linear models** - It uses regression and ANOVA methods to estimate missing data, block and factorial designs shown on a volcano plot. By setting the cut-off thresholds for deducing the linear model for CSB gene data, we obtained the volcano plots showing higher percentage of selected genes in between comparison group plot, but opposite in case of within comparison group.

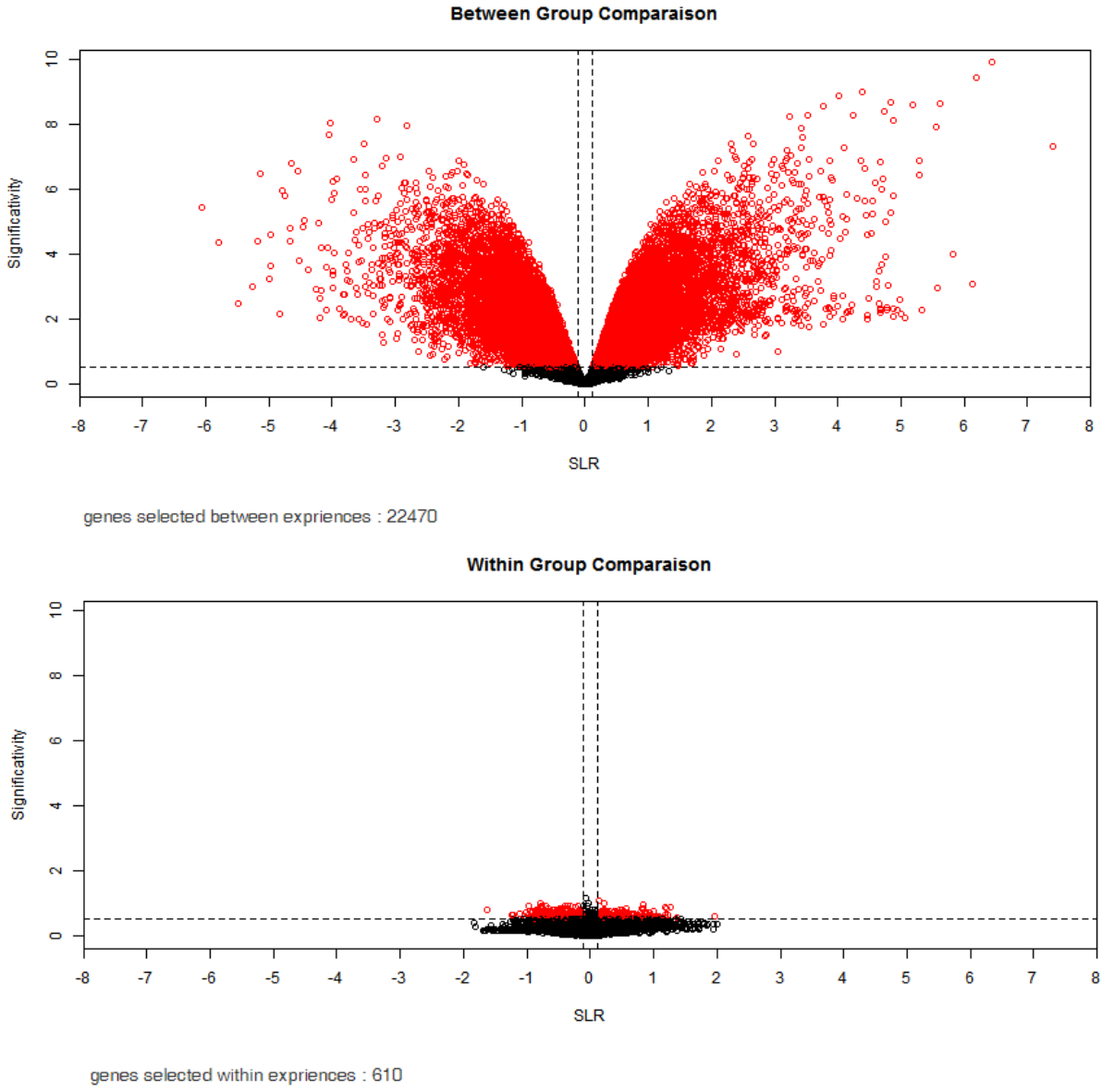


Fig.16: Volcano plots showing Linear model analysis for CSB gene.

- **SAM (Significance analysis of Microarrays)** - This method is used for detecting probes in microarray data so as to determine differential expression between subsets of the conditions. By permutations, the technique deduces the empirical estimates for the FDR (false discovery rates) of differential genes reported. So, for this method, taking delta value to be 2, we got the results as higher occurrence of selected genes in the plot of between experiences, but no selected genes in case of within comparison plot.

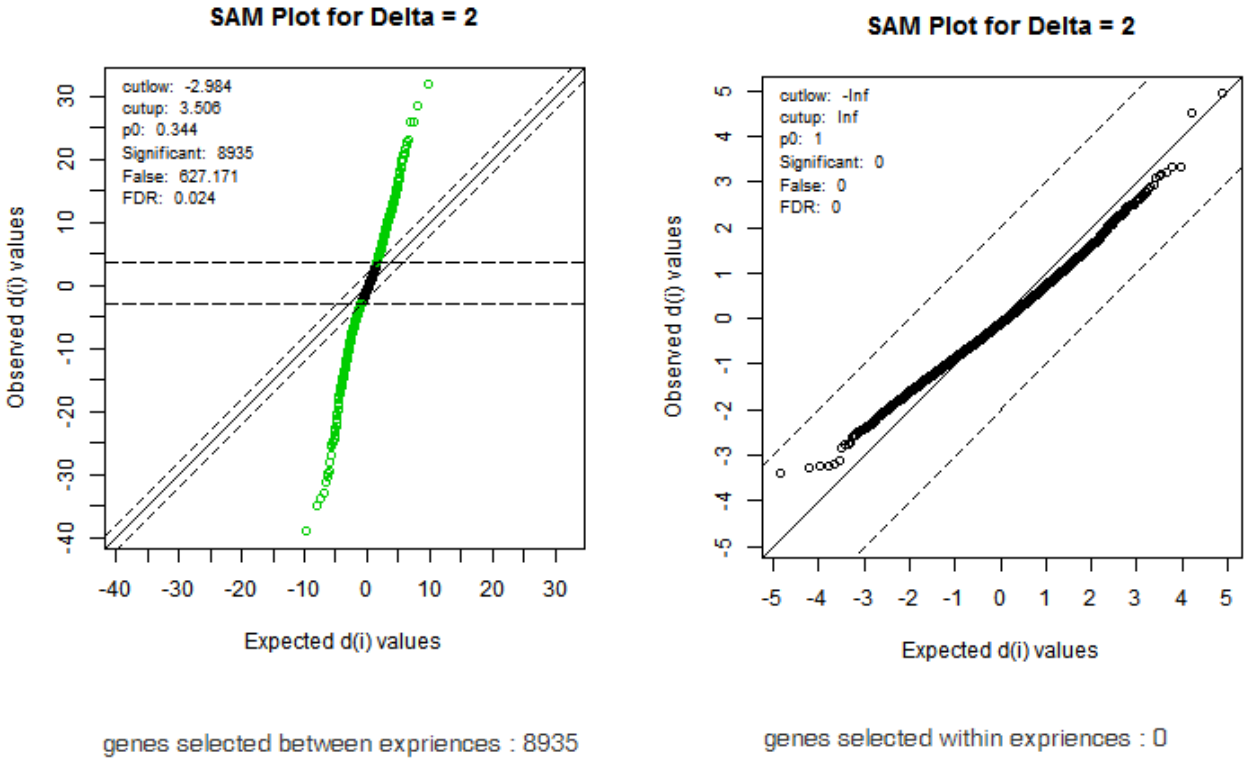


Fig.17: Slope graph obtained for CSB gene by SAM method.

And for comparing the results obtained by these statistical methods, we summarized these results by drawing the Venn Diagram, which shows the number of accepted genes in the colored region of the graphs, along with number of similar genes within overlapped regions.

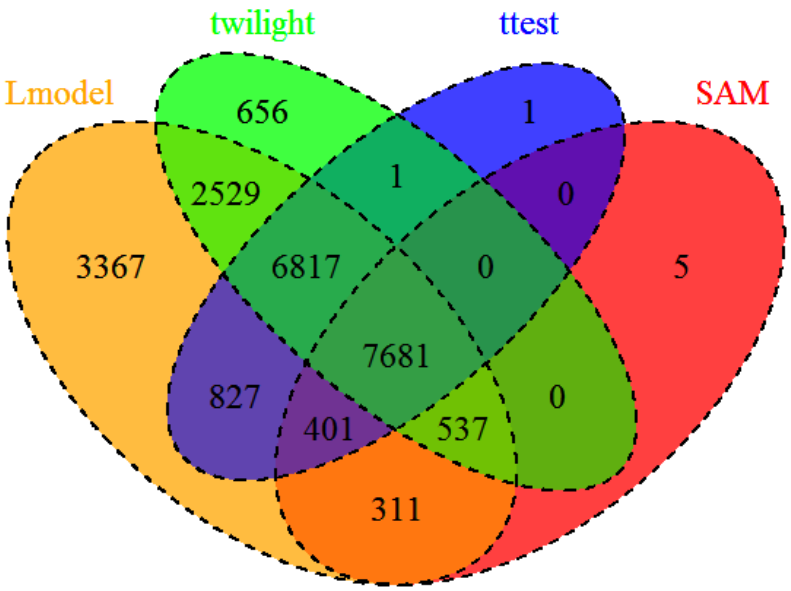


Fig.18: Venn Diagram plotted for all the four statistical methods showing numbers as no. of accepted genes. Overlapped genes from various methods can also be drawn from this figure.

5. Chapter-4

Conclusion and Future Prospects

5.1 Conclusion

SkinCancerDB is one of its kind, literature based NER Specific Skin Cancer database covering not only the Literature, Gene Expression, Molecular and Interaction data but also the Clinical Information like- Manifestations and Epidemiology, Drugs and Therapeutics, Clinical Trials etc, along with the in depth analysis of the collected information.

The following image represents the unique characteristic features of the SkinCancerDB along with its application in the biomedical field:

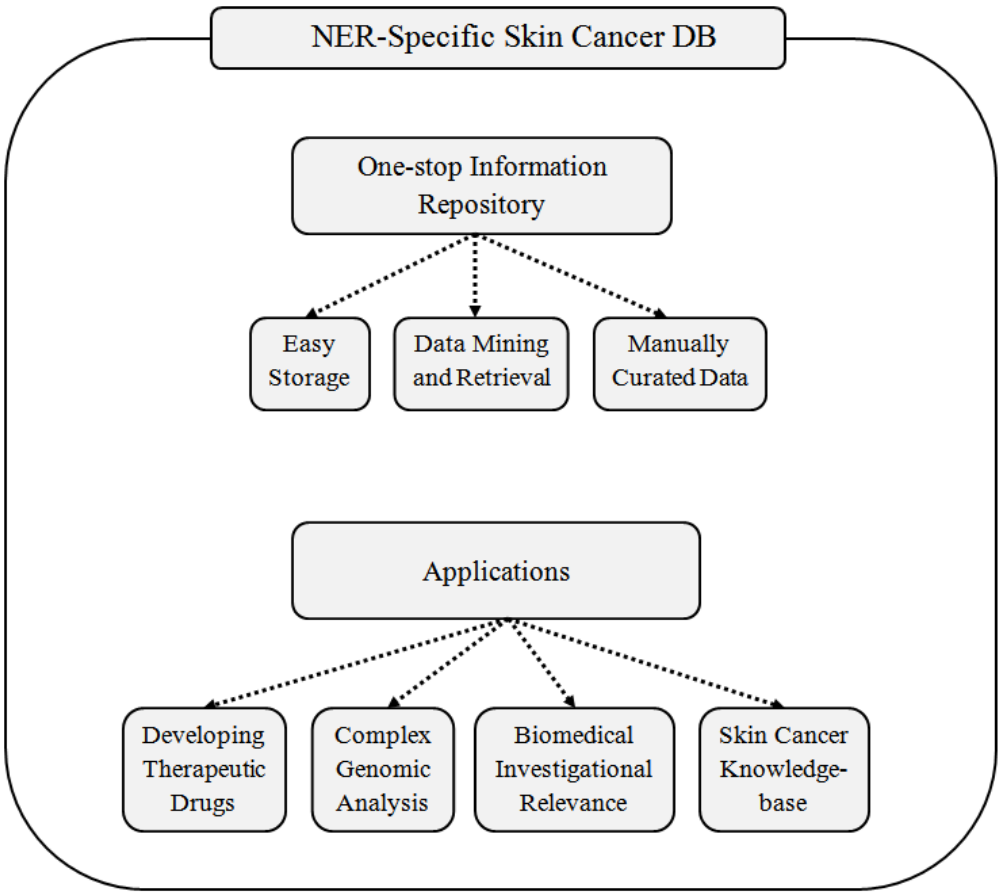


Fig.19: Plan for implementation of future prospects defined for NER-Specific Skin Cancer DB.

The database is primarily based on literature extraction of multi-dimensional data of genomic, expression, molecular, genetic, pathway and clinical information. It is a web accessible user friendly interface which facilitates and supports multi-keyword searching.

It is a database that will continue to grow, add the advancements and up-gradation in biological, medical and technological terms in the above mentioned skin cancers.

5.2 Limitations of this study

The primary limitation of this study is that we were not able to perform bio-computational and expression analysis of cancer associated genes respectively. This was partly due to lack of experimental, scientific and clinical studies conducted on XP and also due to rarity of CS and TTD. This major gap in biological knowledge resulted only in analysis of one gene i.e. CSB, rendering a rather large gap in our previously aspired analytical database to just database.

Per se, there are no further limitations to be discussed other than these but there are future possibilities, which will be discussed in next subsection.

5.3 Future Possibilities

Given that we were not able to create an analytical database for second part (8^{th} Sem), and that there exists on other database focusing on just the NER specific skin cancers, it is safe to assume that SkinCancerDB provides a quick start for the bio-computational analysis itself from the scratch. Moreover, it is possible to add or design the algorithm for in-built analytical tool depending upon the nature of data and analysis required, further with the continued addition of new biological, biomedical and scientific insights SkinCancerDB can be a small dedicated section of already well established and popular databases.

BIBLIOGRAPHY

1. Lehmann AR (2003). DNA repair - deficient diseases, Xeroderma pigmentosum, Cockayne syndrome and Trichothiodystrophy. Biochimie. 85(11):1101-11.
2. DiGiovanna J.J., Kraemer K.H (2012). Shining a light on xeroderma pigmentosum. J. Invest. Dermatol.;132:785–796.
3. Friedberg EC, Walker GC, Siede W, Wood RD, Schultz RA, Ellenberger T (2005). DNA repair and mutagenesis. ASM Press, Washington, DC.
4. Wood RD, Robins P, Lindahl T (1988). Complementation of the xeroderma pigmentosum DNA repair defect in cell-free extracts. Cell 53: 97–106.
5. Huang JC, Svoboda DL, Reardon JT, Sancar A (1992). Human nucleotide excision nuclease removes thymine dimers from DNA by incising the 22nd phosphodiester bond 5′ and the 6th phosphodiester bond 3′ to the photodimer. Proc Natl Acad Sci 89: 3664–3668.
6. Sehgal M, Singh TR (2014). Systems biology approach for mutational and site-specific structural investigation of DNA repair genes for xeroderma pigmentosum. Gene;10;543(1):108-17.
7. Dupuy A, Sarasin A (2015). DNA damage and gene therapy of xeroderma pigmentosum, a human DNA repair - deficient disease. Mutat Res; 776: 2-8.
8. Lahlimi F, Harif M, Elhoudzi J (2016). [Nephroblastoma and xeroderma pigmentosum: A rare association]. Arch Pediatr; 23(1): 75-7.
9. Karentz D (2015). Beyond xeroderma pigmentosum: DNA damage and repair in an ecological context. A tribute to James E. Cleaver. Photochem Photobiol; 91(2): 460-74.
10. Orlando D. Schärer (2013). Nucleotide Excision Repair in Eukaryotes. Cold Spring Harb Perspect Biol. 5(10): a012609. doi: 10.1101/cshperspect.a012609.
11. Bowden NA, Beveridge NJ, Ashton KA, Baines KJ, Scott RJ (2015). Understanding Xeroderma Pigmentosum Complementation Groups Using Gene Expression Profiling after UV-Light Exposure. Int J Mol Sci; 16(7): 15985-96.
12. Feltes BC, Bonatto D (2015). Overview of xeroderma pigmentosum proteins architecture, mutations and post-translational modifications. Mutat Res Rev Mutat Res; 763: 306-20.

13. Santiago KM, França de Nóbrega A, Rocha RM, Rogatto SR, Achatz MI (2015). Xeroderma pigmentosum: low prevalence of germline XPA mutations in a Brazilian XP population. Int J Mol Sci; 16(4): 8988-96.
14. Ikehata H, Chang Y, Yokoi M, Yamamoto M, Hanaoka F (2014). Remarkable induction of UV-signature mutations at the 3'-cytosine of dipyrimidine sites except at 5'-TCG-3' in the UVB-exposed skin epidermis of xeroderma pigmentosum variant model mice. DNA Repair (Amst); 22:112-22.
15. Lehmann J, Schubert S, Emmert S (2014). Xeroderma pigmentosum: diagnostic procedures, interdisciplinary patient care, and novel therapeutic approaches. J Dtsch Dermatol Ges; 12(10):867-72.
16. Lambert WC, Lambert MW (2015).Development of effective skin cancer treatment and prevention in xeroderma pigmentosum. Photochem Photobiol; 91(2): 475-83.
17. Black JO (2016). Xeroderma Pigmentosum. Head Neck Pathol; 10(2): 139-44.
18. Woźniak K, Kuc D, Błasiak J, Kurowska AK, Szaflik J, Szaflik JP (2014). [Ocular manifestations in hereditary diseases with defects in DNA repair]. Klin Oczna; 116(2): 142-5.
19. Reichrath J, Rass K (2014). Ultraviolet damage, DNA repair and vitamin D in nonmelanoma skin cancer and in malignant melanoma: an update. Adv Exp Med Biol; 810: 208-33.
20. Lehmann AR (2015). Xeroderma pigmentosum in the United kingdom. Photochem Photobiol; 91(2): 484-5.
21. Moriel-Carretero M, Herrera-Moyano E, Aguilera A (2015). A unified model for the molecular basis of Xeroderma pigmentosum-Cockayne Syndrome. Rare Dis; 3(1): e1079362.
22. Giustini S, Miraglia E, Berardesca E, Milani M, Calvieri S (2014). Preventive Long-Term Effects of a Topical Film-Forming Medical Device with Ultra-High UV Protection Filters and DNA Repair Enzyme in Xeroderma Pigmentosum: A Retrospective Study of Eight Cases. Case Rep Dermatol; 6(3): 222-6.
23. Fu L, Xu X, Ren R, Wu J, Zhang W, Yang J, Ren X, Wang S, Zhao Y, Sun L, Yu Y, Wang Z, Yang Z, Yuan Y, Qiao J, Izpisua Belmonte JC, Qu J, Liu GH (2016). Modeling xeroderma pigmentosum associated neurological pathologies with patients-derived iPSCs. Protein Cell; 7(3): 210-21.
24. Bensenouci S, Louhibi L, De Verneuil H, Mahmoudi K, Saidi-Mehtar N (2016). Diagnosis of Xeroderma Pigmentosum Groups A and C by Detection of Two Prevalent Mutations in West

Algerian Population: A Rapid Genotyping Tool for the Frequent XPC Mutation c.1643_1644delTG. Biomed Res Int; 2180946.

25. Gonzalez EA, Mudry MD, Palermo AM (2014). DNA repair kinetic of hydrogen peroxide and UVA/B induced lesions in peripheral blood leucocytes from xeroderma pigmentosum patients and healthy subjects. J Environ Pathol Toxicol Oncol; 33(4): 279-93.
26. Sassa A, Kamoshita N, Kanemaru Y, Honma M, Yasui M (2015). Xeroderma Pigmentosum Group A Suppresses Mutagenesis Caused by Clustered Oxidative DNA Adducts in the Human Genome. PLoS One;10(11):e0142218.
27. Kraemer KH, Lee MM, Scotto J (1984). DNA repair protects against cutaneous and internal neoplasia: evidence from xeroderma pigmentosum. Carcinogenesis. 5(4):511-4.
28. Botta E, Nardo T, Broughton BC, Marinoni S, Lehmann AR, Stefanini M (1998). Analysis of mutations in the XPD gene in Italian patients with trichothiodystrophy: site of mutation correlates with repair deficiency, but gene dosage appears to determine clinical severity. Am J Hum Genet. 63(4):1036-48.
29. Cui YP, Chen YY, Wang XM, Wang XL, Nan X, Zhao H (2015). Two Novel Heterozygous Mutations in ERCC8 Cause Cockayne Syndrome in a Chinese Patient. Pediatr Neurol. 53(3):262-5.
30. Koch S, Garcia Gonzalez O, Assfalg R, Schelling A, Schäfer P, Scharffetter-Kochanek K, Iben S (2014). Cockayne syndrome protein A is a transcription factor of RNA polymerase I and stimulates ribosomal biogenesis and growth. Cell Cycle.13(13):2029-37.
31. Wang Y, Chakravarty P, Ranes M, Kelly G, Brooks PJ, Neilan E, Stewart A, Schiavo G, Svejstrup JQ (2014). Dysregulation of gene expression as a cause of Cockayne syndrome neurological disease. Proc Natl Acad Sci U S A.; 111(40):14454-9.