COMPUTATIONAL ANALYSIS OF EXPRESSION AND INTERACTION DATA TO REVEAL ROLE OF DNA MISMATCH REPAIR IN MSI, HNPCC AND CRC

Dissertation submitted in partial fulfillment of the requirement for the degreeof

BACHELOR OF TECHNOLOGY

IN

BIOTECHNOLOGY AND BIOINFORMATICS

By

Arushi Sharma, Sadhika Behl

(141508,141510)

UNDER THE GUIDANCE OF

Dr. Tiratha Raj Singh

Department of Biotechnology and Bioinformatics

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

May,2018

TABLE OF CONTENTS

	Page Number
DECLARATION BY THE SCHOLAR	**I**
CERTIFICATE	**IV**
ACKNOWLEDGEMENT	**V**
LIST OF ABBREVIATIONS	**VI**
List of FIGURES	**VIII**
List of TABLES	**X**
ABSTRACT	**XI**

CHAPTER-1

Introduction 2

1. 1 Problem Statement 7

1. 2 Objectives 8

1. 3 Expected Outcomes 8

CHAPTER-2

Materials and Methods 9

2. 1 Data Analysis 9

2. 2 Database and GUI development, connectivity and data retrieval 16

CHAPTER-3

Results and Discussions **20**

CHAPTER-4

Conclusion **52**

References **53**

DECLARATION BY THE SCHOLAR

I hereby declare that the work reported in the B-Tech thesis entitled **"Computational analysis of expression and interaction data to reveal ole of DNA mismatch repair in MSI, HNPCC and CRC"** submitted at **Jaypee University of Information Technology, Waknaghat India,** is an authentic record of my work carried out under the supervision of **Dr.Tiratha Raj Singh**. The data mentioned in this report was obtained during genuine work done by me.

I therefore declare that data and results are true to the best of my knowledge. I have not submitted this work elsewhere for another degree or diploma.

Arushi Sharma (141508)

Sadhika Behl (141510)

Department of Biotechnology and Bioinformatics

Jaypee University of Information Technology, Waknaghat, India

Date:

CERTIFICATE

This is to certify that the work reported in the B-Tech. thesis entitled **"Computational analysis of expression and interaction data to reveal role of DNA Mismatch Repair in MSI, HNPCC and CRC"** submitted by **Ms. Arushi Sharma and Ms. Sadhika Behl** at **Jaypee University of Information Technology, Waknaghat ,India,** is a bonafide record of his / her original work carried out under my supervision. This work has not been submitted elsewhere for any other degree or diploma.

Signature:

Supervisor: Dr. Tiratha Raj Singh

Designation: Associate Professor

Date:

Acknowledgement

We would take the opportunity to express our gratitude for our mentor and guide Dr. Tiratha Raj Singh, Associate Professor in Department of Biotechnology and Bioinformatics, who initiated us into the domain of research and regulated us with limited patience and without whose significant recommendation and unstinted co-task, the present dissertation would not have been conceivable.

We would express genuine thanks to Dr. SudhirSayal (Head of Department of Biotechnology and Bioinformatics) for his order and support on this task.

Besides, we would like to thank the Administration of the Department of Biotechnology and Bioinformatics for the greater part of their specialized aptitude and support.

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

List of Abbreviations

Abbreviations	Phrase
DNA	Deoxyribonucleic acid
MMR	Mismatch Repair
HNPCC	Hereditary non-polyposis Colorectal Cancer
CRC	Colorectal Cancer
UV	Ultraviolet
mtDNA	Mitochondrial DNA
MSI	Micosatellite Instability
RPA	Replication Protein A
DRMAP	DNA Repair Malignancies Annotation Platform
RNA	Ribonucleic acid
PDB	Protein Data Bank
PFam	Protein Family
STRING	Search Tool for the retrieval of Interacting Genes/Proteins
IHOP	Information Hyperlinked over Proteins
MINT	Molecular INTeraction database
HPRD	Human Protein Reference Database
WB-DEGS	Within and Between Group Comparisons for Differentially Expressed Gene Selection Gene
RMA	Robust Multiarray Averaging
SAM	Significant Analysis of Microarray
FDR	False Discovery Rate
DAVID	Database for Annotation, Visualization and Integrated Discovery
PANTHER	Protein Analysis Through Evolutionary Relationships
WebGestalt	WEB-based GEneSeTAnaLysis Toolkit
GO	Gene Ontology
GSE	Gene expression Series
CEL	

GLAD4U	Gene list automatically derived for you
HPA	High Performance Analysis
FANTOM5	Functional Annotation of the Mouse/Mammalian Genome
GTEx	Genotype Tissue Expression
FC	Fold Change
PonyORM	Pony Object Relational Mapper
HTML	Hyper Text Markup Language
CSS	Cascading Style Sheet
PHP	Hypertext Preprocessor
JS	Java Script
P	Biological Process
M	Molecular Function
C	Cellular Component
TPM	Tags Per Million
RPKM	Read Kilo Per Million
DDR	DNA damage Response
TGF	Transforming Growth Factor
GEO	Gene Expression Omnibus
HPA	Human Protein Atlas

LIST OF FIGURES

Figure Number	Caption
1.1	Mechanism of DNA MMR in Eukaryotes
1.2(a)	HNPCC is an autosomal dominant disease.
1.2(b)	Symbols in the figure 1.2(a)
2.1	Various tools and databases being utilized in data collection and analysis
2.2	Systematic representation of steps used to analyze the data
2.3	Workflow of WB-DEGS
2.4(a)	Uploaded Files
2.4(b)	Gene Expression matrix is obtained when the files are uploaded
2.4(c)	Selection of Background Correction
2.4(d)	Background Correction Method and Normalisation Method
2.4(e)	Two groups are added as diseased and normal
2.5	After retrieving the WB-DEGS data, these steps are followed
2.6	Quantitative Analysis for the set of genes to know their role in HNPCC and CRC
2.7	ER diagram using PonyORM
2.8	CSS implemented for background
2.9	PHP deployed to set up the connection
3.2	Volcano Plot for between group comparison for significance (y axis) versus fold change (x axis)

3.3	Venn Diagram generated after applying the normalisation methods
3.4	Annotation tree obtained from the WebGestalt for various regulatory biological process
3.5	Graphs for the GO for three parameters: P, C and F
3.6	RNA expression of Colon cells in MSH2 using HPA, GTEx and FANTOM5
3.7	Home page on the website showing the available search options
3.8	About page displaying the objectives and Future Aspects
3.9	Displays the dropdown for the Uniprot IDs
3.10	Dropdown for the Ensemble IDs is displayed
3.11	Shows the Gene as dropdown
3.12	Results are retrieved for the options Uniprot IDs, ensemble Ids and gene
3.13	Dropdown and results for Cellular Components
3.14	Dropdown and results for Molecular Function
3.15	Dropdown and results for biological process
3.16	Interaction results from STRING.
3.17	Intact results for interactions
3.18	RNA and protein expression

LIST OF TABLES

Table Number	Caption	Page Number
1.1	Enzymes involved in the MMR system	
3.1	Functional database GLAD4U applied to get the results related to HNPCC and CRC.	
3.2	RNA expression in Colon cells of genes	
3.3	protein expression in colon tissues	
3.4	Gene along with their proteins and functions.	

Abstract

Microsatellite instability (MSI) is an error mechanism associated with DNA mismatch repair (MMR) system constituting a set of genes. If MMR fails, MSI may lead to various forms of cancers such as hereditary non polyposis colorectal cancer (HNPCC). In this study, we explored the gene expression and network data to reveal the significance of MSI in HNPCC. Genes and proteins were observed for their specific role in HNPCC with respect to MSI and MMR. Besides standard markers, more genes were identified as putative markers having significant contribution in the regulation of the mechanisms associated with MSI and MMR for HNPCC. Experimental validation of these genes will prove to a promising outcome for further research and will aid in the maintenance of the disease. The crucial genes obtained from the study were assimilated in the database and the website.

CHAPTER-1

INTRODUCTION

DNA MMR (Mismatch Repair) system is the nethermost researched topics of the last decade. Its connection with HNPCC (Hereditary Non-Polyposis Colorectal Cancer and other malignancies makes it a comprehensive subject of interest. DNA is the basic genetic material in any organism and it is often subjected to many chemical alterations can cause a damage to it(Li et al., 2007). Mutations in the DNA may be due to insertion, deletion and misincorporation of the bases that may be during the synthesis, replication and recombination in DNA (Leach et al., 1993, H. Lynch et al., 1996, C. Bronner et al., 1994)

A compendium of MMR as a database is warranted especially in terms of understanding the next generation sequencing technologies and their applications.

Chemical alterations occurs as a result of oxygen radicals that may arise during respiration, x-ray and gamma rays that produce ionizing radiations, UV radiations, aromatic hydrocarbons, plant and microbial products and the chemicals used for chemo therapies.

These results in the misincorporationof bases due to which they do not show complementary base pairing resulting in mismatches (Iyer et al., 2006)e.g., C could be mispaired to A and T could be mispaired to G. DNA damage may be due to single base alteration (depurination and deamination), two base alteration-pyrimidine’s dimer, chain breaks-ionizing radiation and cross linkages that occur between bases(T. Kunkel et al., 2015) There are various enzymes that aid in repairing the MMR (*Table 1*)(M. Schofield et al., 2003, R. Lahue et al., 1989).

Table 1.1: Enzymes involved in the MMR system

Organisms and Enzymes			Functions
E. Coli	S. cerevisiae	Human	Functions of Eukaryotic Proteins
MutS	MSH2	MSH2	MutSα(with MSH6; 80-90%); MutSβ(with MSH3)
MutS	MSH3	MSH3	MutSβ(with MSH2); repair of larger loops
MutS	MSH6	MSH6	MutSα(with MSH2); repair of mismatches and small loops
MutL	MLH1	MLH1	Forms heterodimers with the other three MutL homologs
MutL	PMS1	PMS2	MutLα(90%); Mismatch repair; endonuclease motif
MutL	MLH2	PMS1	MutLβ; Role unknown
MutL	MLH3	MLH3	MutLϒ; Mismatch repair; endonuclease

MMR system excisemispaired bases that are a result of replication errors and recombination between faulty complemented sequences in DNA(G. Li., 2007). The mutations are detected by the polymerase and the errors can be resolved using the system. Inactivation of MMR system may lead to HNPCC, aging, Rothmundthomson syndrome and many other diseases. MLH1 is a key player in mitochondrial DNA (mtDNA) MMR system and on how they perform their crucial roles (Sehgal et al., 2015, F. Kadyrov et al., 2006).

DNA MMR is strongly conserved biological pathway which helps in maintaining the genomic stability (Yang, 2000). DNA damage leads to cell death or enough mutations which may lead to reduced fitness, poor regulation of transcription patterns, andeventually the aging phenotype. MMR maintains the stability of the microsatellite which is disturbed due to the ageing (V. Gorbunova et al., 2007). The DNA repair if not

treated can lead to the genomic instability referred to as microsatellite instability(MSI) which is often seen in the form of cancers such as HNPCC.

The mechanism of MMR in eukaryotes includes many genes for the correction of mismatch (Kolodner et al., 1999). The Mismatch is recognized by the MutSalpha i.e., a complex of MSH2 and MSH6. MutLalpha includes MLH1 and PMS2 that incise the part where there is a mismatch. Exonuclease 1 cuts the part to be corrected. Replication Protein A (RPA) binds the strand that stabilises the process. DNA polymerase resynthesise the strand with the help of PCNA. DNA ligase links the two strands which corrects the mismatch. In this way the whole process works (*figure 1*).

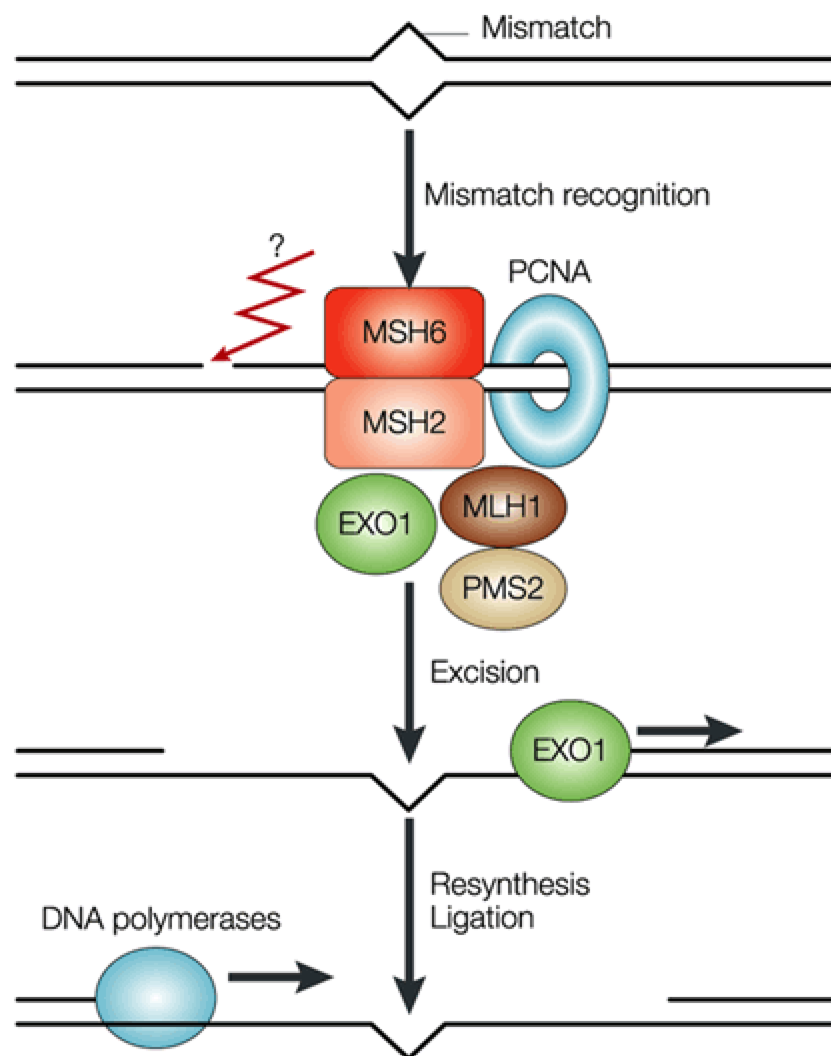


Figure 1.1: Mechanism of DNA MMR in Eukaryotes

Many researchers have been working in this field and have found significant results in this context. Our focus is on Colorectal Cancer and its hereditary form.HNPCCis an autosomal dominant syndrome as a consequence of defective MMR genes (shown in

Figure 3) (Sehgal et al., 2014).It is a heterogeneous disease and proceeded when there are germline defects in one of atleast four MMR genes.CRC is the third largest cancer in the world. The global risk of it is 1 in 22 in men and 1 in 24 in women (*Navarro M et al., 2016)*.Since defectsin the MSH2 quality may represent upwards of 60% of HNPCC cases, and imperfections in the MLH1 quality may assume a part in up to 30%, abandons in these 2 qualities likely record for by far most of HNPCC.Many other genes are also responsible for the disease (Shukla et al., 2016).

HNPCC is distinguished into Lynch syndrome I (familial colon cancer) and Lynch syndrome II (HNPCC associated with other cancers of the gastrointestinal [GI] or reproductive system).

In addition to this, we also studied the colorectal cancer. The expanded danger of malignancy is because of acquired transformations that corrupt the self-repair ability of DNA. The intestine involves a series of pits or crypts, bound by a monolayer epithelial cell. The primary signs of cancer are aberrant crypt foci; these growths impact only a few crypts. Abnormal and disordered growth or dysplasiais the next stage in tumourprogression. Dysplastic aberrant crypt foci may evolve as polyp – benign tumour masses that protrude against the epithelium. Some polyps maintain normal cell architecture and morphology, whereas others have abnormalities in inter- and intra-cellular organisation.Adenomatous polyps or adenomas are the abnormal .Majority people aged 70 or above will have generated at least one colorectal adenoma, but it will usually be asymptomatic. The following stage of cancer progression is when adenomas develop into adenocarcinomas, aggressive tumours. Metastasis is the process by which cancer cells break away from a tumour and expand around the body, normally in the bloodstream, to cause secondary tumours away. Itis generally diagnosed when adenocarcinomas have expanded, but there is excellent prognosis for cases captured before metastasis, surgery and chemo/radiotherapy being quite effective.

However, the presence of multiple genes and the heterogeneity of mutations present challenges to the development of diagnostic tests for this disease.

hMSH2germline mutations were identified at chromosomal positions in HNPCC families.(Fishel et al., 1993) The discoveries made in the field of MMR linked to MSI and HNPCC could bring clinical relevance with respect to research area.

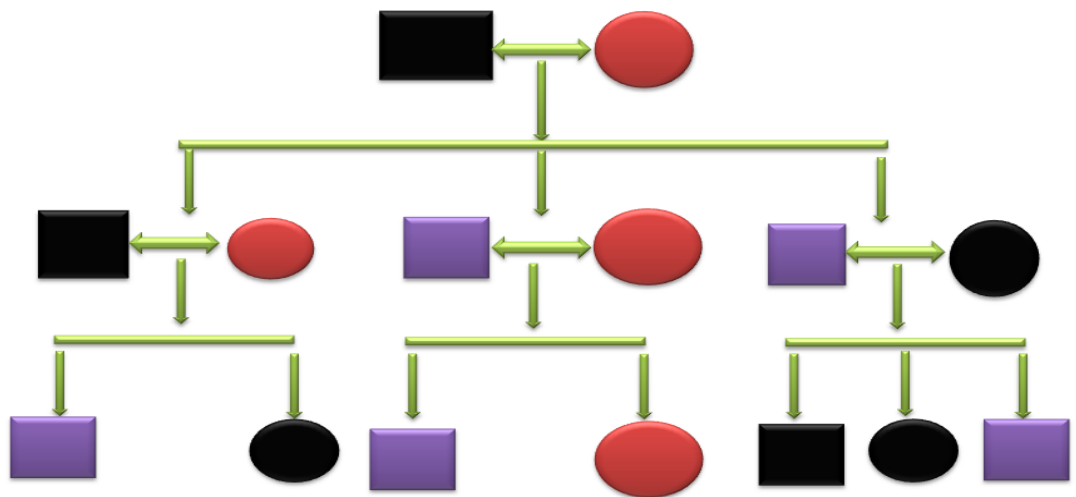
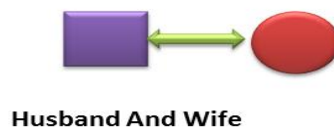


Figure 1.2(a): HNPCC is an autosomal dominant disease.



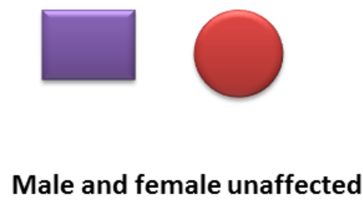


Figure 1.2(b): Symbols in the figure 3(a)

In fig 1.2(a) & 1.2(b), one inherited copy of the mutated gene in each cell is abundant to grow the risk of cancer. It is necessary to know that people inherit an increased cancer

risk, not the disease itself.Not all individuals who acquire mutations in these genes will emergetumour. Anybody can get affected irrespective of their gender.

1.1 Problem Statement

DNA damage is the major reason for the colorectal cancer and other associated malignancies. DNA Repair Malignancies Annotation Platform(DRMAP) focuses primarily on MSI, HNPCC and CRC. DRMAP is a curated and exclusive repository of the above mentioned disorders. The main purpose of it is to bring into focus the genes that are involved in MMR, CRC, MSI, HNPCC. To encourage the improvement and revelation of new analytic and prognostic treatments, and for the characterisation of these tumours, it is important to utilize the scattered information on the above diseases accessible through productions, tests, specialized reports, clinical reports, databases and so on.Keeping in all the existing gaps in knowledge we have designed DRMAP so as to concentrate on the gaps for an information intensive enriched database, which could be of enormous use to the scientific community.

1.2 Objectives

We aim to achieve the followingobjectivesfor the successful fulfilment of the project:

- To give the information related to MMR specific genes related to MSI, CRC and HNPCC.
- To develop one of its kind MMR specific MSI, CRC, HNPCC covering the Interactions, gene ontology i.e., cellular component, molecular functions, biological processes, RNA and protein expression in colon tissues, sub-cellular and chromosomal location of the genes.
- Creating a database that will be updated from time-to-time and will include all the advancements of the aforementioned disorders.
- Study of the expression of these genes in the disease.

1.3 Expected Outcomes

- The premier outcome of the database is the easy repository, retrieval, mining and annotation of the data related to MMR, HNPCC and CRC in a capable and structured manner that may be used in various investigations related to it.
- The data that has been collected can be used to study the insights of the disorders and prove helpful for the discovery of the drugs, derived via understanding the gene ontology of various genes involved in the process.
- The expression of the genes in the colon tissues can prove useful for the study of the data collected during the annotation.
- All the data can act as a support for the storage and establishment of other computational databases, so as to change the data appropriate for meaningful analysis which can further be deployed in large scale projects. The data will be updated regularly to include the latest information in it.

CHAPTER-2

MATERIALS AND METHODS

2.1 Data Analysis

This chapter incorporates the Materials and methods of the project where all the computational and annotational steps are beingdiscussed. The extensive study and understanding of the MMR mechanism and its relation to HNPCC, CRC and MSI was done by reviewing various literatures through PubMed (Iyer et al. 2006., Li , 2008, Kunkel et al. 2015).

The tools and databases were determined in compliance with the type of data that is related to the MMR, HNPCC, CRC and MSI and can be used for the clinical trials in future analysis (figure 2.1). After completion of the starting procedure, we formulated a methodology where we devised a work flow that explained all the sequential steps required in the development of the DRMAP (DNA Repair Malignancies Annotation Platform).We pooled all the genes related to the topic of study along with their description, structure and function from Uniprot (Magraneet al., 2010)and PDB (FUJII et al., 1996) respectively. They give the deep insights and aid in recognizing the role of MMR and MSI pathways in HNPCC and CRC. The orthologs were found that assisted us in knowing the families of the genes and the evolutionary process and learning about the genes with same biological functions. The domains were recognised from NCBI to have knowledge about the structural and functional role of the proteins and their involvement in the MMR, HNPCC, CRC and MSI was also distinguished. The sub-cellular and chromosomal location was derived from Uniprot and Ensemble(Varadi et al., 2015)

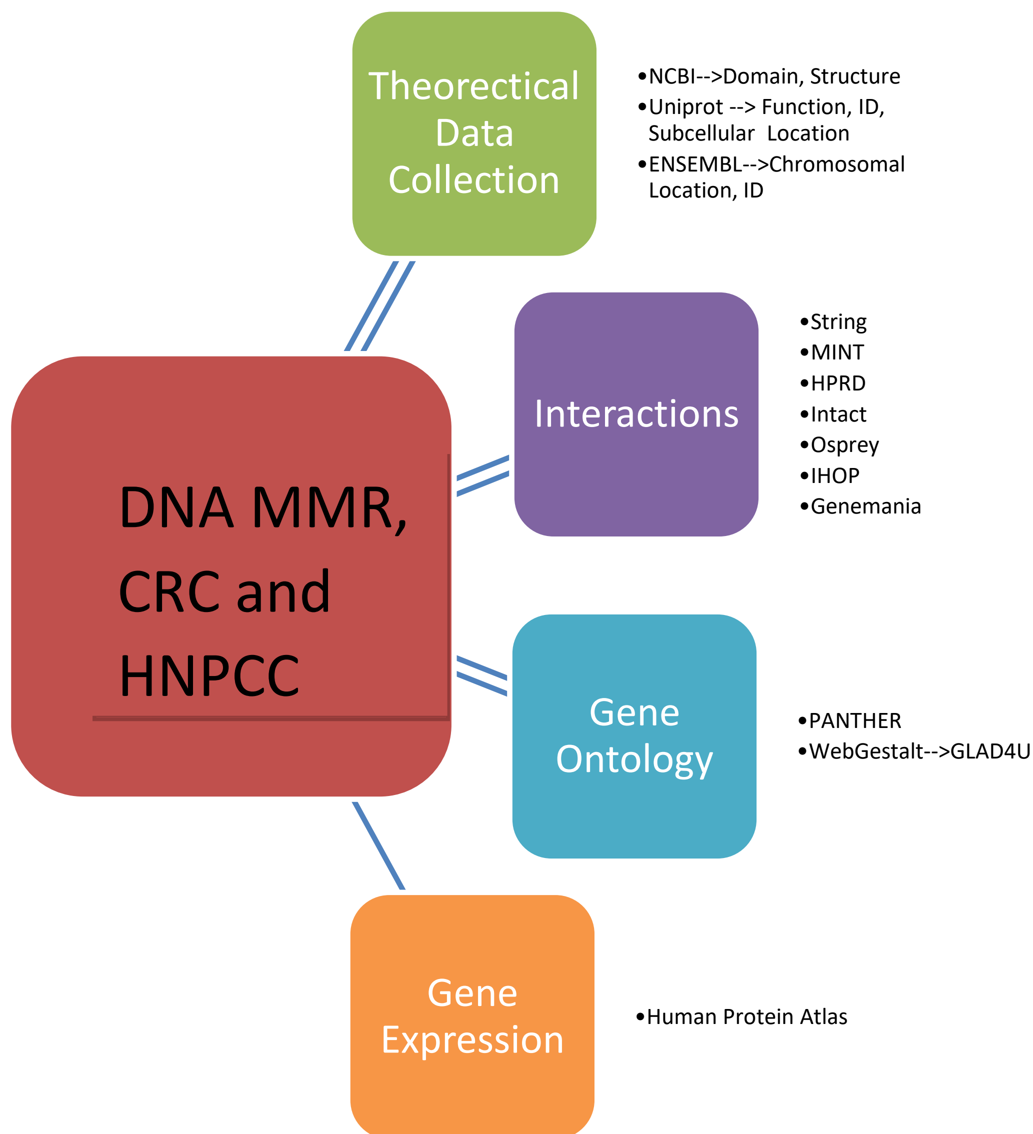


Figure 2.1: Various tools and databases being utilized in data collection and analysis

Considering how the genes interacted with each other and with other genes helps to discover those proteins which are directly or indirectly involved in the cancer causing pathways along with type of interaction that occur among them. To analyse the type of interaction and associations between the genes we used STRING (Search Tool for the retrieval of Interacting Genes/Proteins)(Szklarczyk et al., 2014), IHOP (Information

Hyperlinked over Proteins)(Hoffmann et al., 2004), MINT (Molecular INTeraction database)(Ceol et al., 2009), HPRD (Human Protein Reference Database), Genemania (Montojo et al., 2014), Osprey (Breitkreutz et al., 2003) and IntAct(Hermjakob et al., 2004)databases and tools. To analyse the protein-protein interactions for every gene we used the STRING database that displayed the results on the basis of experimental evidence, co- occurrence and co -expression with 0.400 as the minimum interaction score and 1 as the maximum score. Ingenemania, we analysed the interactions which our set of genes had with each other along with other genes keeping the search specific to their physical interaction and co-occurrence. In IHOP we extracted the data for each gene using different approaches like two hybrid, experimentally verified, not experimentally verified, coimmunoprecipitation, pull down etc. It shows the interacting proteins and the evidences in the results. Applying intact, we drew the interactions with two hybrid approaches to find the kind of association a particular gene is displaying with the other genes. HPRD displays the results of the interacting proteins, evidence (*in vitro, in vivo* and yeast two hybrid) and the kind of interactions (direct, complex) that a gene have with other genes and proteins. MINT emphasises on the experimentally verified and curated protein-protein interactions through literatures and displays the interaction, interaction type and the detection methods which have been used. Osprey displays the results on the experimental basis for the interactions between a set of genes and other genes (figure 2.2).

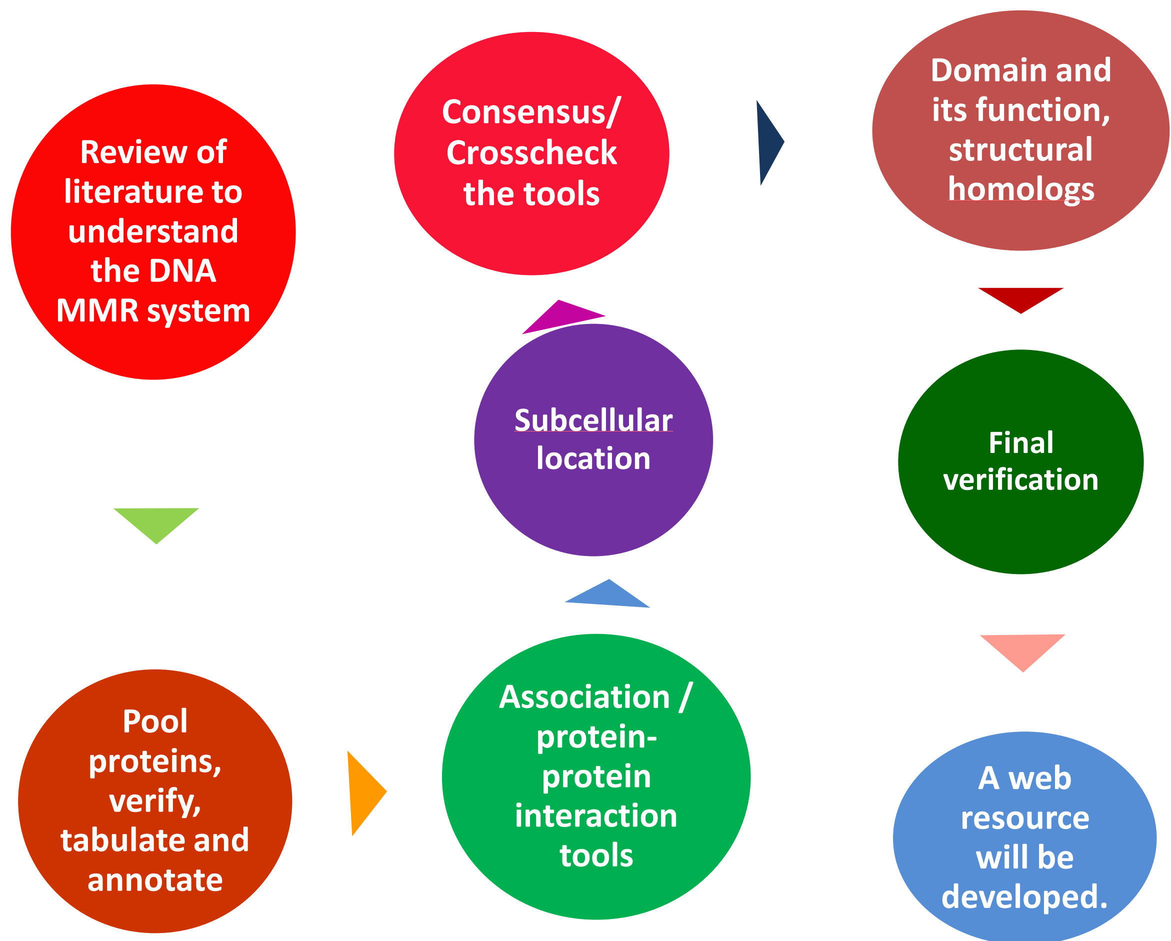


Figure 2.2: Systematic representation of steps used to analyze the data

After obtaining the results from the aforementioned steps, we downloaded the microarray data of MMR involved in HNPCC (Hereditary Non Polyposis Colon Cancer) to compare the common/significant genes from the previous results that are involved in the cancer. We used WB-DEGS (Within and Between Group Comparisons for Differentially Expressed Gene Selection) (Unpublished work) (Moussa et al., 2012) to preprocess, visualize, and select genes with exactness to limit the false positive rate. It also includes some classical gene selection methods (*see figure 2.3*). We uploaded the selected .CEL files from GEO (Gene Expression Omnibus) for series GSE24514 "Candidate driver genes in microsatellite - unstable colorectal cancer", and performed preprocessing of the data (Alhopuro et al., 2012). Then we performed statistical analysis and mapped the overexpressed and underexpressed genes.

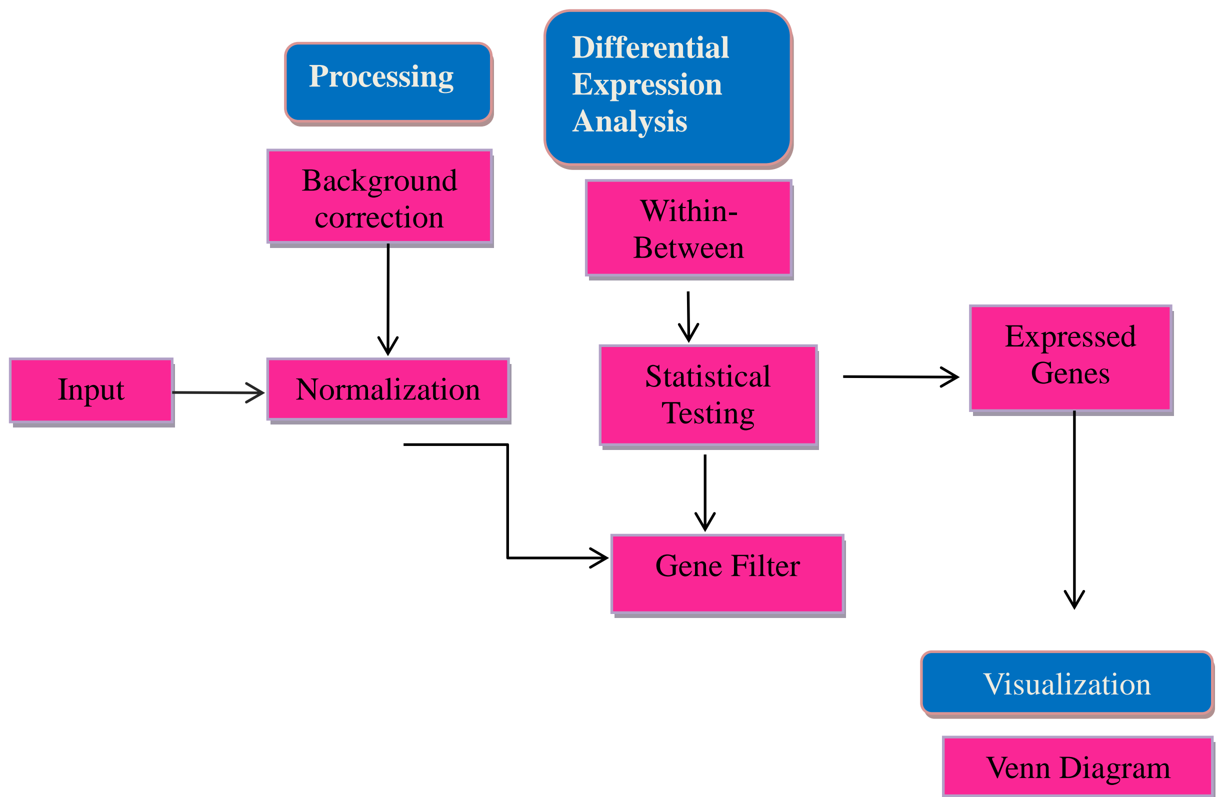


Figure 2.3: Workflow of WB-DEGS

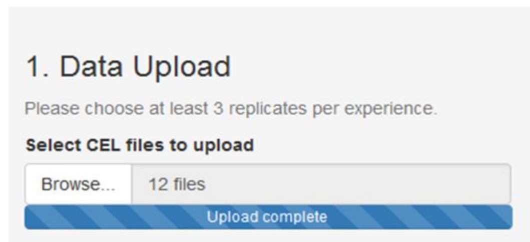


Figure 2.4(a)

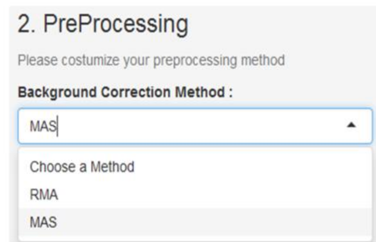


Figure 2.4(c)

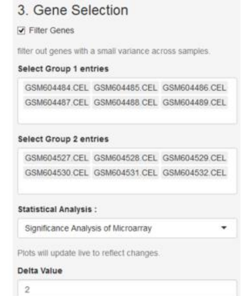


Figure 2.4 (e)

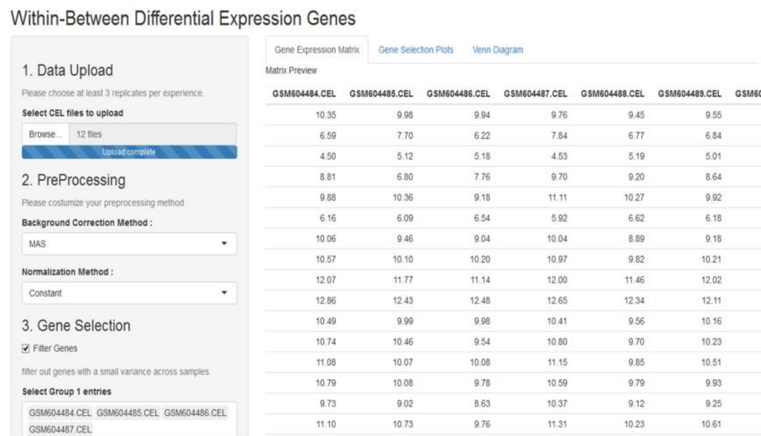


Figure 2.4(b)

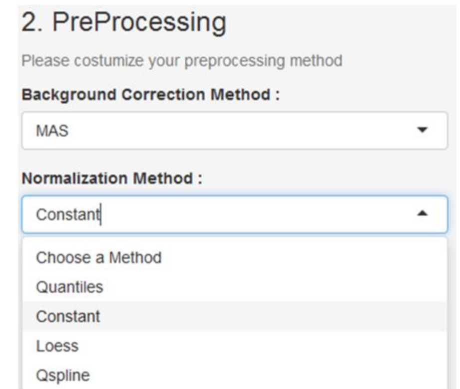


Figure 2.4 (d)

Figure 2.4(a): Uploaded Files,**2.4(b)**: Gene Expression matrix is obtained when the files are uploaded, **2.4(c)**: Selection of Background Correction, **2.4(d)**:Background Correction Method and Normalisation method and **2.4(e)**: Two groups are added as diseased and normal

We uploaded the selected files of the series in Data Upload option and then used RMA (Robust Multiarray Averaging) and MAS (Affymetrix Microarray Suite)as a background correction method with quantiles as normalisation method. After preprocessing of data, we divided the samples into two groups: a) Group 1: test group, b) Group 2: control group based on the curated sample data from the given study (Alhopuro et al., 2011). In the final step, we applied statistical analysis for the estimation of local and global FDR and mapped the overexpressed and underexpressed genes by Significance Analysis of Microarray (SAM).

After mapping all the intersected probe ids in the Venn diagram, to their respective gene ids using DAVID (Database for annotation, visualisation and integrated discovery)(Huang et al., 2007), we performed the comparative interactions using

aforesaid tools. We found the consensus of the interactions at the end for the final annotation purpose (Figure 2.5).

The gene ontology consists of Biological Process, Cellular Component, and Molecular Function which tellspathways and processes that is an aftereffect of the activities of multiple gene products, where gene products are active and molecular activities of gene products for the involved genes. PANTHER (Protein Analysis Through Evolutionary Relationships) (Mi et al., 2009)and the WebGestalt(WEB-based GEneSeTAnaLysis Toolkit) (J. Wang et al., 2013) realised the GO i.e., the gene function.

Figure 2.5: After retrieving the WB-DEGS data, these steps are followed.

More sets ofVenn diagram were examined and 2119 genes were mapped against their gene ids. Their gene ontology and interactions was also found in the similar way as previously done.

The genes that were common from the different datasets and weresignificantly involved in the MMR, HNPCC and CRC were drawn out. The important genes that were obtained from the former analysis, were also incorporated. Their overrepresentation Enrichment Analysis was carried out by utilizing diseases as functional database. The GLAD4U (Gene list automatically derived for you) (Jourquin et al., 2012) as functional database name gave the genes that were linked with MMR, HNPCC or CRC. Both set of genes were manually checked for their involvement in disease and their regulatory processes.

The protein and RNA expression of the various genes in rectum was procured through Human Protein Atlas(Pontén et al., 2008). The three methods HPA, FANTOM and GTEx were employed to get the RNA expression. The strategies ascertained in light of the gene expression levels over all tissues and incorporate tissue improved, amass advanced, tissue upgraded, communicated on the whole, blended and not recognized. The methods ascertained based on the gene expression levels over all tissues and include tissue enriched, group advanced, tissue enhanced, expressed in all, mixed and not recognised. The expressions of the genes were studied from the literature in HNPCC and CRCand various conclusions were pulled out. The files obtained after normalisation were used to fetch the P-value and FC (Fold Change) value which is one of the parameter to determine the significant genes(Figure 2.6).

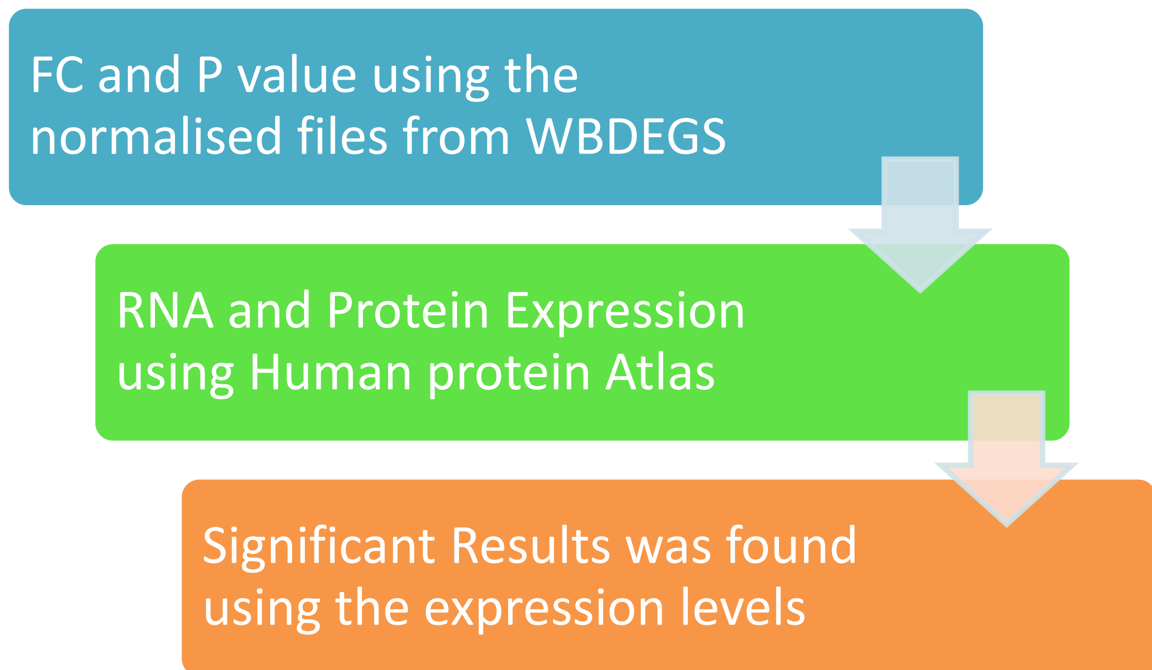


Figure 2.6: Quantitative Analysisfor the set of genes to know their role in HNPCC and CRC

2.2 Database and GUI development, connectivity and data retrieval

In this step, we have designed an outline of the database using PonyORM (Object Relational Mapper) where the structure of the database has been formed (Figure2.7).

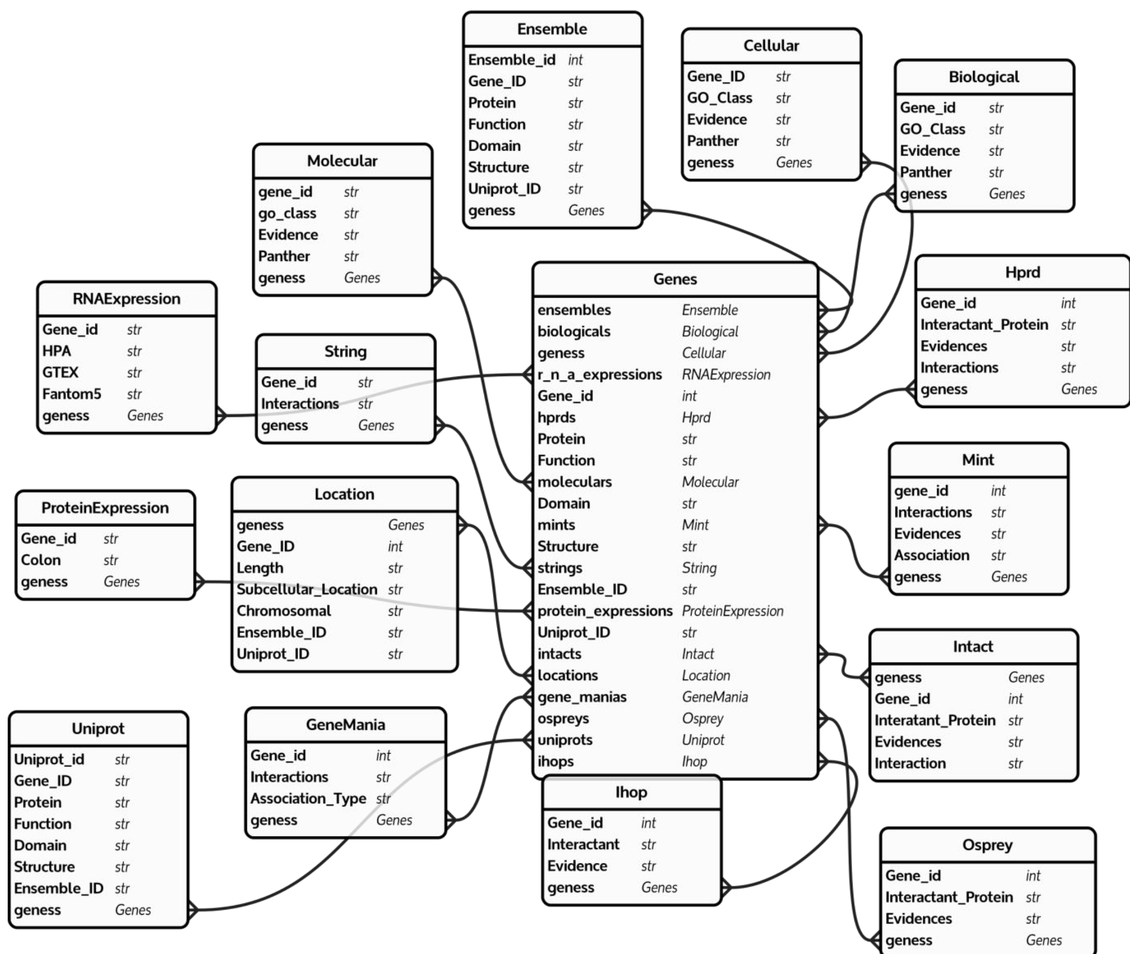


Figure 2.7: ER diagram using PonyORM

The data has been gathered from different databases and based upon various parameters 14 tables were designed.

GUI(Graphical User Interface) is in the form of DRMAP website used with the XAMPP server 3.2.2(Apache server and MySQL) to store the data. HTML (Hyper Text Markup Language), CSS(Cascading Style Sheet), JavaScript and PHP (Hypertext Preprocessor) are being utilized cumulatively for the final disposition of GUI as websie and its connectivity with the normalised database in the background.

HTML is used to create the web pages and the interfaces for applications. It defines the structure semantically and we can insert scripts written in JS. DHTML makes a page more interactive and animations can be added by implementing it. It uses a combination of HTML, CSS and JS.

CSS is a style sheet language deployed for formatting the document. It aids in creating interactive pages and changing the graphic design of a document (Figure 2.8)

```
Body
 {
 font-family: "Open Sans", Arial, sans-serif;  font-weight: 400;  font-size: 16px;  line-height:
    1.7;  color: #777;  background: #fff;
}
#page
 {
position: relative;  overflow-x: hidden;  width: 100%;  height: 100%;  -webkit-transition: 0.5s;
    -o-transition: 0.5s;  transition: 0.5s;
}
.offcanvas
#page
{
overflow: hidden;  position: absolute;
}
.offcanvas #page:after
{
  -webkit-transition: 2s;  -o-transition: 2s;  transition: 2s;  position: absolute;  top: 0;
    right: 0;  bottom: 0;  left: 0;  z-index: 101;  background: rgba(0, 0, 0, 0.7);  content: "";
}
a {
 color: #66D37E;  -webkit-transition: 0.5s;  -o-transition: 0.5s;  transition: 0.5s;
}
a:hover, a:active, a:focus
 {
 color: #66D37E;  outline: none;  text-decoration: none;
}
P
 {
margin-bottom: 20px;
 }
```

Figure 2.8: CSS implemented for background

JS is an object and prototype-based, interpreted language also known as scripting language for Web pages.

PHP is a server-side scripting language which helps to develop web. (Figure 2.9)

The various tables that have been created were uploaded on the server and stored in the database. The connection was set and results were then retrieved using PHP through various search options provided to the user on the website.

Database along with the website is available for academic and research purpose at: http://www.bioinfoindia.org/drmap.

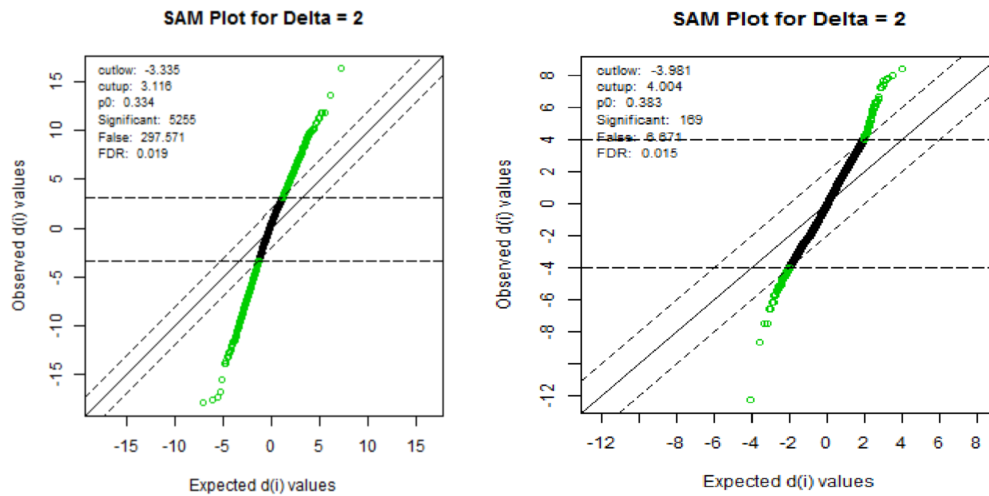
```php
<?php
$servername = "localhost";
$username = "root";
$password = "";
$dbname = "drmap";
$conn = mysqli_connect($servername, $username, $password, $dbname);
if (!$conn) {
    die("Connection failed: " . mysqli_connect_error());
}
$key=$_GET['user'];
echo "<table>";
echo "<tr>";
echo "<th align=center>Gene_ID</th>";
echo "<th align=center>Interactant_Protein</th>";
echo "<th align=center>Evidences</th>";
echo "</tr>";
if($key=='All')
{
     $sql="SELECT * FROM osprey";
}
else
{
     $sql = "SELECT * FROM osprey WHERE Gene_ID='$key'";
}
$count=0;
$result = mysqli_query($conn, $sql);

if (mysqli_num_rows($result) > 0)
{
    while($row = mysqli_fetch_assoc($result)) {
             $count=$count+1;
             echo "<tr>";
             echo "<td>". $row["Gene_ID"]. "</td>";
             echo "<td>". $row["Interactant_Protein"]. "</td>";
             echo "<td>". $row["Evidences"]. "</td>";
             echo "</tr>";
}} else {
    echo "0 results";
}
echo "</table>";
?>
```

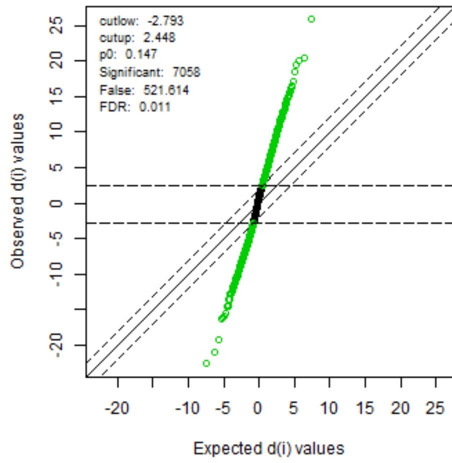
Figure 2.9: PHP deployed to set up the connection

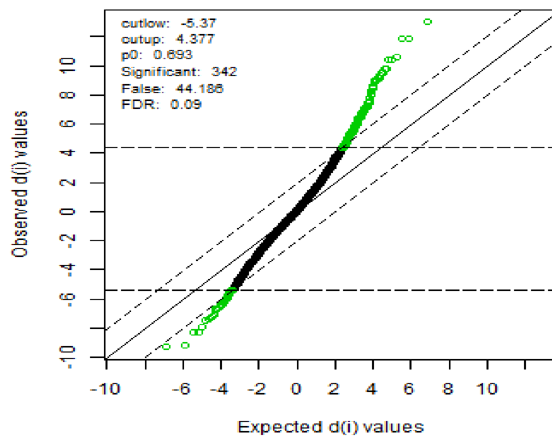
Chapter-3

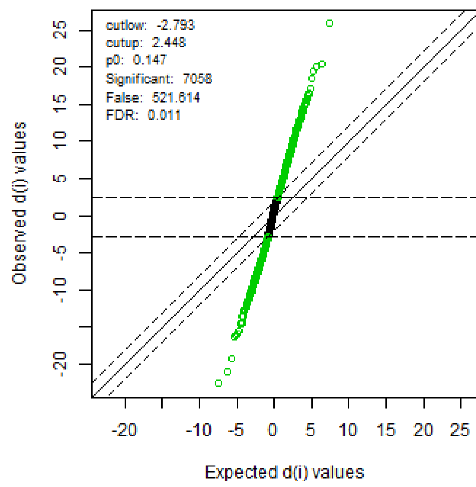
Results and Discussions

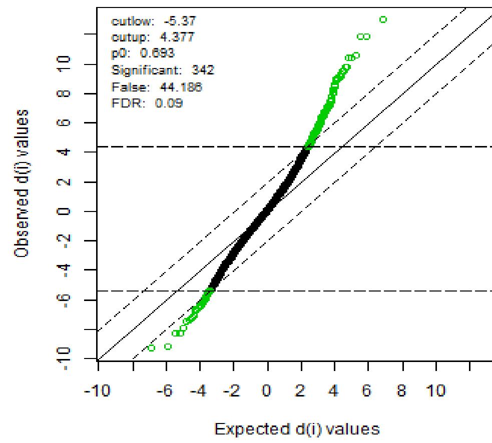
Significant Analysis of Microarray (SAM) is a technique for deciding if changes in gene expression are measurably critical.It identifies genes with statistically convincingchanges in expression using particular t-tests and figure a statistic dj for each gene j which evaluates strength relationship between gene expression and a response variable. Several plots were generated for SAM using different values of delta but significant results were obtained when the value of delta was 2. The thresholds were +4 and -4. The genes that were above +4 are termed as positive gene set i.e, genes have higher expression and genes below -4 have lower gene expression. It can be seen in the *figure 15(a)*. The p-value is decided by the False Discovery Rate (FDR). Similar SAM plots were acquired from the analysis (figure 15) which gave us many probe ids which were converted to Gene ids using DAVID.




SAM Plot for Delta = 2
cutlow: -2.793
cutup: 2.448
p0: 0.147
Significant: 7058
False: 521.614
FDR: 0.011
Observed d(i) values
Expected d(i) values


SAM Plot for Delta = 2
cutlow: -5.37
cutup: 4.377
p0: 0.693
Significant: 342
False: 44.186
FDR: 0.09
Observed d(i) values
Expected d(i) values


SAM Plot for Delta = 2
cutlow: -2.793
cutup: 2.448
p0: 0.147
Significant: 7058
False: 521.614
FDR: 0.011
Observed d(i) values
Expected d(i) values


SAM Plot for Delta = 2
cutlow: -5.37
cutup: 4.377
p0: 0.693
Significant: 342
False: 44.186
FDR: 0.09
Observed d(i) values
Expected d(i) values

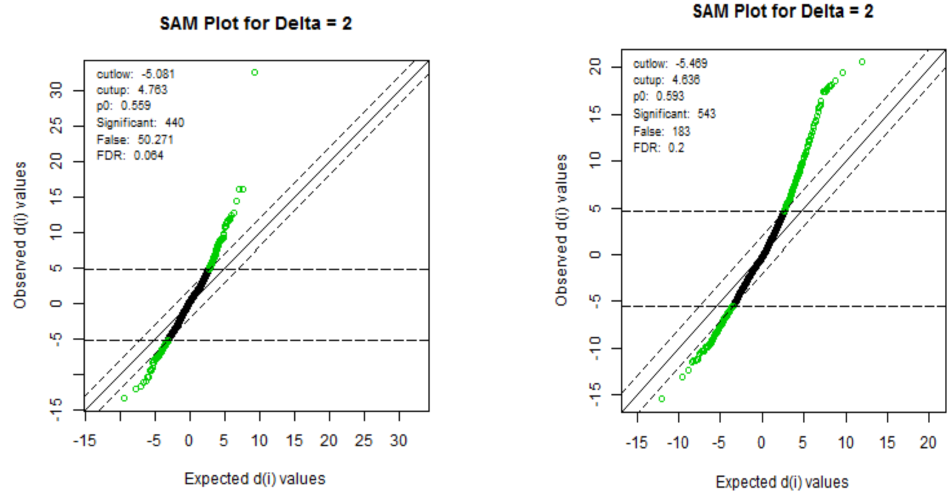
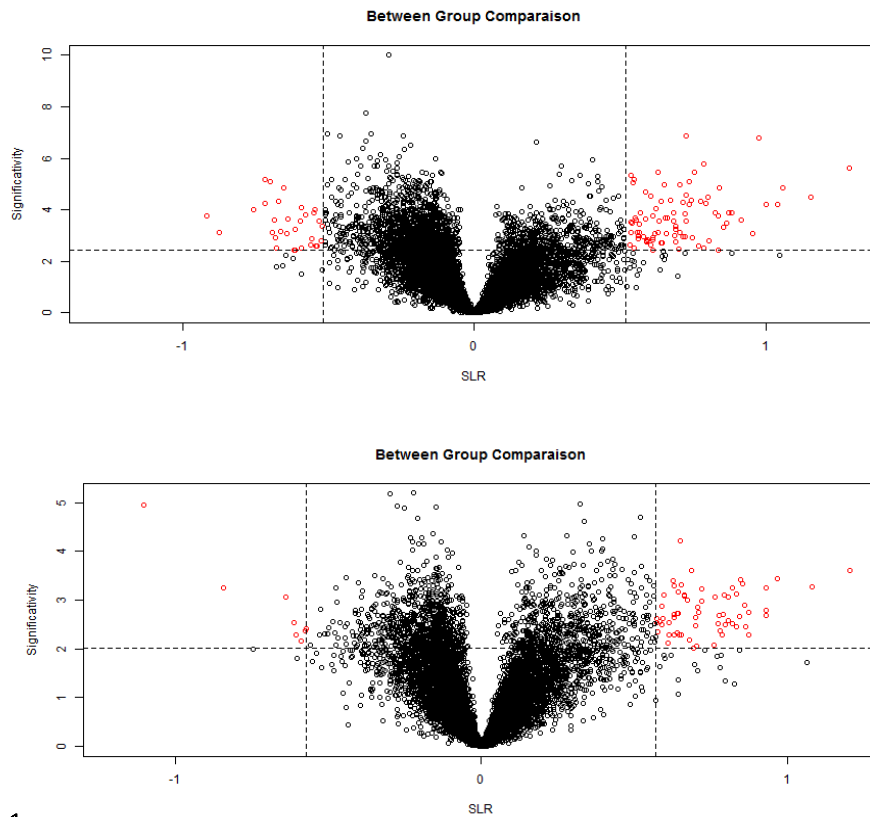
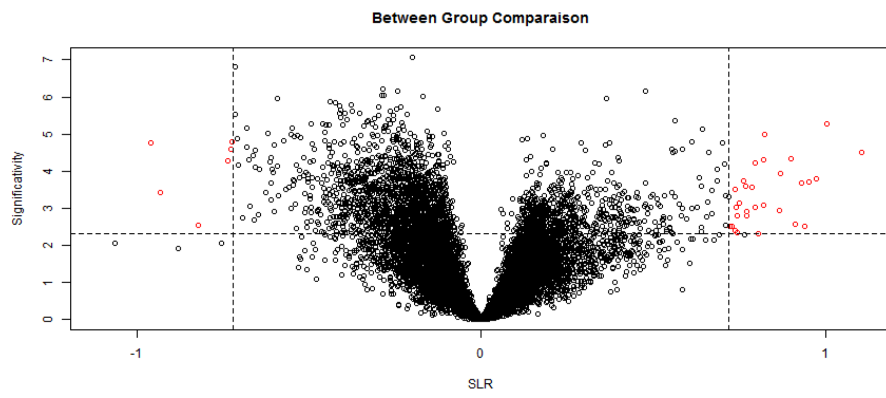
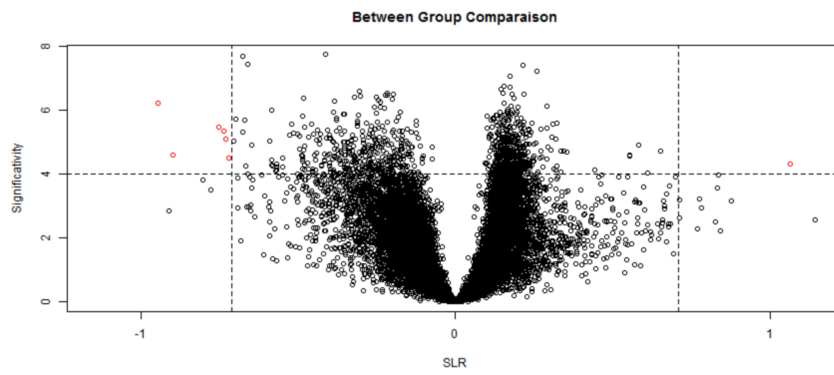


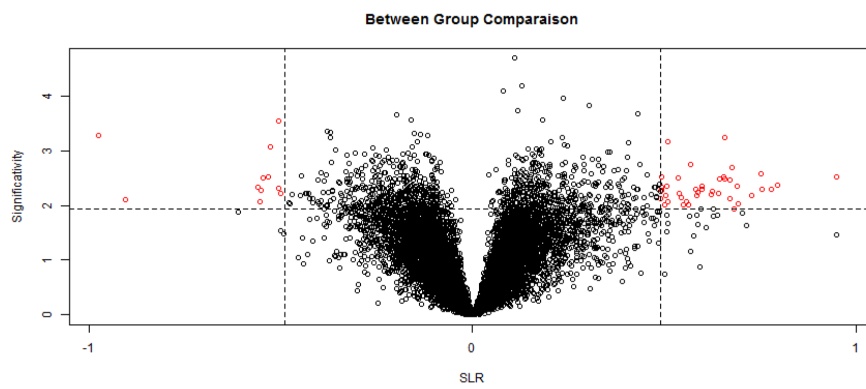
Figure 3.1: SAM plot for expected and obtained dataset for this respective expression levels. Overexpressed and underexpressed genes were selected through respective threshold values.

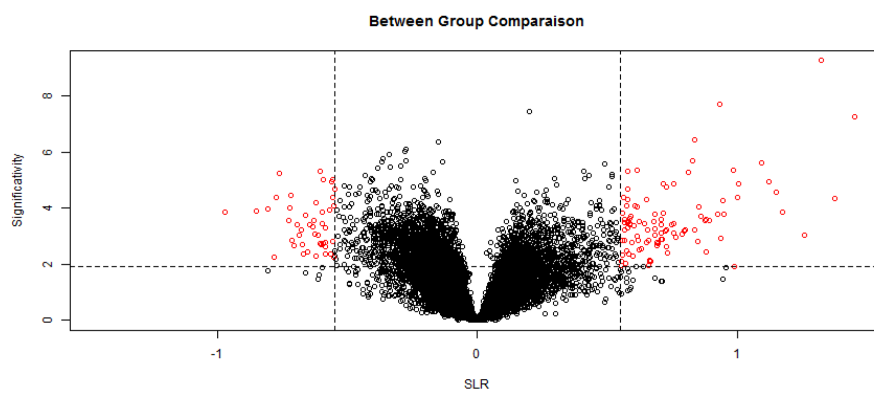


1


Between Group Comparaison
Significativity
SLR
-1
0
1


Between Group Comparaison
Significativity
SLR
-1
0
1


Between Group Comparaison
Significativity
SLR
-1
0
1


Between Group Comparaison
Significativity
SLR
-1
0
1

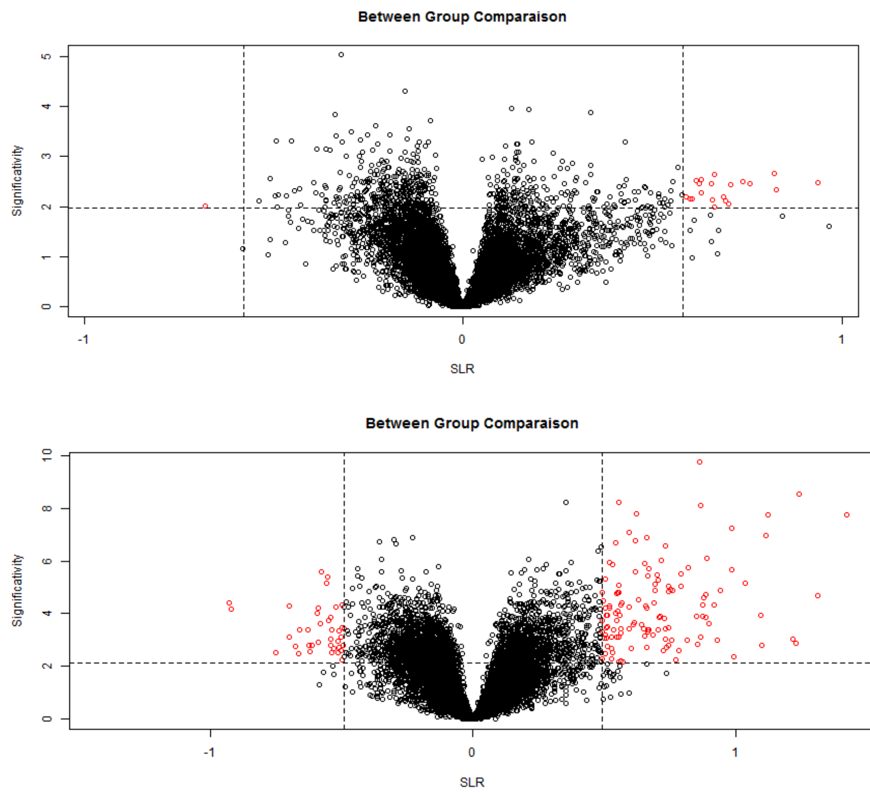
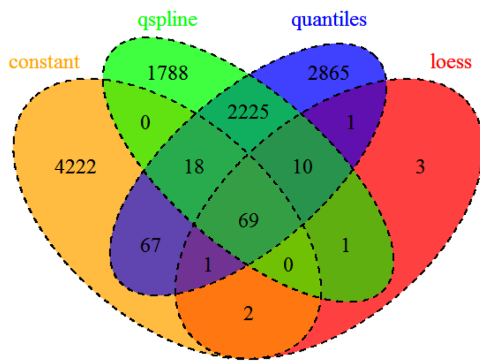
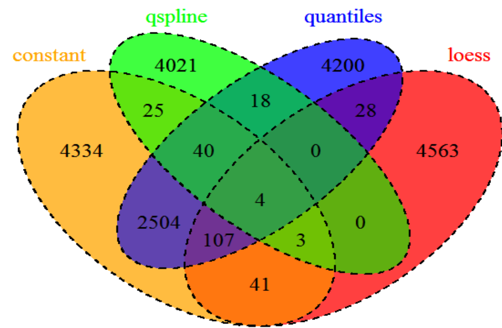
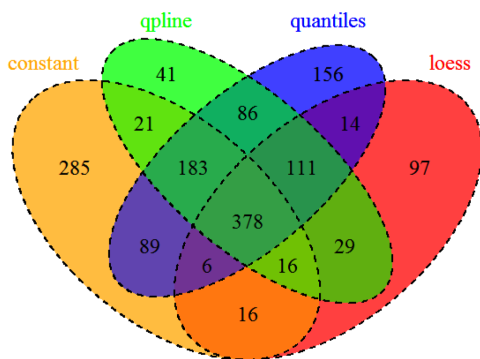


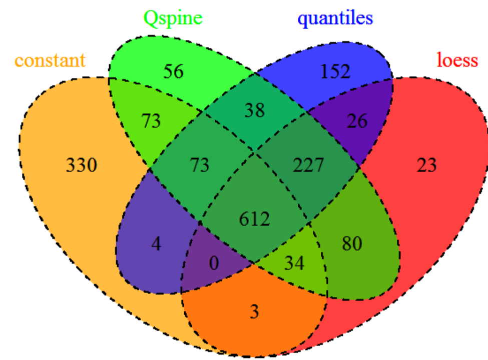
Figure 3.2: Volcano Plot for between group comparisons for significance (y axis) versus fold change(x axis)

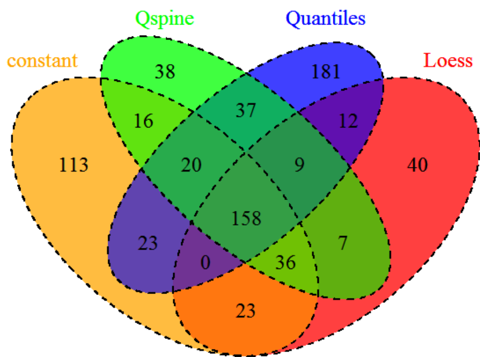
Volcano plot is a plot between the Signal Log Ratios and Significance. It was used to measure the background noise amongst the genes. It was being plotted for within the group and between group comparisons, but more differentiated results were obtained in the between group comparisons. It gave the over expressed and under expressed genes based on the thresholds. Alike outcomes came out which are displayed in the figure 3.2.

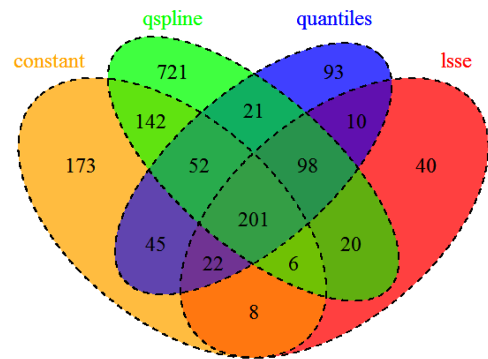


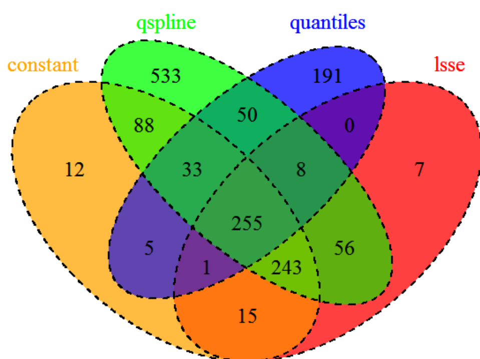












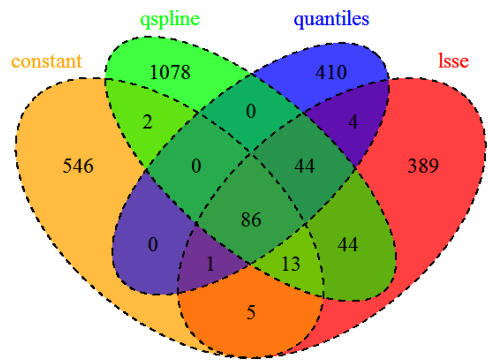


Figure 3.3: VennDiagramgenerated after applying the normalisation methods

Venn diagrams were obtained when the four normalisation techniques - constant, qspline, quantiles and loess was applied and 2119 common genes were attained as show in *figure 3.3*. The genes that we got are linked to MMR system.

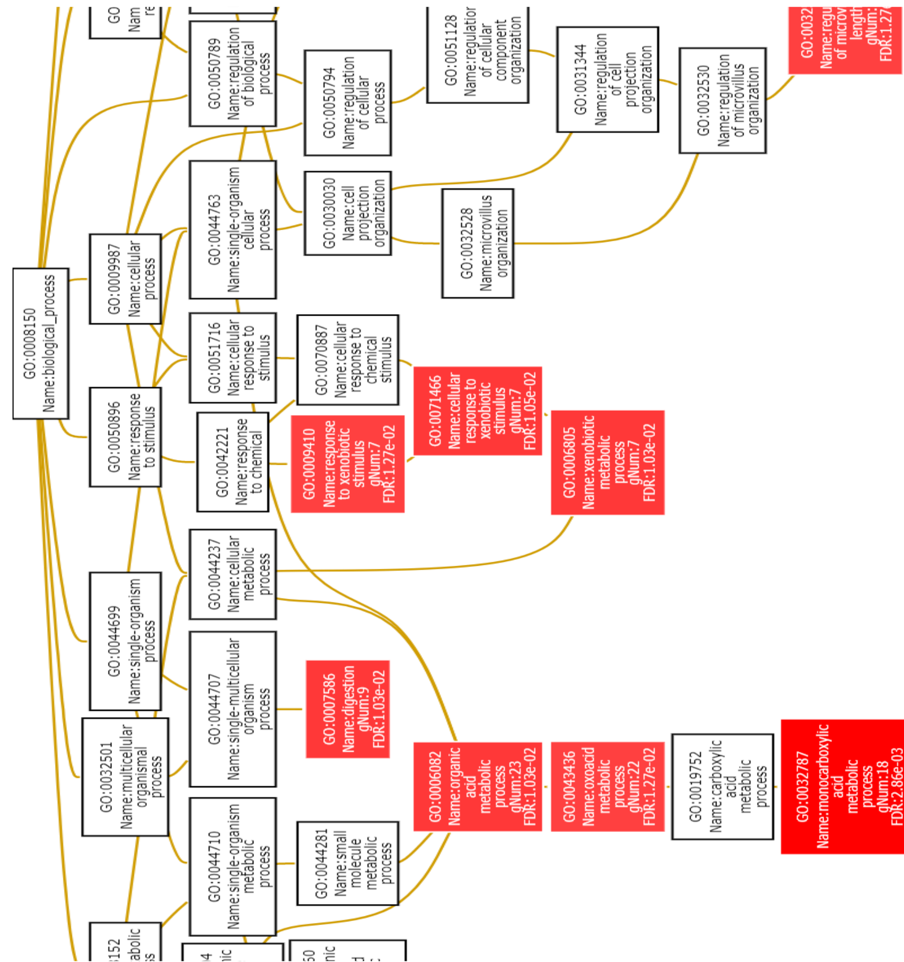


Figure 3.4: Annotation tree obtained from the WebGestalt for various regulatory biological processes

The gene ontology (GO), applied in WebGestalt gave us many trees that showed different processes where genes are involved. Genes that are associated with MMR and HNPCC from the tree were identified.

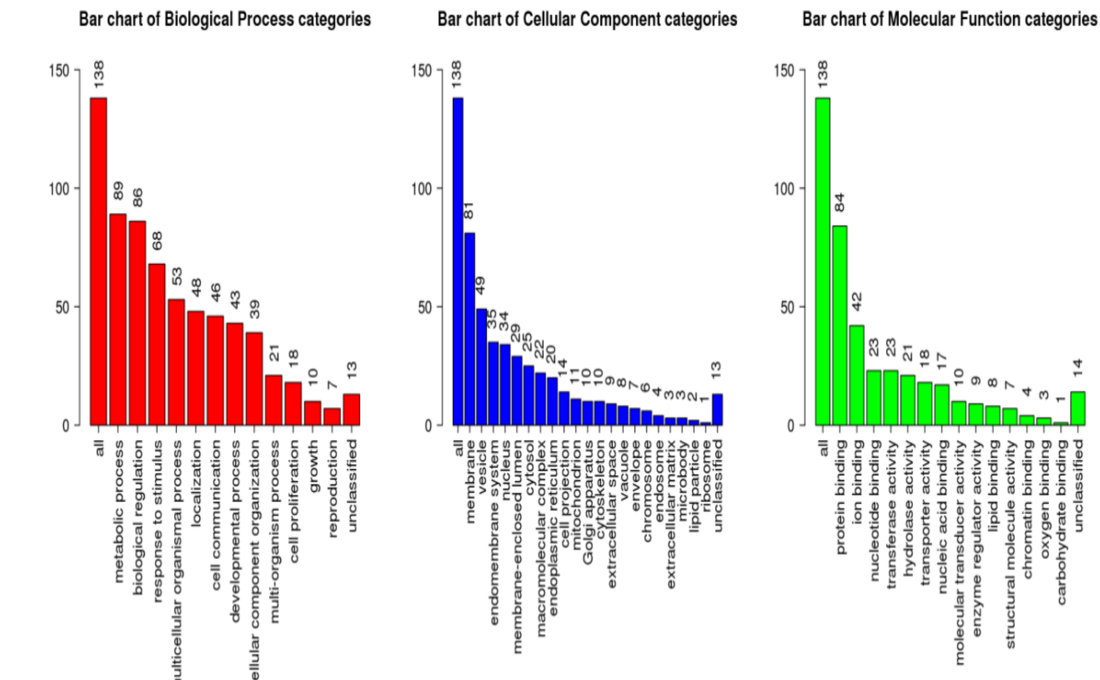


Figure 3.5: Graphs for the GO for three parameters: P, C and F.

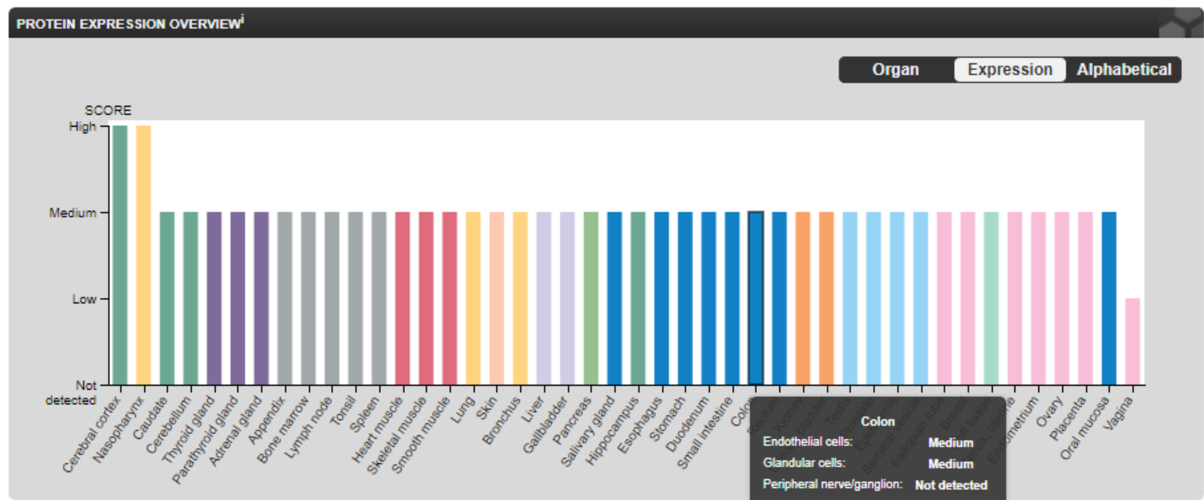
This graph shows the cellular component (C) that include the cell or its parts, the biological process (P) are the molecular events relevant to functioning of the integrated living units and molecular functions (F) are the elemental activities of a gene product at the molecular level. All the genes taken are displayed in the graph with their location, function and process in the cell.

Table 3.1: Functional database GLAD4U applied to get the results related to HNPCC and CRC.

ID	Name	#Gene	FDR
PA443754	Colonic Diseases	14	2.78e-05
PA444632	Intestinal Diseases	15	2.78e-05
PA443756	Colonic Neoplasms	11	7.1e-05
PA443899	Diarrhea	7	1.07e-03
PA443265	Adenocarcinoma	13	1.45e-03
PA444256	Gastrointestinal Diseases	14	1.45e-03
PA446108	Colorectal Neoplasms	11	2.5e-03
PA443749	Colitis	9	3.79e-03
PA165108442	Neoplasm of unspecified nature of digestive system	14	3.79e-03
PA444154	Fatty Liver	8	6.02e-03

When the diseases are chosen as the functional database and GLAD4U as the name of the database, we get the results as shown in Table 3.1. We found many genes in the database and compared it with the previous data. We found crucial results which had repetitive appearances of the genes.

The protein and RNA expression levels were investigated using HPA where there was a great variation in the appearance ofthe genes in the colon tissues (Figure3.6). Some important results were drawn that showed a crucial level in the disease. The genes are being displayed in the table which have a variation in their expression. In the figure showing the RNA expression in colon cells using HPA is 13.6 Tags per Million (TPM) in which the gene length is normalised first and then depth sequencing. In the other figure of GTEx, the reading is 3 Read Kilo perMillion (RPKM) that was made for single RNA seq. FANTOM5 shows a value of 9.4 TPM.The Protein and RNAexpression of genes in colon tissues have been incorporated in the table 3.2 and 3.3.



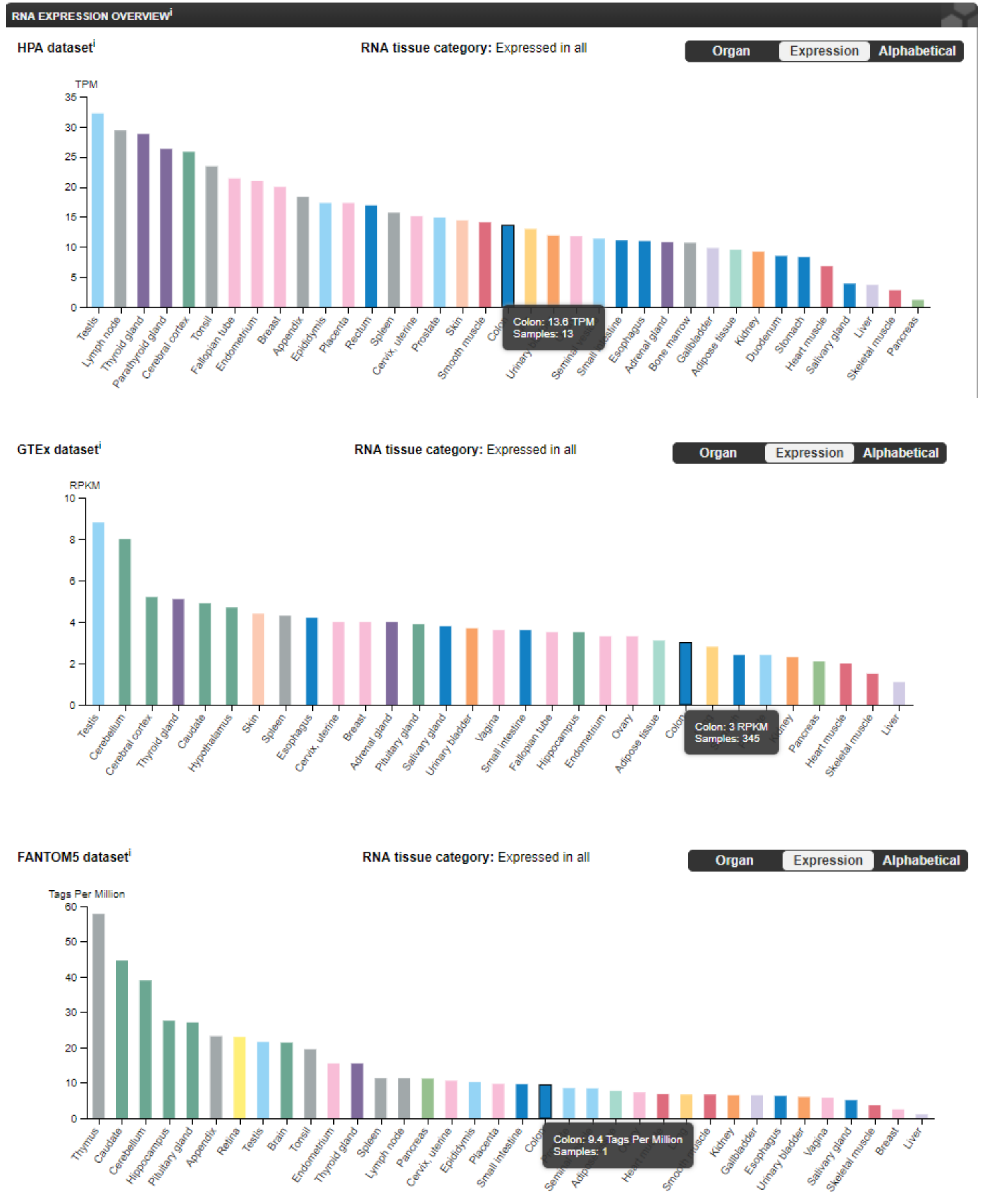


Figure 3.6: Protein and RNA expression of Colon cells in MSH2 using HPA, GTEx and FANTOM5

Table 3.2 :RNA expression in Colon cells of genes.

Gene	HPA(TPM)	GTEx(RPKM)	FANTOM5(Tags per million)

XRCC1	11.9	10.1	20.7
LGR5	3.4	0.6	12.9
SELE	1.2	2.6	2.2
MTHFR	9.4	4.5	23.2
HMGCR	46.2	12	28.3
WNT5B	6.1	5.7	6.8
NR1I2	10.3	4.6	43.8
ABCC2	0.3	0.3	28.7
KRAS	28.6	7.5	34.8
EGF	0.8	0.2	0
FCGR3A	12.3	3.8	0.7
UGT1A6	1.8	0	1
ABCC1	3	9.6	19
ABCG1	13.8	5.1	37.9
PTGER4	17.5	7.6	63
TYMS	4	0.5	7.9
EGFR	12.5	7	35.2
ABCB1	17.1	4.1	121.9
ABCC5	17.7	8.1	10.4
GSTP1	221.5	161	347.5
ABCG2	19.7	2.8	51.8
ALOX12	0.4	0.5	0.1
IL23R	0.8	0	0.5
HLA-G	0	0.5	0.9
UGT1A7	0.1	0	0
DPYD	9.6	4.8	96.8
SLC29A1	21.5	44.1	102.6
SLCO1B1	0	0	0
SHMT1	20.9	8.6	62.9
PTGES	4.6	3.1	18.9
AREG	18.4	1.2	17.6
MGAT4A	24.2	4.5	43.9
PARD3B	4.4	4.9	23.3
UGT1A9	0.4	NA	0
ENOSF1	11.4	11.4	28
CXCR4	31.7	10.9	63.9
ADCY2	0.7	1.4	13.2
TGFB1	11.9	19.2	80.1
RGS5	52.1	41.5	177.7
KLC1	33.7	24.9	52.1
VEGFA	26.7	25	127.3
TGFBR2	61.9	60.3	367.6
XRCC3	7.4	4.7	1.6
MSH2	13.6	3	9.4
BRAF	6	4.2	38.7
PTEN	31.3	10.9	275.7
GSTM1	10.6	10.1	30.9

MDM2	22.8	4.7	57.2
APC	6.7	3.3	35.2
MSH6	13.6	4.2	31.9
HRAS	5.3	14.3	17.9
MUTYH	5.6	5.7	10.9
DNMT3B	0.7	0.5	0.8
CA9	0.2	0.8	18
FBXW7	9.8	3.4	21.5
PMS2	4.9	2.8	5.1
MSH3	6.4	2.2	33.9
MLH2	7.5	3.2	9.9
PCNA	81.5	24.7	33
RPA1	20.8	13.1	43
H2AX	10.3	12	73.3
GTF3C3	15.6	4.4	27.7
TAF15	29.9	0	135.1
ATF7IP	11.2	3.4	46.7
MYST2	14.9	9.9	76.8
HDAC2	47.5	5.9	55.6
HDAC3	22.3	21.4	4.8
HDAC5	10.6	13.7	45.1
HDAC10	3.8	8	21.1
SUV39H2	3	1.1	9.7
JARID2	4.4	3.1	44.8
BMI1	25.7	15.8	39.5
MGMT	6	5.6	25.6
NTHL1	8.2	9.5	11.9
OGG1	9.5	5.1	35.3
ERCC3	16.6	17.7	24.3
ERCC4	2	1.1	9.7
RAD50	6.4	6.5	39.8
XRCC5	89	46.3	251.8
SMUG1	20.3	4.4	1.9
XRCC4	5.9	1.6	7.8
YB1	450.1	238.5	61.1
ATR	4.6	2.5	15.5
PINK1	27	22.3	52.1
POLG	10.4	136	30.1
RAD21	71.3	30.5	102.5
RFC1	14.2	7.7	54.8
RFC3	8.5	2.2	7.5
MERTK	8.5	6.7	90.5
NBS1	17.7	8.9	38.7
BRCA1	6.9	0.9	11.6
EPCAM	659.9	140.8	1116.8
CD44	86.7	19.1	205.2
CA4	253.5	24.2	18.4

ACACB	10.2	18.3	130.3
DST	41.2	16.5	109
ENO1	368	208	436.1
FHL1	175.6	168.1	118.2
GUCA2A	768.8	141	235.9
HNRNPL	66.3	65.9	171.7
MYH11	235.2	1737.7	2191.1
PPIB	304.3	120.8	364.7
SET	158.2	45.5	309.5
SLC26A3	1075	196.3	227.7
SORD	11.6	2.1	37.3
TMEM97	14.5	4.6	22.6
UGT1A10	49.7	8	3.8
TP53	28.6	9.3	30
PMS1	7.5	3.2	9.9
MLH1	19.4	7.6	25.2
CHEK2	8.2	1.8	8.8
RFC3	8.5	2.2	7.5
LIG1	6.6	3.9	15.5
AURKA	10.3	2.1	6.8
CCND1	19.2	9.7	50.1
POLD1	8	5.4	11.3
HMGB1	324.8	37.4	114.9
H2AFX	10.3	12	73.3
ERCC1	24.2	10.5	60.3
ERCC2	5.3	3.8	2.8
SLC19A1	3.1	1.7	14.6
PTGS2	3.8	9.2	9.7

Table 3.3 : protein expression in colon tissues

Gene ID	Endothelial Cells	Glandular Cells	Peripheral nerve/ganglion
XRCC1	Medium	High	Medium
LGR5	Low	Low	Low
SELE	Medium	Not Detected	Not Detected
MTHFR	NA	NA	NA
HMGCR	Medium	Medium	Medium
WNT5B	NA	NA	NA
NR1I2	NA	NA	NA
ABCC2	Not Detected	Medium	Low
KRAS	Low	High	Not Detected
EGF	NA	NA	NA
FCGR3A	Not Detected	Not Detected	Not Detected
UGT1A6	Not Detected	Not Detected	Not Detected
ABCC1	Medium	Not Detected	Medium

ABCG1	Low	Medium	Medium
PTGER4	Medium	High	High
TYMS	Low	Medium	Not Detected
EGFR	NA	Not Detected	NA
ABCB1	Not Detected	Low	Not Detected
ABCC5	Not Detected	Low	Not Detected
GSTP1	Low	Not Detected	Not Detected
ABCG2	Not Detected	Medium	Low
ALOX12	Medium	Medium	Medium
IL23R	Not Detected	Not Detected	NA
HLA-G	Not Detected	Not Detected	Not Detected
UGT1A7	NA	NA	NA
DPYD	Not Detected	Not Detected	Not Detected
SLC29A1	Medium	High	Medium
SLCO1B1	Not Detected	Not Detected	NA
SHMT1	Not Detected	Medium	NA
PTGES	Not Detected	Not Detected	Not Detected
AREG	Not Detected	Not Detected	Not Detected
MGAT4A	Medium	Medium	Not Detected
PARD3B	Medium	Medium	Not Detected
UGT1A9	NA	NA	NA
ENOSF1	Low	Medium	Not Detected
CXCR4	NA	NA	NA
ADCY2	Low	Low	Low
TGFB1	Not Detected	Not Detected	Not Detected
RGS5	NA	NA	NA
Gene	Endothelial Cells	Glandular Cells	Peripheral nerve/ganglion
KLC1	NA	NA	NA
VEGFA	Low	High	Not Detected
TGFBR2	NA	NA	NA
XRCC3	NA	NA	NA
MSH2	Medium	Medium	Not Detected
BRAF	Not Detected	High	High
PTEN	Low	Low	NA
GSTM1	High	High	High
MDM2	High	High	High
APC	Low	Medium	NA
MSH6	Medium	High	Medium
HRAS	Low	High	Not Detected
MUTYH	Low	Medium	Low
DNMT3B	Low	Medium	NA
CA9	NA	NA	NA
FBXW7	High	High	Medium
PMS2	Not Detected	Medium	Medium
MSH3	NA	NA	NA
MLH2	Low	Low	Low

PCNA	Not Detected	High	NA
RPA1	Medium	Medium	Medium
H2AX	Medium	Medium	Medium
GTF3C3	NA	NA	NA
TAF15	High	High	Medium
ATF7IP	Medium	Medium	Medium
MYST2	Medium	Medium	Medium
HDAC2	Medium	High	Medium
HDAC3	Low	Medium	Not Detected
HDAC5	Medium	High	High
HDAC10	Medium	High	Medium
SUV39H2	Not Detected	Not Detected	Not Detected
JARID2	Medium	Medium	Medium
BMI1	Low	Medium	Low
MGMT	High	High	Medium
NTHL1	Medium	Medium	Medium
OGG1	Medium	Medium	Not Detected
ERCC3	High	High	High
ERCC4	Low	Not Detected	Not Detected
RAD50	Medium	High	Medium
XRCC5	High	High	NA
SMUG1	NA	NA	NA
XRCC4	High	High	NA
YB1	Low	High	Low
ATR	NA	NA	NA
PINK1	Not Detected	Medium	Medium
Gene	Endothelial Cells	Glandular Cells	Peripheral nerve/ganglion
POLG	Not detected	Medium	NA
RAD21	High	High	High
RFC1	Medium	Medium	Low
RFC3	Medium	Medium	Low
MERTK	NA	NA	NA
NBS1	Low	Medium	Medium
BRCA1	Low	Medium	Low
EPCAM	Not Detected	High	Not Detected
CD44	Low	Medium	Medium
CA4	Not Detected	Medium	Not Detected
ACACB	Not Detected	Low	Not Detected
DST	Low	Medium	Low
ENO1	Low	Medium	Low
FHL1	Medium	Not Detected	Not Detected
GUCA2A	Not Detected	Low	Not Detected
HNRNPL	High	High	High
MYH11	NA	NA	NA
PPIB	Medium	Medium	Medium
SET	Medium	High	Medium

SLC26A3	Not Detected	Medium	NA
SORD	Not Detected	Low	Not Detected
TMEM97	NA	NA	NA
UGT1A10	NA	NA	NA
TP53	NA	NA	NA
PMS1	Low	Low	Low
MLH1	High	High	NA
CHEK2	NA	High	NA
RFC3	Medium	Medium	Low
LIG1	Medium	High	Medium
AURKA	NA	Low	NA
CCND1	Low	Low	Low
POLD1	Medium	High	Medium
HMGB1	High	High	NA
H2AFX	Medium	Medium	Medium
ERCC1	Medium	Low	High
ERCC2	High	Medium	NA
SLC19A1	NA	NA	NA
PTGS2	NA	NA	NA

The results then gave us the remarkable or common genes from the above analysis and most of them are directly linked in the MMR system (Table 3.4)

Gene ID	Protein	Function
MSH2	DNA mismatch repair protein Msh2	Component of the post-replicative DNA mismatch repair system (MMR).
BRAF	Serine/threonine-protein kinase B-raf	Protein kinase involved in the transduction of mitogenic signals from the cell membrane to the nucleus.
PTEN	Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN	Acts as a dual-specificity protein phosphatase, dephosphorylating tyrosine-, serine- and threonine-phosphorylated proteins.
GSTM1	Glutathione S-transferase Mu 1	Conjugation of reduced glutathione to a wide number of exogenous and endogenous hydrophobic electrophiles.
MDM2	E3 ubiquitin-protein ligase Mdm2	component of the TRIM28/KAP1-ERBB4-MDM2 complex which links growth factor and DNA damage response pathways
APC	Adenomatous polyposis coli protein	required for the localization of MACF1 to the cell membrane and this localization of MACF1 is critical for its function in microtubule stabilization.
MSH6	DNA mismatch repair protein Msh6 isoform 1	post-replicative DNA mismatch repair system (MMR)
HRAS	GTPase HRas	Involved in the activation of Ras protein signal transduction
MUTYH	Adenine DNA glycosylase	Involved in oxidative DNA damage repair.
DNMT3B	DNA (cytosine-5)-methyltransferase 3B	Required for genome-wide de novo methylation and is essential for the establishment of DNA methylation patterns during development.
CA9	Carbonic anhydrase 9	May be involved in the control of cell proliferation and transformation.
FBXW7	F-box/WD repeat-containing protein 7	Recognizes and binds phosphorylated sites/phosphodegrons within target proteins and thereafter bring them to the SCF complex for ubiquitination
PMS2	MMR endonuclease PMS2 isoform a	Post-replicative DNA mismatch repair system
MSH3	DNA mismatch repair protein Msh3	Post-replicative DNA mismatch repair system (MMR)
MLH2	PMS1 protein homolog 1 isoform a	repair of mismatches in DNA
PCNA	proliferating cell nuclear antigen	In response to DNA damage,this protein is ubiquitinated and is involved in the RAD6-dependent DNA repair pathway
RPA1	Replication protein A1	stabilizes single-stranded DNA intermediates, that form during DNA replication or upon DNA stress
H2AX	histone H2AX	marking sites of DNA damage, protects DNA from getting damaged with the UV radiations of sun.
GTF3C3	General Transcription Factor 3C polypeptide 3 Isoform 1	Involved in RNA polymerase III-mediated transcription.
TAF15	TATA-binding protein-associated factor 2N isoform 1	TAF15 is important for cellular proliferation and regulates the expression of a subset of cell cycle genes through miRNAs

ATF7IP	Activating Transcription factor7 interacting protein isoform 1	Chromatin formation,required to stimulate histone methyltransferase activity of SETDB1
MYST2	histone acetyltransferase	gene regulation, DNA replication, and repair
HDAC2	Histone deacetylases 2	transcriptional regulation, cell cycle progression and developmental events
HDAC3	Histone deacetylases 3	deacetylation of lysine residues on the N-terminal part of the core histones
HDAC5	Histone deacetylases 5	deacetylation of lysine residues on the N-terminal part of the core histones
HDAC10	Histone deacetylases 10	deacetylates MSH2 at Lys-73
XRCC4	DNA repair protein XRCC4 isoform 2	Involved in DNA non-homologous end joining (NHEJ) required for double-strand break repair and V(D)J recombination.
YB1	nuclease-sensitive element-binding protein 1	Promotes separation of DNA strands that contain mismatches or are modified by cisplatin.
ATR	serine/threonine-protein kinase ATR	activates checkpoint signaling upon genotoxic stresses such as ionizing radiation (IR), ultraviolet light (UV), or DNA replication stalling, thereby acting as a DNA damage sensor.
PINK1	Serine/threonine-protein kinase PINK1	Protects against mitochondrial dysfunction during cellular stress by phosphorylating mitochondrial proteins.
POLG	DNA polymerase subunit gamma-1	Involved in the replication of mitochondrial DNA. Associates with mitochondrial DNA.
SUV39H2	Histone-lysine N-methyltransferase SUV39H2	cell cycle regulation, transcriptional repression and regulation of telomere length
JARID2	Protein Jumonji	modulates histone methyltransferase activity and promotes the recruitment of histone methyltransferase complexes
BMI1	Polycomb complex protein BMI-1	chromatin remodeling and modification of histones
MGMT	Methylated-DNA--protein-cysteine methyltransferase	Repairs alkylated guanine in DNA
NTHL1	Endonuclease III-like protein 1	base excision repair
OGG1	N-glycosylase/DNA lyase isoform 1a	DNA repair enzyme that incises DNA at 8-oxoG residues
ERCC3	TFIIH basal transcription factor complex helicase XPB subunit isoform a	Nucleotide excision repair (NER) of DNA
ERCC4	DNA repair endonuclease XPF	Involved in homologous recombination that assists in removing interstrand cross-link.
RAD50	DNA repair protein RAD50	Required to bind DNA ends and hold them in close proximity
XRCC5	X-ray repair cross-complementing protein 5	involved in stabilizing broken DNA ends and bringing them together
SMUG1	single-strand selective monofunctional uracil DNA glycosylase isoform1	Recognizes base lesions in the genome and initiates base excision DNA repair.

RAD21	Double-strand-break repair protein rad21 homolog	Cleavable component of the cohesin complex, involved in chromosome cohesion during cell cycle, in DNA repair, and in apoptosis.
RFC1	Replication factor C subunit 1	play a role in DNA transcription regulation as well as DNA replication and/or repair.
RFC3	Replication factor C subunit 3	elongation of primed DNA templates by DNA polymerase delta and epsilon requires the action of the accessory proteins proliferating cell nuclear antigen (PCNA) and activator 1.
MERTK	MERTK protein	ATP binding
NBS1	Cell cycle regulatory protein p95	MRE11-RAD50-NBN (MRN complex) which plays a critical role in the cellular response to DNA damage and the maintenance of chromosome integrity.
BRCA1	BRCA1	E3 ubiquitin-protein ligase that specifically mediates the formation of 'Lys-6'-linked polyubiquitin chains and plays a central role in DNA repair by facilitating cellular responses to DNA damage.
EPCAM	Epithelial cell adhesion molecule	May act as a physical homophilic interaction molecule between intestinal epithelial cells (IECs) and intraepithelial lymphocytes (IELs) at the mucosal epithelium for providing immunological barrier as a first line of defense against mucosal infection.
TP53	Cellular tumor antigen p53	induces growth arrest or apoptosis depending on the physiological circumstances and cell type.
PMS1	PMS1 protein homolog1 isoform a	repair of mismatches in DNA
MLH1	DNA mismatch repair protein Mlh1 isoform 1	*Heterodimerizes with PMS2 to form MutL alpha, a component of the post-replicative DNA mismatch repair system (MMR)
CHEK2	Checkpoint Kinase 2	Serine/threonine-protein kinase which is required for checkpoint-mediated cell cycle arrest, activation of DNA repair and apoptosis in response to the presence of DNA double-strand breaks.
RFC3	Replication factor C subunit 3	elongation of primed DNA templates by DNA polymerase delta and epsilon requires the action of the accessory proteins proliferating cell nuclear antigen (PCNA) and activator 1.
LIG1	DNA ligase1 isoform 1	repairing single strand breaks in double stranded DNA of an organism
AURKA	AURKA	acts as a key regulatory component of the p53/TP53 pathway, and particularly the checkpoint-response pathways critical for oncogenic transformation of cells, by phosphorylating and stabilizing p53/TP53.
CCND1	G1/S-specific cyclin-D1	Regulatory component of the cyclin D1-CDK4 (DC) complex that phosphorylates and inhibits members of the retinoblastoma (RB) protein family including RB1 and regulates the cell-cycle during G1/S transition.
POLD1	DNA polymerase delta catalytic subunit	As the catalytic component of the Pol-delta3 complex and te DNA polymerase delta complexes (Pol-delta4 complex), plays a crucial role in high fidelity genome replication, including in lagging strand synthesis, and repair.

HMGB1	Highly mobility group box 1	repairs small lesions or inappropriate bases on DNA
H2AFX	Histone H2AX	Variant histone H2A which replaces conventional H2A in a subset of nucleosomes.
ERCC1	DNA excision repair protein ERCC-1 isoform 3	Non-catalytic component of a structure-specific DNA repair endonuclease responsible for the 5'-incision during DNA repair.
ERCC2	TFIIH basal transcription factor complex helicase XPD subunit isoform1	Involved in nucleotide excision repair (NER) of DNA by opening DNA around the damage
SLC19A1	Folate transporter 1 isoform 1	DNA replication and repair
PTGS2	Prostaglandin G/H synthase	Up-regulation of PTGS2 is also associated with increased cell adhesion, phenotypic changes, resistance to apoptosis and tumor angiogenesis.
CD44	CD44 antigen	Mediates cell-cell and cell-matrix interactions through its affinity for HA, and possibly also through its affinity for other ligands such as osteopontin, collagens, and matrix metalloproteinases (MMPs).
CA4	Carbonic anhydrase 4	Reversible hydration of carbon dioxide. May stimulate the sodium/bicarbonate transporter activity of SLC4A4 that acts in pH homeostasis.
ACACB	Acetyl-CoA carboxylase 2	role in regulation of mitochondrial fatty acid oxidation through malonyl-CoA-dependent inhibition of carnitine palmitoyltransferase 1
DST	Dystonin	Acts as an integrator of intermediate filaments, actin and microtubule cytoskeleton networks.
ENO1	Alpha-enolase	Multifunctional enzyme that, as well as its role in glycolysis, plays a part in various processes such as growth control, hypoxia tolerance and allergic responses.May be a tumor suppressor.
FHL1	Four and a half LIM domains protein 1	May have an involvement in muscle development or hypertrophy.
GUCA2A	Guanylin	It stimulates this enzyme through the same receptor binding region as the heat-stable enterotoxins
HNRNPL	Heterogeneous nuclear ribonucleoprotein L	Component of the heterogeneous nuclear ribonucleoprotein (hnRNP) complexes and associated with most nascent transcripts
MYH11	Myosin-11	Cellular myosin that appears to play a role in cytokinesis,cell shape, secretion & capping).
PPIB	Peptidyl-prolyl cis-trans isomerase B	PPIases accelerate the folding of proteins. It catalyzes the cis-trans isomerization of proline imidic peptide bonds in oligopeptides.
SLC26A3	Chloride anion exchanger	Mediates the efficient absorption of chloride ions in the colon, participating in fluid homeostasis.
SORD	Sorbitol dehydrogenase	Part of the polyol pathway that plays an important role in sperm physiology.
TMEM97	Sigma intracellular receptor 2	Intracellular orphan receptor that binds numerous drugs and which is highly expressed in various proliferating cancer cells .
UGT1A10	UDP-glucuronosyltransferase 1-10	UDPGT is of major importance in the conjugation and subsequent elimination of potentially toxic xenobiotics and endogenous compounds.

XRCC1	DNA repair protein XRCC1	Involved in DNA single-strand break repair by mediating the assembly of DNA break repair protein complexes.
LGR5	Leucine-rich repeat-containing G-protein coupled receptor 5	Receptor for R-spondins that potentiates the canonical Wnt signaling pathway and acts as a stem cell marker of the intestinal epithelium and the hair follicle.
SELE	E-selectin	Cell-surface glycoprotein having a role in immunoadhesion.
MTHFR	Methylenetetrahydrofolate reductase	Catalyzes the conversion of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate, a co-substrate for homocysteine remethylation to methionine.
HMGC R	3-hydroxy-3-methylglutaryl-coenzyme A reductase	Transmembrane glycoprotein that is the rate-limiting enzyme in cholesterol biosynthesis as well as in the biosynthesis of nonsterol isoprenoids that are essential for normal cell function including ubiquinone and geranylgeranyl proteins.
WNT5B	Protein Wnt-5b	Ligand for members of the frizzled family of seven transmembrane receptors.
NR1I2	Nuclear receptor subfamily 1 group I member 2	Nuclear receptor that binds and is activated by variety of endogenous and xenobiotic compounds.
ABCC2	Canalicular multispecific organic anion transporter 1	Mediates hepatobiliary excretion of numerous organic anions.
KRAS	GTPase KRas	Ras proteins bind GDP/GTP and possess intrinsic GTPase activity.
EGF	Pro-epidermal growth factor	EGF stimulates the growth of various epidermal and epithelial tissues in vivo and in vitro and of some fibroblasts in cell culture.
FCGR3 A	Low affinity immunoglobulin gamma Fc region receptor III-A	Receptor for the Fc region of IgG. Binds complexed or aggregated IgG and also monomeric IgG.
UGT1A6	UDP-glucuronosyltransferase 1-6	DPGT is of major importance in the conjugation and subsequent elimination of potentially toxic xenobiotics and endogenous compounds.
ABCC1	Multidrug resistance-associated protein 1	Mediates export of organic anions and drugs from the cytoplasm.
ABCG1	ATP-binding cassette sub-family G member 1	Transporter involved in macrophage lipid homeostasis.
PTGER 4	Prostaglandin E2 receptor EP4 subtype	The activity of this receptor is mediated by G(s) proteins that stimulate adenylate cyclase.
TYMS	Thymidylate synthase	Contributes to the de novo mitochondrial thymidylate biosynthesis pathway.
EGFR	Epidermal growth factor receptor	Receptor tyrosine kinase binding ligands of the EGF family and activating several signaling cascades to convert extracellular cues into appropriate cellular responses.
ABCB1	Multidrug resistance protein 1	Energy-dependent efflux pump responsible for decreased drug accumulation in multidrug-resistant cells.
ABCC5	Multidrug resistance-associated protein 5	Acts as a multispecific organic anion pump which can transport nucleotide analogs.
GSTP1	Glutathione S-transferase P	conjugation of reduced glutathione to a wide number of exogenous and endogenous hydrophobic electrophiles.
ABCG2	ATP-binding cassette sub-family G member 2	plays a role in porphyrin homeostasis as it is able to mediates the export of protoporhyrin IX (PPIX) both from mitochondria to cytosol and from cytosol to extracellular space, and cellular export of hemin, and heme.
ALOX12	Arachidonate 12-lipoxygenase, 12S-type	Plays a role in apoptotic process, promoting the survival of vascular smooth muscle cells for instance.

IL23R	Interleukin-23 receptor	Associates with IL12RB1 to form the interleukin-23 receptor.
HLA-G	HLA class I histocompatibility antigen, alpha chain G	nvolved in the presentation of foreign antigens to the immune system.
UGT1A7	UDP-glucuronosyltransferase 1-7	UDPGT is of major importance in the conjugation and subsequent elimination of potentially toxic xenobiotics and endogenous compounds.
DPYD	Dihydropyrimidine dehydrogenase [NADP(+)]	Involved in pyrimidine base degradation.
SLC29A1	Equilibrative nucleoside transporter 1	Mediates both influx and efflux of nucleosides across the membrane (equilibrative transporter).
SLCO1B1	Solute carrier organic anion transporter family member 1B1	Mediates the Na+-independent uptake of organic anions.
SHMT1	Serine hydroxymethyltransferase, cytosolic	Interconversion of serine and glycine
PTGES	Prostaglandin E synthase	Catalyzes the oxidoreduction of prostaglandin endoperoxide H2 (PGH2) to prostaglandin E2 (PGE2)
AREG	Amphiregulin	Ligand of the EGF receptor/EGFR.
MGAT4A	Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase A	Glycosyltransferase that participates in the transfer of N-acetylglucosamine (GlcNAc) to the core mannose residues of N-linked glycans.
PARD3B	Partitioning defective 3 homolog B	Putative adapter protein involved in asymmetrical cell division and cell polarization processes.
UGT1A9	UDP-glucuronosyltransferase 1-9	This isoform has specificity for phenols.
ENOSF1	Mitochondrial enolase superfamily member 1	Plays a role in the catabolism of L-fucose, a sugar that is part of the carbohydrates that are attached to cellular glycoproteins.
CXCR4	C-X-C chemokine receptor type 4	Receptor for the C-X-C chemokine CXCL12/SDF-1 that transduces a signal by increasing intracellular calcium ion levels and enhancing MAPK1/MAPK3 activation.
ADCY2	Adenylate cyclase type 2	Catalyzes the formation of the signaling molecule cAMP in response to G-protein signalling
TGFB1	Transforming growth factor beta-1	Multifunctional protein that controls proliferation, differentiation and other functions in many cell types.
RGS5	Regulator of G-protein signaling 5	Inhibits signal transduction by increasing the GTPase activity of G protein alpha subunits thereby driving them into their inactive GDP-bound form.
KLC1	Kinesin light chain 1	Kinesin is a microtubule-associated force-producing protein that may play a role in organelle transport.
VEGFA	Vascular endothelial growth factor A	Growth factor active in angiogenesis, vasculogenesis and endothelial cell growth.
TGFBR2	TGF-beta receptor type-2	Transmembrane serine/threonine kinase forming with the TGF-beta type I serine/threonine kinase receptor, TGFBR1, the non-promiscuous receptor for the TGF-beta cytokines TGFB1, TGFB2 and TGFB3
XRCC3	DNA repair protein XRCC3	Involved in the homologous recombination repair (HRR) pathway of double-stranded DNA, thought to repair chromosomal fragmentation, translocations and deletions.

Table 3.4: Gene along with their proteins and functions.

The work proposed the rate of MMR in the DNA damage response (DDR) that activates cell cycle seize and in some occurrence apoptosis. The focus is on the genes that are directly linked to the MMR, HNPCC and CRC. The MMR promotes a DDR mediated key signalling events in return to various types of DNA damage incorporating those that are experienced in radiation/chemotherapy.

The connection was set up on the server and the genes were uploaded on the website i.e., DRMAP. The GUI was accomplished by HTML, CSS, JS and PHP and the search on the platform can be done by Gene ID, Uniprot ID, Ensemble ID, Expression, Interactions, Locations and Gene Ontology. There are various options of pages- HOME, ABOUT, CONTACT and HELP (see figures 3.7-3.18).

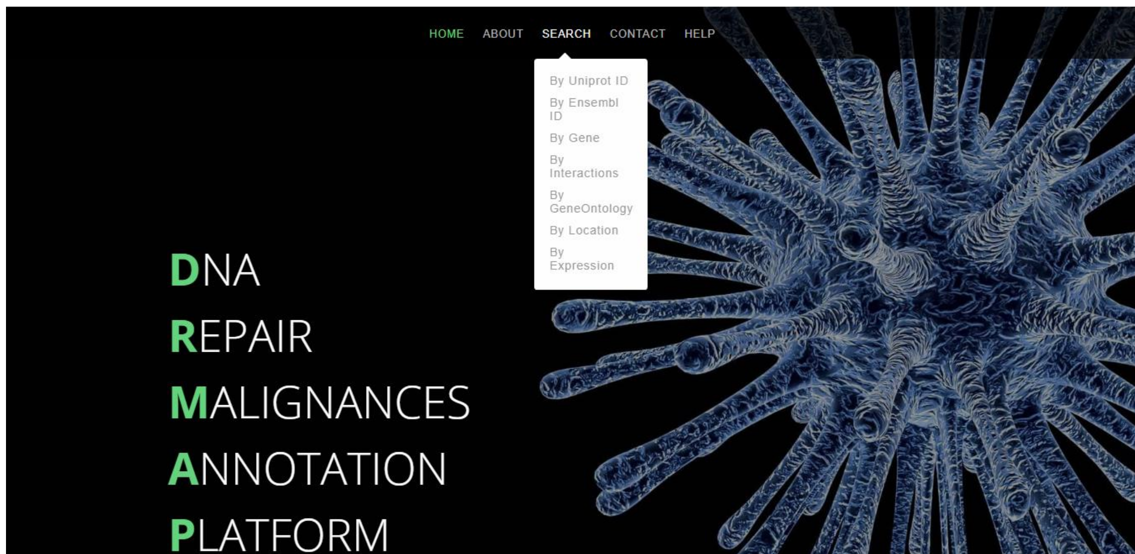


Figure 3.7: Home page on the website showing theavailable search options

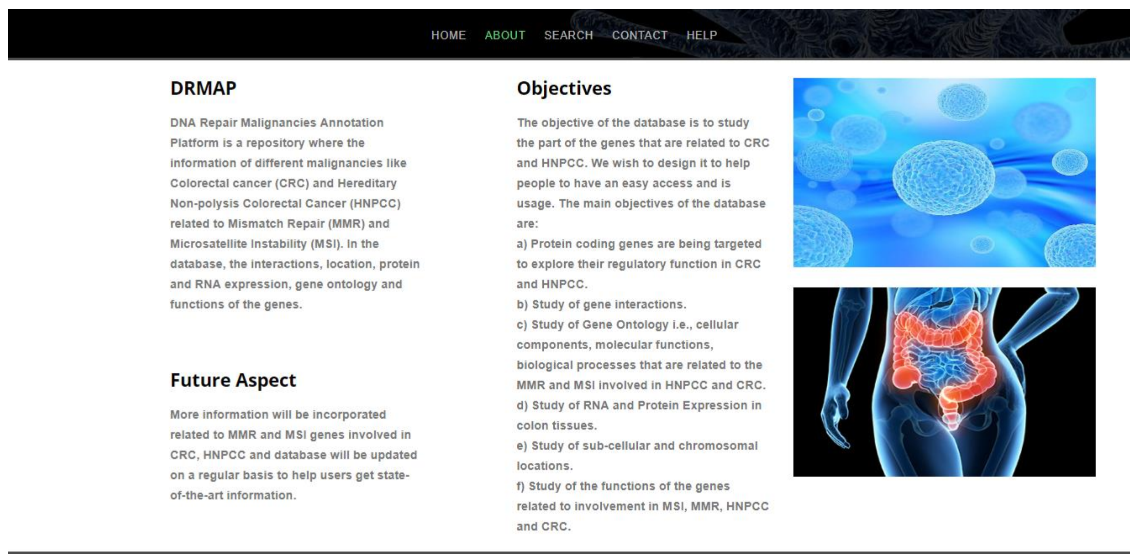


Figure 3.8: About page displaying the objectives and Future Aspects

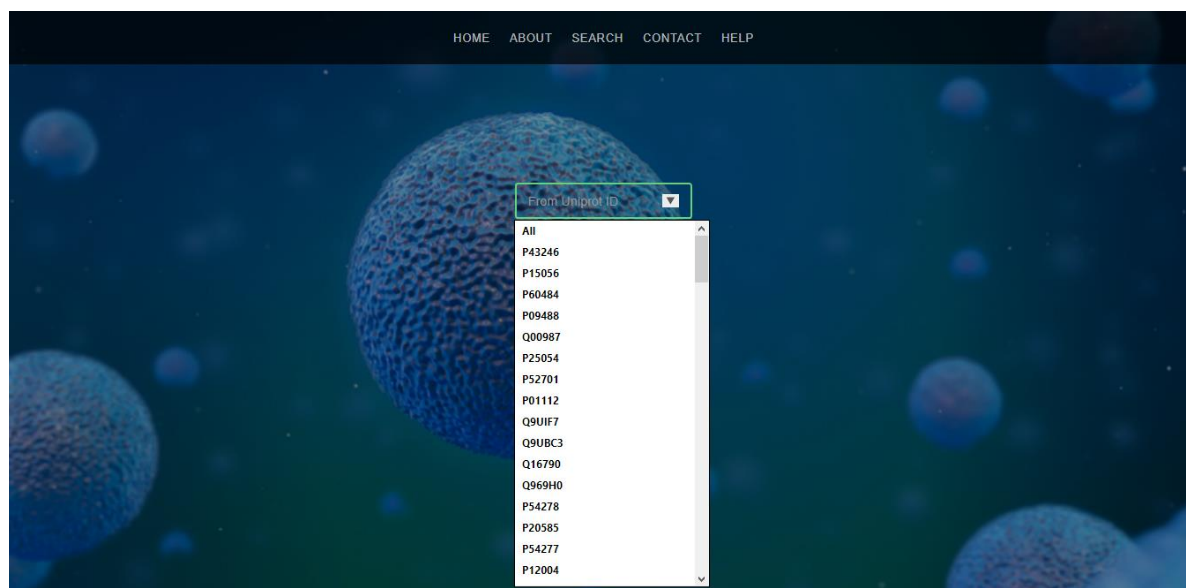


Figure 3.9: Displays the dropdown for the Uniprot IDs

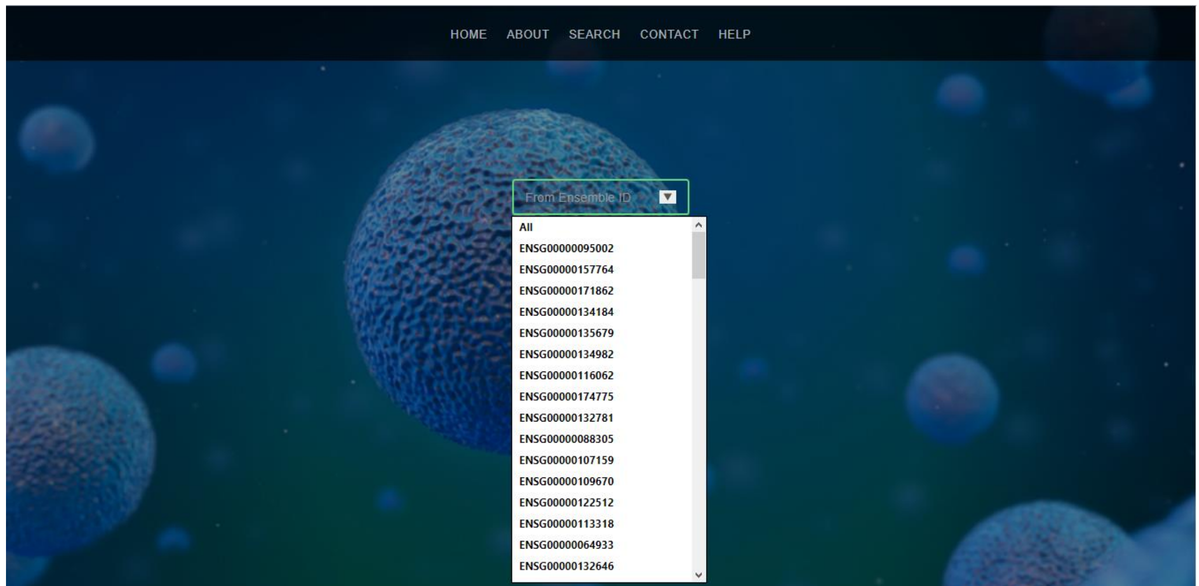


Figure 3.10: Dropdown for the Ensemble IDs is displayed

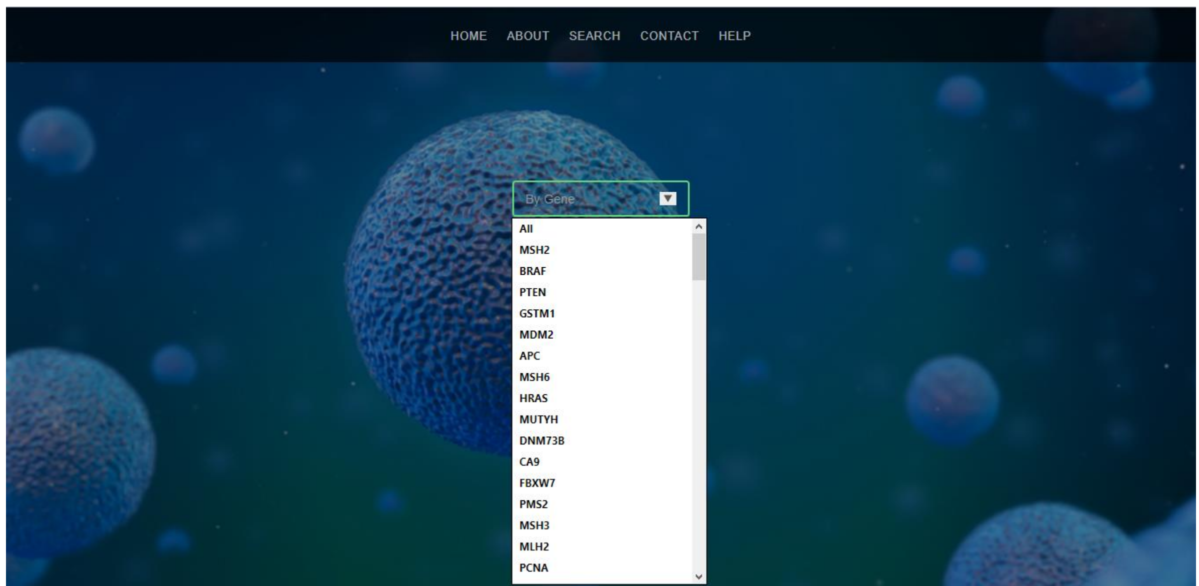


Figure 3.11: Shows the Gene as dropdown

HOME ABOUT SEARCH CONTACT HELP

From Uniprot ID

Search

Uniprot_ID	Gene_ID	Protein	Function	Domain	Structure	Ensemble_ID
P43246	MSH2	DNA mismatch repair protein Msh2	Component of the post-replicative DNA mismatch repair system (MMR).	MutS_1, MutS_III superfamily	3THZ	ENSG00000095002
P15056	BRAF	Serine/threonine-protein kinase B-raf	Protein kinase involved in the transduction of mitogenic signals from the cell membrane to the nucleus.	STKc_Raf,	5J17	ENSG00000157764
P60484	PTEN	Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN	Acts as a dual-specificity protein phosphatase, dephosphorylating tyrosine-, serine- and threonine-phosphorylated proteins.	PTEN_C2	4O1V	ENSG00000171862
P09488	GSTM1	Glutathione S-transferase Mu 1	Conjugation of reduced glutathione to a wide number of exogenous and endogenous hydrophobic electrophiles.	GST_C_Mu	2F3M	ENSG00000134184
Q00987	MDM2	E3 ubiquitin-protein ligase Mdm2	component of the TRIM28/KAP1-ERBB4-MDM2 complex which links growth factor and DNA damage response pathways	zf-RanBP	5TRF	ENSG00000135679
P25054	APC	Adenomatous polyposis coli	required for the localization of MACF1 to the cell membrane and this localization of MACF1 is critical for	Arm_APC_u3	4YK6	ENSG00000134982

Figure 3.12: Results are retrieved for the options Uniprot IDs, ensemble Ids and gene

HOME ABOUT SEARCH CONTACT HELP

From Biological Process

Search

From Cellular compone

All
mismatch repair complex
damaged DNA binding
mismatched DNA binding
DNA ligase activity
histone h2a pthr23430
histone h2a pthr23431
histone h2a pthr23432
histone h2a pthr23433
histone h2a pthr23434

ABOUT US SEARCH GET IN TOUCH

HOME ABOUT SEARCH CONTACT HELP

From Cellular componer

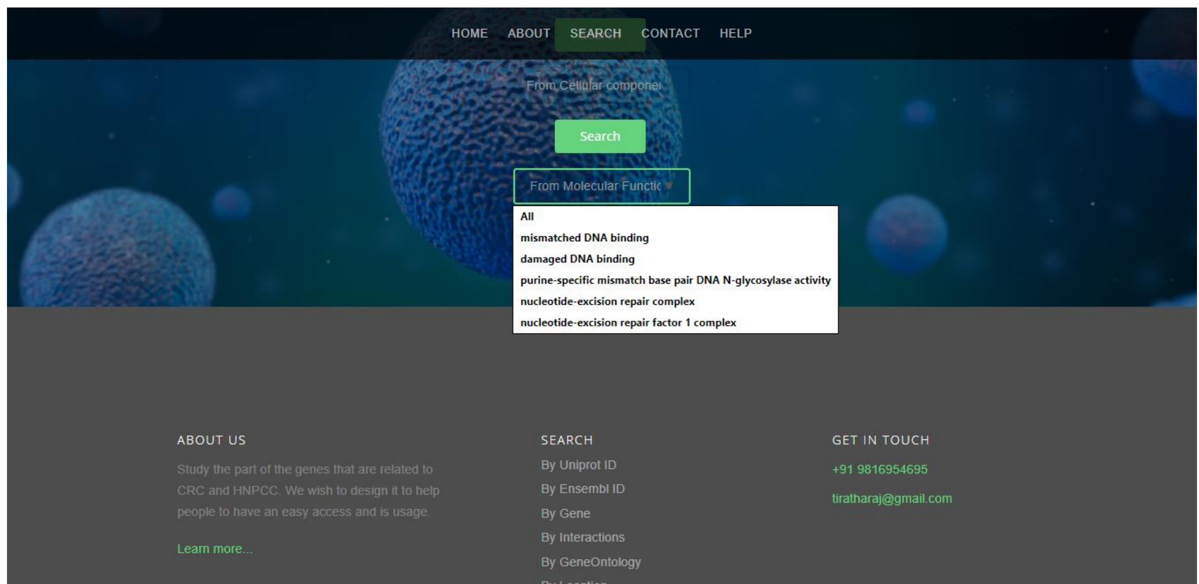
Search

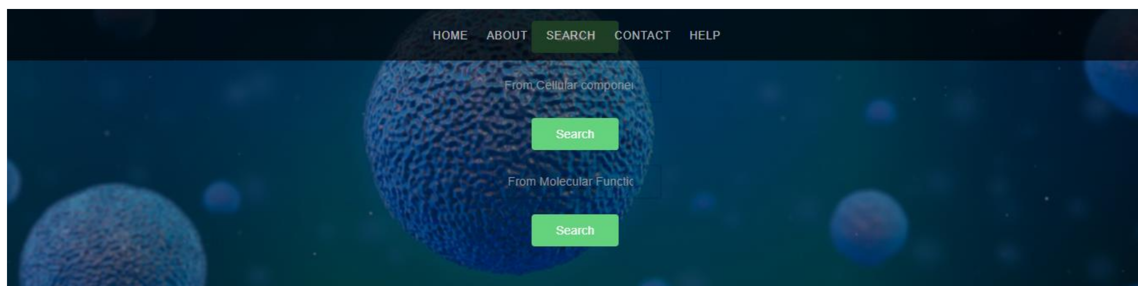
From Molecular Functic

Search

Gene_ID	GO_Class	Evidence	Panther
MSH2	mismatch repair complex	IBA	dna mismatch repair muts related proteins pthr11361
MSH6	mismatch repair complex	IBA	dna mismatch repair muts related proteins pthr1136
PMS2	mismatch repair complex	IBA	dna mismatch repair protein (mlh, pms, mutl) pthr10073
MSH3	mismatch repair complex	IBA	dna mismatch repair muts related proteins pthr11361
TP53	damaged DNA binding	IBA	cellular tumor antigen p53 pthr11447
POLD1	damaged DNA binding	IDA	dna polymerase catalytic subunit pthr103
HMGB1	damaged DNA binding	ISS	swi/snf-related chromatin binding protein pthr13711
ERCC1	damaged DNA binding	IDA	excision repair cross-complementing 1 ercc1pthr12749
PMS1	mismatched DNA binding	IEA	dna mismatch repair protein (mlh, pms, mutl) pthr10073
LIG1	DNA ligase activity	IMP,IDA,TAS,IBA	dna ligase pthr10459
H2AFX	histone h2a pthr23430	NA	NA
H2AFX	histone h2a pthr23431	NA	NA

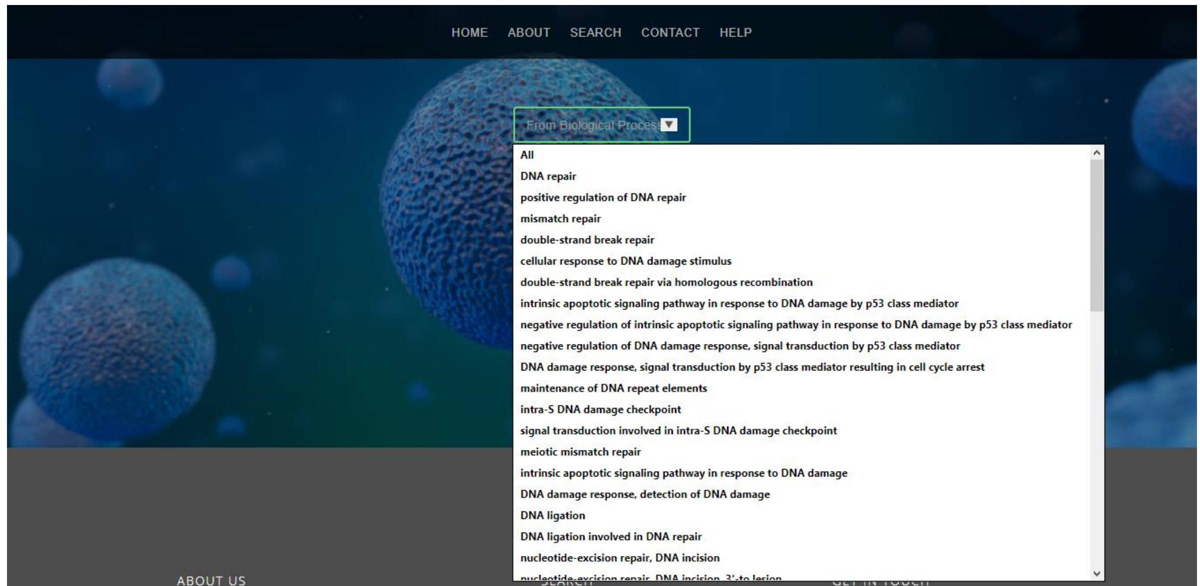
Figure 3.13: Dropdown and results for Cellular Components

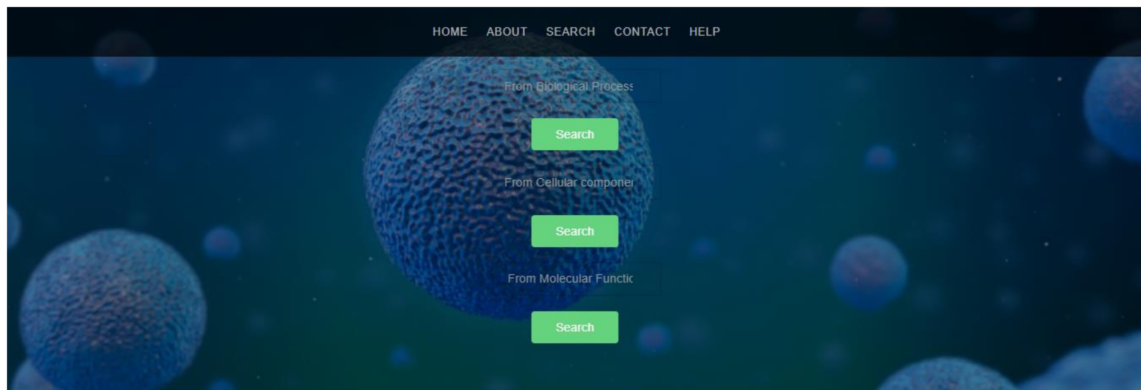




Gene_ID	GO_Class	Evidence	Panther
MSH2	mismatched DNA binding	IDA	dna mismatch repair muts related proteins pthr11361
MSH6	mismatched DNA binding	IDA	dna mismatch repair muts related proteins pthr1136
MSH3	mismatched DNA binding	IDA,IPI,IBA	dna mismatch repair muts related proteins pthr11361
MSH2	damaged DNA binding	IBA	dna mismatch repair muts related proteins pthr11361
MSH6	damaged DNA binding	IBA	dna mismatch repair muts related proteins pthr1136
MSH3	damaged DNA binding	IBA	dna mismatch repair muts related proteins pthr11361
PCNA	damaged DNA binding	IDA	proliferating cell nuclear antigen pthr11352
RPA1	damaged DNA binding	IDA	replication factor a 1, rfa1 pthr23273
H2AX	damaged DNA binding	IEA	histone h2a pthr23430
OGG1	damaged DNA binding	IDA	n-glycosylase/dna lyase pthr10242
ERCC3	damaged DNA binding	NAS	rad25/xp-b dna repair helicase pthr11274

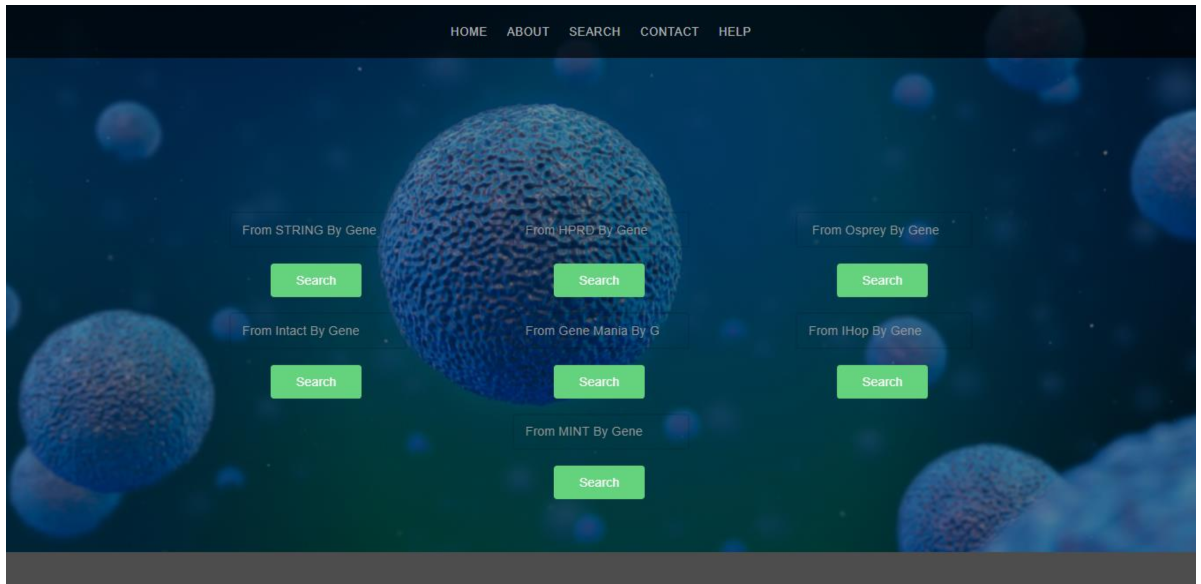
Figure 3.14: Dropdown and results for Molecular Function

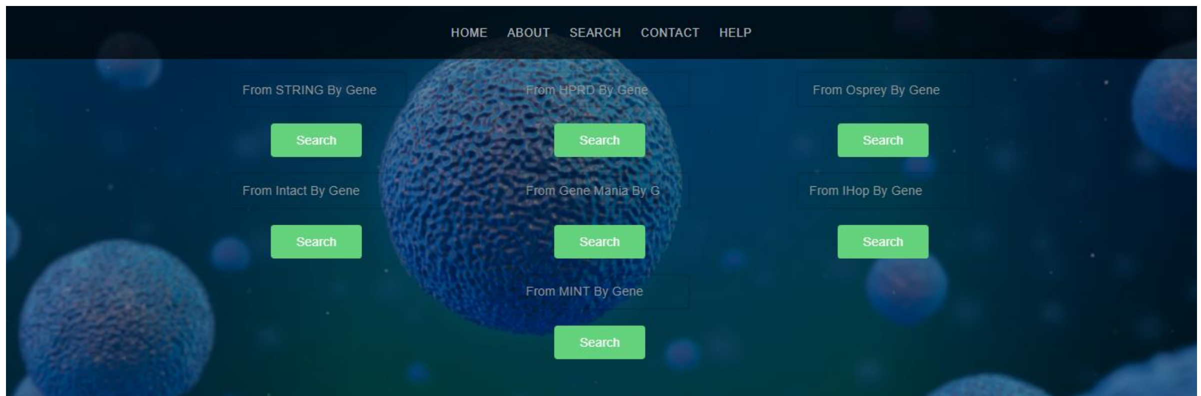




Gene_ID	GO_Class	Evidence	Panther
MSH2	DNA repair	IDA	dna mismatch repair muts related proteins pthr11361
MSH6	DNA repair	IDA	dna mismatch repair muts related proteins pthr11361
MUTYH	DNA repair	IBA,TAS	a/g-specific adenine glycosylase/endonuclease iii pthr10359
MSH3	DNA repair	IDA	dna mismatch repair muts related proteins pthr11361
RPA1	DNA repair	IMP	replication factor a 1, rfa1 pthr23273
ERCC3	DNA repair	IMP	rad25/xp-b dna repair helicase pthr11274

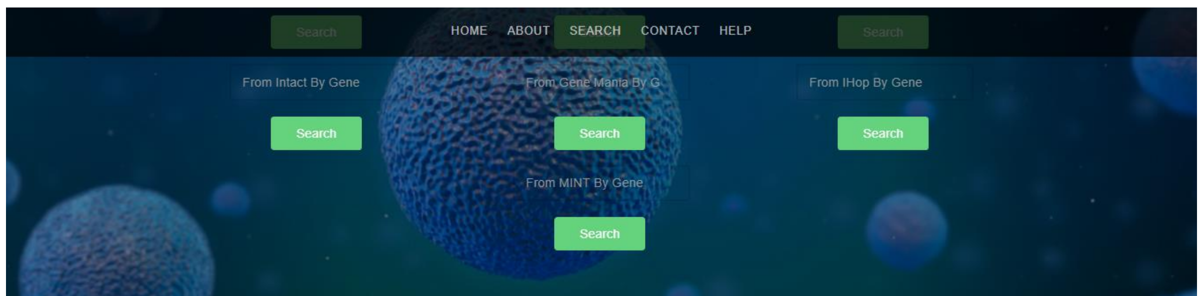
Figure 3.15: Dropdown and results for biological process





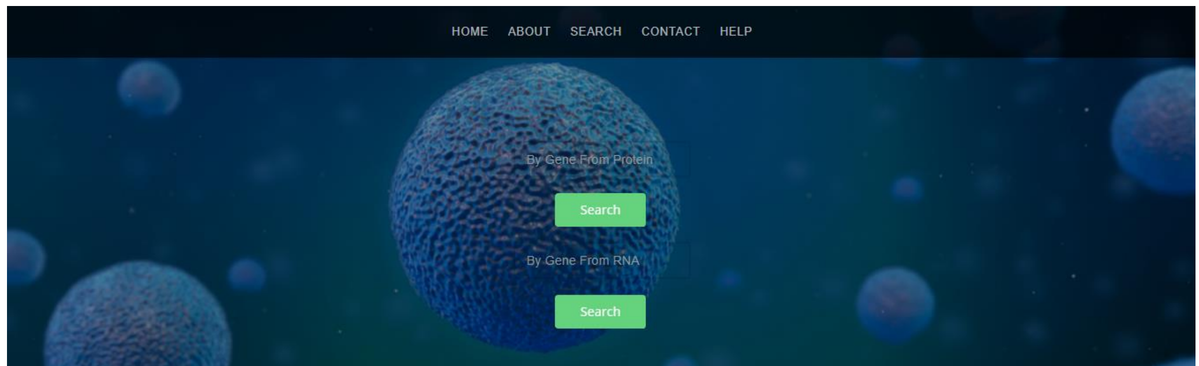
Gene_ID	Interactions
MSH2	POLA1,POLA2,MCM4,MCM5,RPA1,MSH6
BRAF	RAP1A, NRAS, RAP1B, PRKCE, PRKCA, RPS6KB2, RPS6KA2, MAP2K1, MAP3K1, MAP2K2
PTEN	MLH3, MAST2, GYG1, GNPNAT1, MARCH2, MARCH3, SPRTN, HIST1,H2BJ, GYG2
GSTM1	GSTM2, UCP1, BECN1
MDM2	MDM4, IGF1R, CASP3, TP53, BCAS3, CCNG1, IFNAR1, GADD45A,CDKN1A, AKT1
APC	UCKL1, FAM123B, CTNNB1, UPRT, UCKL1, CTNNA1, JUP, AXIN2, AXIN1, GSK3B
[illegible]	[illegible]

Figure 3.16: Interaction results from STRING.

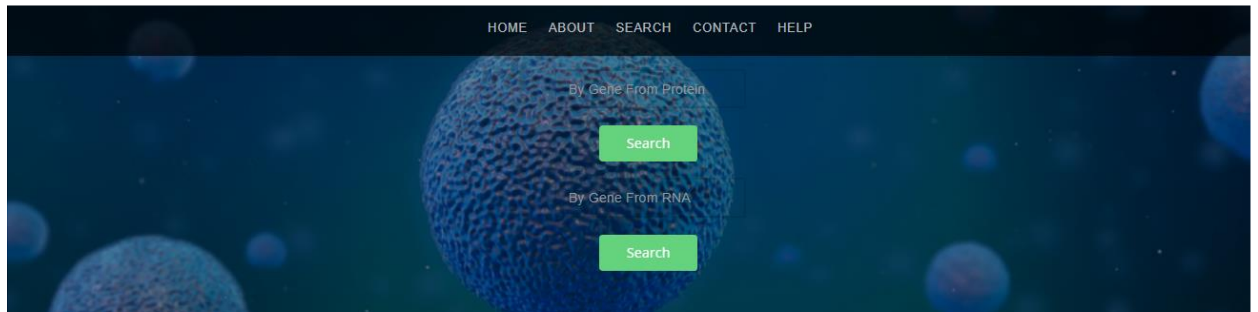


Gene_ID	Interactant_Proteins	Evidences	Interactions
MSH2	CREBBP, Proliferating cell nuclear antigen, SMC1, MutS homolog 3,Exonuclease 1,MSH6	In Vivo ; In Vitro; yeast two hybrid	Direct
BRAF	14-3-3 Eta, 14-3-3 theta, 14-3-3 zeta, Opa interacting protein 5	yeast two hybrid	Direct
PTEN	NA	NA	NA
GSTM1	MAP3K1, ASK1	in vitro	direct
MDM2	Ataxin 1 ubiquitin like interacting protein, Ubiquitin B, Ubiquitin specific protease 2, Transcription factor IID, CDKN2A, MDM4, Numb homolog, DNA polymerase epsilon, catalytic subunit A, Ribosomal protein L11, Telethonin, p53, Dihydrofolate reductase	yeast two hybrid	direct
APC	Protein phosphatase 2, regulatory subunit B (B56), alpha, SIAH1, Microtubule associated protein, RP/EB family, member 1, AXIN1, Axin 2, Rho guanine nucleotide exchange factor 4, Protein tyrosine phosphatase, nonreceptor 13,	yeast two hybrid	direct, complex

Figure 3.17: Intact results for interactions



Gene_ID	Colon
CD44	Endothelial Cells:Low
CD44	Glandular Cells:Medium
CD44	Peripheral nerve/ganglion:Medium
CA4	Endothelial cells:Not detected
CA4	Glandular cells:Medium
CA4	Peripheral nerve/ganglion:Not detected
ACACB	Endothelial cells:Not detected
ACACB	Glandular cells:Low



Gene	HPA(TPM)	GTEx(RPKM)	FANTOM5(Tags per million)
XRCC1	11.9	10.1	20.7
LGR5	3.4	0.6	12.9
SELE	1.2	2.6	2.2
MTHFR	9.4	4.5	23.2
HMGCR	46.2	12	28.3
WNT5B	6.1	5.7	6.8
NR1I2	10.3	4.6	43.8
ABCC2	0.3	0.3	28.7
KRAS	28.6	7.5	34.8

Figure 3.18: RNA and protein expression

From the literature review, there was a study on XRCC1 on Kashmiri population thatshowed a correlation between XRCC1 A870 polymorphism and risk of CRC(Nissar et al., 2015).KRAS mutations hasadistinct patternin colorectal carcinomas indicated by the germlineMMR defects. The BRAF is associated inversely to mutations in KRAS and directly linkedto MSI. EGFR and ABCG2 are overexpressedin this condition(Oliveira et al., 2014). TYMS and SLC29A1resulted in up-regulation of theCRC (Valentin et al., 2013, Vecchio et al., 2017)WNT5B shows a reduced expression in CRCwhereas an elevated level of the WNT5A was seen (S. Li, 2008). PARD3B protein interact with SMAD3in TGF-beta signalling pathway(Rozadilla et al., 2012).

Colorectal adenocarcinoma development have beenlinked to the mutations in SMAD3. The polymorphisms in ERCC2 may be attached to the risk of developing CRC(Kabzinski et al., 2015).High levels of H2AFX, AURKA and CXCR4 affects theprognosis in CRC while high expression of TP53 had a good prognosis in CRC(Li et al.,2017,Lee et al., 2015,Goos et al., 2013,Adrover et al.,1999). There is reduced expression in MSI cancer of MYH11i.e., less smooth muscle component. Disruption of LIG1 may result in the disorder (Colorectal Cancer Atlas). Inhibition of PTGS2 prevents thetumour growth (Wang et al., 2011). ADYC2 is one of the gene in MetastasisSignalling Network of CRC(FANGet al., 2013). ALOX12 may be one of the predisposing factors to CRC(Küry et al.,2008). Expression of AREG is significantly higher in the left-sided CRCs biomarkers(Zarkavelis et al.,2017). VEGF1 is predominant angiogenicfactors in CRC (Bendardaf et al., 2008).UGT1A9 polymorphismpredicts response and toxicity in CRC patients (Carlini et al., 2005).CCND1 was not possessed by the normal people but by the patients (Balcerczak et al., 2005).HLA-G can be beneficial in predicting the prognosis of CRCs (Zeestraten et al., 2013).

Chapter-4

Conclusion

The MMR systems in mammalian cells are primarily known to understand the fidelity of DNA sequences which might be occurred at the time of DNA replication. But under some stressful condition, defect in MMR leads to severe disorders like HNPCC and sporadic cancers. Our study provided an additional set of genes to MMR and HNPCC which were previously either unexplored or uncovered. Furthermore, the vital genes are screened and their role is found. A new resource named DRMAP was developed through the information generated from bioinformatics analysis. Other information from standard resources was also collected and incorporated into DRMAP. We hope that this resource would serve as a useful training dataset for predicting whether or not any proteins are related to MMR. It is anticipated that experimental validation of these genes will provide a promising set of bio markers for the maintenance of the disease. In the future, miRNAs, lncRNAs, network and pathways could be analysed to evaluate the role of other genes in the disease. Similar approaches can be applied for other diseases. There is a need to investigate other genes involved in the disease and their specific regulatory role through computational analysis followed by experimental validation.

References

A. Moussa, Brigitte Vannierand MounirMaouene, "New Algorithm for Gene Selection in Microarray Data Analysis Workflow." Computer Technology and Application 3(2): 169-174, 2012.

A. Ceol, A. ChatrAryamontri, L. Licata et al., "MINT, the molecular interaction database: 2009 update", Nucleic Acids Research, vol. 38, no. 1, pp. D532-D539, 2009.

A. Shukla , A. Moussa and T. Singh, "DREMECELS: A Curated Database for Base Excision and Mismatch Repair Mechanisms Associated Human Malignancies", PLOS ONE, vol. 11, no. 6, p. e0157031, 2016.

B.J Breitkreutz, C. Stark, M. Tyers, "Osprey: a network visualization system": 2003 *Genome Biology*, vol. *4,* no .3, pp. R22, 2003.

C. Bronner, S. Baker, P. Morrison et al., "Mutation in the DNA mismatch repair gene homologue hMLH 1 is associated with hereditary non-polyposis colon cancer", Nature, vol. 368, no. 6468, pp. 258-261, 1994.

C. Fernandez-Rozadilla, J. Cazier, V. Moreno et al., "Pharmacogenomics in colorectal cancer: a genome-wide association study to predict toxicity after 5-fluorouracil or FOLFOX administration", The Pharmacogenomics Journal, vol. 13, no. 3, pp. 209-217, 2012.

C. Oliveira, "Distinct patterns of KRAS mutations in colorectal carcinomas according to germline mismatch repair defects and hMLH1 methylation status", Human Molecular Genetics, vol. 13, no. 19, pp. 2303-2311, 2004.

D. Huang, B. Sherman, Q. Tan et al., "DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists", Nucleic Acids Research, vol. 35, no. 2, pp. W169-W175, 2007.

D. Szklarczyk, A. Franceschini, S. Wyder et al., "STRING v10: protein–protein interaction networks, integrated over the tree of life", Nucleic Acids Research, vol. 43, no. 1, pp. D447-D452, 2014.

D. Wang, D. Xia and R. DuBois, "The Crosstalk of PTGS2 and EGF Signaling Pathways in Colorectal Cancer", Cancers, vol. 3, no. 4, pp. 3894-3908, 2011.

E. Adrover, M. Maestro, M. Sanz-Casla et al., "Expression of high p53 levels in colorectal cancer: a favourable prognostic factor", British Journal of Cancer, vol. 81, no. 1, pp. 122-126, 1999.

E. Balcerczak, G. Pasz-Walczak, P. Kumoret al., "Cyclin D1 protein and CCND1 gene expression in colorectal cancer", European Journal of Surgical Oncology (EJSO), vol. 31, no. 7, pp. 721-726, 2005.

E. Zeestraten, M. Reimers, S. Saadatmand et al., "Combined analysis of HLA class I, HLA-E and HLA-G predicts prognosis in colon cancer patients", British Journal of Cancer, vol. 110, no. 2, pp. 459-468, 2013.

F. Del Vecchio, V. Mastroiaco, A. Di Marco et al., "Next-generation sequencing: recent applications to the analysis of colorectal cancer", Journal of Translational Medicine, vol. 15, no. 1, 2017.

F. Kadyrov, L. Dzantiev, N. Constantin et al., "Endonucleolytic Function of MutLα in Human Mismatch Repair", Cell, vol. 126, no. 2, pp. 297-308, 2006.

F. Leach, N. Nicolaides, N. Papadopoulos et al., "Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer", Cell, vol. 75, no. 6, pp. 1215-1225, 1993.

F. Li, Q. Liu, Y. Chen et al., "Escherichia coli mismatch repair protein MutL interacts with the clamp loader subunits of DNA polymerase III", Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, vol. 637, no. 1-2, pp. 101-110, 2008.

F. Pontén, K. Jirström and M. Uhlen, "The Human Protein Atlas-a tool for pathology", The Journal of Pathology, vol. 216, no. 4, pp. 387-393, 2008.

Li, "Mechanisms and functions of DNA mismatch repair", Cell Research, vol. 18, no. 1, pp. 85-98, 2007.

G. Zarkavelis, "Current and future biomarkers in colorectal cancer", Annals of Gastroenterology, 2017.

H. Hermjakob, "IntAct: an open source molecular interaction database", Nucleic Acids Research, vol. 32, no. 90001, pp. 452D-455, 2004.

H. Lynch, T. Drouhard, H. Vasen, et al., "Genetic counseling in a navajo hereditary nonpolyposis colorectal cancer kindred", Cancer, vol. 77, no. 1, pp. 30-35, 1996.

H. Mi and P. Thomas, "PANTHER Pathway: An Ontology-Based Pathway Database Coupled with Data Analysis Tools", Methods in Molecular Biology, pp. 123-140, 2009.

J. Goos, V. Coupe, B. Diosdado et al., "Aurora kinase A (AURKA) expression in colorectal cancer liver metastasis is associated with poor prognosis", British Journal of Cancer, vol. 109, no. 9, pp. 2445-2452, 2013.

J. Jourquin, D. Duncan, Z. Shi et al., "GLAD4U: deriving and prioritizing gene lists from PubMed literature", BMC Genomics, vol. 13, no. 8, p. S20, 2012.

J. Kabzinski, K. Przybylowska, L. Dziki et al., "An association of selected ERCC2 and ERCC5 genes polymorphisms, the level of oxidative DNA damage and its repair efficiency with a risk of colorectal cancer in Polish population", Cancer Biomarkers, vol. 15, no. 4, pp. 413-423, 2015.

J. Montojo, K. Zuberi, H. Rodriguez et al., "GeneMANIA: Fast gene network construction and function prediction for Cytoscape", F1000Research, 2014.

J. Wang, D. Duncan, Z. Shi et al., "WEB-based GEneSeTAnaLysis Toolkit (WebGestalt): update 2013", Nucleic Acids Research, vol. 41, no. 1, pp. W77-W83, 2013.

L. Carlini, N. Meropol and J. Bever et al., "UGT1A7 and UGT1A9 polymorphisms predict response and toxicity in colorectal cancer patients treated with capecitabine/irinotecan.", Clinical Cancer Research, vol. 11, no. 3, pp. 1226-36, 2005.

L.T FANG, S. LEE, H. CHOI et al., "Comprehensive genomic analyses of a metastatic colon cancer to the lung by whole exome sequencing and gene expression analysis", International Journal of Oncology, vol. 44, no. 1, pp. 211-221, 2013.

M. Magrane and U. Consortium, "UniProt Knowledgebase: a hub of integrated data", Nature Precedings, 2010.

M. Schofield and P. Hsieh, "DNA Mismatch Repair: Molecular Mechanisms and Biological Function", Annual Review of Microbiology, vol. 57, no. 1, pp. 579-608, 2003.

M. Sehgal and T. Singh, "DR-GAS: A database of functional genetic variants and their phosphorylation states in human DNA repair systems", DNA Repair, vol. 16, pp. 97-103, 2014.

M. Sehgal, R. Gupta, A. Moussa et al., "An Integrative Approach for Mapping Differentially Expressed Genes and Network Components Using Novel Parameters to Elucidate Key Regulatory Genes in Colorectal Cancer", PLOS ONE, vol. 10, no. 7, p. e0133901, 2015.

M. Valentin, C. Therkildsen, S. Veerla et al., "Distinct gene expression signatures in Lynch syndrome and familial colorectal cancer type X", BMC Proceedings, vol. 7, no. 2, p. P52, 2013.

M. Varadi and P. Tompa, "The Protein Ensemble Database", Advances in Experimental Medicine and Biology, pp. 335-349, 2015.

M. Navarro, A. Nicolas, A. Ferrandez et al., "Colorectal cancer population screening programs worldwide in 2016: An update", World Journal of Gastroenterology, vol. 23, no. 20, p. 3632, 2017.

P. Alhopuro, H. Sammalkorpi, I. Niittymäki et al., "Candidate driver genes in microsatellite-unstable colorectal cancer", International Journal of Cancer, vol. 130, no. 7, pp. 1558-1566, 2011.

R. Bendardaf, A. Buhmeida and M. Hilksa et al., "VEGF-1 expression in colorectal cancer is associated with disease localization, stage, and long-term disease-specific survival.", Anticancer Research, vol. 28, no. 6, pp. 3865-70, 2008.

R. Fishel, M. Lescoe, M. Rao et al., "The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer", Cell, vol. 75, no. 5, pp. 1027-1038, 1993.

R. Hoffmann and A. Valencia,"A gene network for navigating the literature", Nature Genetics, vol. 36, no. 7, pp. 664-664, 2004.

R. Iyer, A. Pluciennik, V. Burdett et al., "DNA Mismatch Repair: Functions and Mechanisms", ChemInform, vol. 37, no. 20, 2006.

R. Kolodner and G. Marsischky, "Eukaryotic DNA mismatch repair", Current Opinion in Genetics & Development, vol. 9, no. 1, pp. 89-96, 1999.

R. Lahue, K. Au and P. Modrich,"DNA mismatch correction in a defined system", Science, vol. 245, no. 4914, pp. 160-164, 1989.

S. Küry, B. Buecher, S. Robiou-du-Pont et al., "Low-penetrance alleles predisposing to sporadic colorectal cancers: a French case-controlled genetic association study", BMC Cancer, vol. 8, no. 1, 2008.

S. Li, Profiling of gene expression changes in human colon crypt maturation and study of their dysregulation in tumourigenesis. University of Hong Kong (Pokfulam Road, Hong Kong), 2008.

S. Nissar, A. Sameer, R. Rasool et al., "Polymorphism of the DNA Repair Gene XRCC1 (Arg194Trp) and its role in Colorectal Cancer in Kashmiri Population: a Case Control Study", Asian Pacific Journal of Cancer Prevention, vol. 16, no. 15, pp. 6385-6390, 2015.

T. Kunkel and D. Erie, "Eukaryotic Mismatch Repair in Relation to DNA Replication", Annual Review of Genetics, vol. 49, no. 1, pp. 291-313, 2015.

V. Gorbunova, A. Seluanov, Z. Mao et al., "Changes in DNA repair during aging", Nucleic Acids Research, vol. 35, no. 22, pp. 7466-7474, 2007.

W. Yang, "Structure and function of mismatch repair proteins", Mutation Research/DNA Repair, vol. 460, no. 3-4, pp. 245-256, 2000.

Y. Lee, T. Yin and Y. Chen et al., "High Expression of Phospho-H2AX Predicts a Poor Prognosis in Colorectal Cancer", Anticancer research, vol. 35, no. 4, pp. 2447-53, 2015.

Y. Li, J. Pang, S. Gao et al., "Role of CXCR4 and SDF1 as prognostic factors for survival and the association with clinicopathology in colorectal cancer: A systematic meta-analysis", Tumor Biology, vol. 39, no. 6, p. 101042831770620, 2017.