SRI VENKATESWARA INTERNSHIP PROGRAM FOR RESEARCH IN ACADEMICS (SRI-VIPRA)

SRI-VIPRA

Project Report of 2025: SVP-2515

"Pan-Cancer Analysis Of DNA Repair Genes Reveals Interesting Relationship Between Their Expression And Patient Survival"

IQAC
Sri Venkateswara College
University of Delhi
Benito Juarez Road, Dhaula Kuan, New Delhi
New Delhi -110021

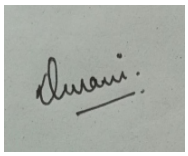
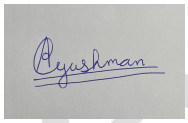
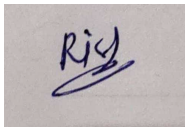
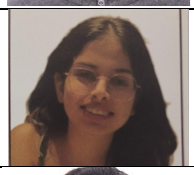
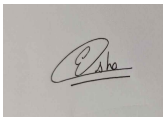
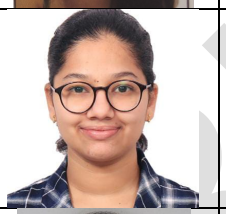
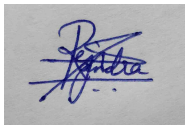
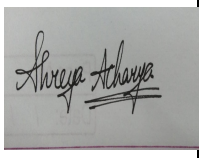
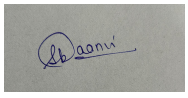
SRIVIPRA PROJECT 2025

Title : Pan-Cancer Analysis Of DNA Repair Genes Reveals Interesting Relationship Between Their Expression And Patient Survival.

Name of Mentor: Dr. Ravindra Varma Polisetty **Name of Department: Biochemistry** **Designation: Assistant Professor**	

List of students under the SRIVIPRA Project

S.No	Photo	Name of the student	Roll number	Course	Signature
1		Kritika Chaturvedi	1323035	BSc (Hons) Biological Sciences	
2		Sukriti Sitaraman	1323053	BSc (Hons) Biological Sciences	

3		Durani Sikha Kakoty	1323001	BSc (Hons) Biological Sciences	
4		Ayushman Saini	1223012	BSc (Hons) Biochemistry	
5		Nilesh Tyagi	1224036	BSc (Hons) Biochemistry	
6		Rishabh Garg	1224008	BSc (Hons) Biochemistry	
7		Esha Mulchandani	1224040	BSc (Hons) Biochemistry	
8		Sandra Reji	1224005	BSc (Hons) Biochemistry	
9		Shreya Acharya	1224015	BSc (Hons) Biochemistry	
10		Saanvi Baweja	1224043	BSc (Hons) Biochemistry	

Signature of Mentor

Certificate of Originality

This is to certify that the aforementioned students from Sri Venkateswara College have participated in the summer project SVP-2515 titled- **"Pan-Cancer Analysis Of DNA Repair Genes Reveals Interesting Relationship Between Their Expression And Patient Survival"**.The participants have carried out the research project work under my guidance and supervision from 1st July, 2024 to 30th September 2025. The work carried out is original and carried out in an online/offline/hybrid mode.

P. Ravindra Varma

Signature of Mentor

Acknowledgements

We would like to extend our deepest gratitude to **Dr. Ravindra Varma** of the Biochemistry Department at Sri Venkateswara College, University of Delhi, for providing us with the invaluable opportunity to undertake this summer internship project.His unwavering guidance, mentorship, and support have been instrumental in our academic growth and development.

We are also profoundly thankful to the entire Biochemistry Department team for their technical expertise and unwavering assistance throughout the project. Their patience, willingness to share their knowledge, and collaborative spirit have been invaluable assets to our research endeavours.

We would also like to express our sincere regards to our esteemed principal, Prof. Vajala Ravi, for conducting SRI- VIPRA and for his unwavering support and to make this research project a reality. His commitment to fostering a conducive environment for academic excellence has been instrumental in our success."

Also we would like to express our heartfelt gratitude to co-ordinators who contributed to the successful conduction of SRI VIPRA.

Collective contributions from the intern students comprising of Kritika Chaturvedi, Sukriti Sitaraman, Durani Sikha Kakoty, Ayushman Saini, Rishabh Garg, Nilesh Tyagi, Esha Mulchandani, Sandra Reji, Shreya Acharya, Saanvi Baweja are greatly acknowledged, each member equally helped in successfully completing the project reflecting their diligence and significant efforts.

TABLE OF CONTENTS

1. ABSTRACT

This study presents a pan-cancer analysis of DNA repair genes (DRGs) to assess their involvement in cancer progression and prognosis. From multiple databases and literature sources, 269 validated DRGs were identified and classified into major repair pathways. Using the GEPIA2 platform, expression and survival data from TCGA and GTEx were analyzed across 33 cancer types. Differentially expressed genes ($|\log_2 FC| > 1$, $p < 0.05$) were further evaluated through Kaplan–Meier survival analysis.

Most DRGs showed elevated expression across cancers, with 157 upregulated and 25 downregulated genes. Twenty-four were significantly overexpressed in multiple cancer types, suggesting enhanced repair activity supports tumour survival under genomic stress. Elevated expression of several DRGs correlated with poor prognosis, indicating possible "DNA repair addiction" in tumor cells. However, gene-specific patterns, such as the prognostic significance of FANCI in several cancers, highlight that DNA repair functions are context-dependent and vary with cancer type and molecular environment.

2. INTRODUCTION

In order for living organisms to continue their existence and properly perform, their genomes should be in their normal state. DNA-damaging agents of both endogenous and exogenous types, such as chemical mutagens, ultraviolet (UV) radiation, ionizing radiation, reactive oxygen species generated during metabolism, and incorrect replication, are always present in cells. cells have introduced a complex network of DNA repair mechanisms to "fight" these hazards. The genes of DNA repair (DRGs), which are responsible for proteins that can identify, signal, and repair the different kinds of DNA damage, are the main components of DNA repair which make the accurate duplication of genetic information in cell division possible. These mechanisms function in conjunction with one another to perform the coordination of these processes. The mutation process can be initiated by even one unrepaired lesion which in turn may lead chromosomal instability, or cell death, thus, if the DNA repair is inefficient.

Under regular conditions, DRGs protect the genome through a multitude of paths. As an illustration, the base excision repair (BER) pathway, which is carried out by genes like OGG1 and XRCC1, is a method of repairing small base modifications, whereas the nucleotide excision repair (NER) pathway with XPA and ERCC1, gets rid of DNA fragments that are tightly bound due to UV light. The mismatch repair (MMR) system which is under the control of MLH1, MSH2, and MSH6, fixes the errors of replication like base mismatches and insertion–deletion loops. The more serious damage, for example, double-strand breaks are fixed either by homologous recombination (HR) that is facilitated by BRCA1, BRCA2, and RAD51 or by non-homologous end joining (NHEJ) which depends on KU70/80 and DNA-PKCs. Altogether, these pathways constitute a strong defense network, and their correct operation is mandatory for genome stability.

In case of mutations or epigenetic silencing of DRGs, the repair system fails, leading to the presence of damaged DNA. As a result, these mutations will pile up over time and this will promote genomic instability, which is termed as a feature of cancer. The genes that are responsible for DNA repair and are defective, are the main reason for the occurrence of both hereditary and sporadic cancers, besides other factors. A good example would be the germline mutations in BRCA1 and BRCA2 that remarkably elevate the chances of breast and ovarian cancers, whereas changes in MLH1 or MSH2 make the person vulnerable to Lynch syndrome and colorectal cancer. In the same way, modifications in TP53 (produces p53 protein), known as the "guardian of the genome", are seen in more than half of all types of human cancers. The instability of cancer genomes is, in fact, a major source of clinical complications. However,

one of the important positive effects that this phenomenon brings is the enhanced sensitivity of cancer cells to DNA-targeting agents. Cancer cells with defects in the DNA repair gene groups are unable to do DNA-repair sufficiently after therapy-induced damage, so they are selectively attacked by DNA-damaging drugs. This statement is especially applicable to those agents that lead to double-strand breaks (DSBs), since their repair is mostly dependent on the HR and NHEJ pathways. The first group of DNA-targeting agents includes non-covalent DNA-binding agents, while covalent DNA-reactive agents represent the second group. Examples of the former are minor groove binders and intercalating agents, whereas alkylating agents and platinum-based drugs are illustrators of the latter. Although these drugs present with some setbacks like resistance to therapy and the risk of secondary carcinogenesis, they still hold a significant place in the current cancer treatment regimen, especially in combination with targeted therapy or immunotherapy. Doctors can increase the effectiveness of these drugs and at the same time reduce the harm to normal cells by identifying the DNA repair defects in cancer cells, such as BRCA1/2 mutations, and applying the drugs accordingly. Thus, the relationship of cell damage due to errors in DNA repair along with DNA-targeting drugs illustrates the vulnerabilities as well as the ability of cancer cells to adjust. Knowledge of these interactions has enabled the coming of new methods of treatment, for instance, the use of synthetic lethality strategies with PARP inhibitors in BRCA-mutated cancers and the employment of immune checkpoint inhibitors in MMR-deficient tumors. These improvements collectively indicate the double role of DNA repair genes, as the primary stabilizers of the genome in normal conditions, and as the potential sources of weakness in cancer therapy.

i) DRGs and DRG-Associated Genes:

DNA repair genes (DRGs) are the genes that code directly for the proteins, which are responsible for the recognition, signaling, and repairing of DNA damage. They are the main components in maintaining the genome, and their function is necessary to prevent mutations and chromosomal instability. There are several DRGs that have been thoroughly researched, such as BRCA1, BRCA2, MLH1, MSH2, ATM, and TP53, and each of them is a major player in the main repair mechanisms like homologous recombination, mismatch repair, and nucleotide excision repair. A mutation of one of these genes is most of the time the trigger of hereditary and sporadic cancers, thus they are the main biomarkers of genetic susceptibility and the targets for therapy.

On the other hand, DRG-associated genes refer to genes which do not directly work in DNA repair but have an impact on or regulate the activity of DRGs and their pathways. These are genes encoding proteins involved in chromatin remodeling (SMARCA4), post-translational modification of repair proteins (PARP1), as well as checkpoint signaling (CHEK1, CHEK2). Although they are not "core repair machinery," these proteins can interact with the cellular repair system in many ways, including the efficiency, regulation, or even the localization of DRG proteins to the damage sites. In conclusion, the defects in DRG-associated genes can unintentionally affect DNA repair capacity, and as a result, those cells with genomic instability can become sources for cancer progression or be impacted by cancer therapies in a way that those cells become more sensitive to DNA-targeting agents.

The main distinction between DRGs and DRG-associated genes is in their function: DRGs perform DNA repair as the main agents, thus they directly fix the DNA damage, while DRG-associated genes are the genes that code for the proteins which act as modulators or regulators that influence the efficiency, the accuracy, and the choice of the DNA repair pathway. Both sets of genes are important in a clinical context, as changes in either can lead to the development of cancer and impact treatment outcomes, however, DRGs are more directly linked to the repair processes while DRG-associated genes have an indirect role through regulation and pathway support.

ii) Comparative Overview of Functional and Defective DRGs:

Under normal circumstances, DNA repair genes (DRGs) maintain the stability of the genome by recognizing DNA damage, initiating a checkpoint signaling process, and thus, they allow the cells to perform a correct repair by a specialized pathway such as base excision repair, mismatch repair, nucleotide excision repair, and double-strand break repair through homologous recombination or non-homologous end joining. The efficiency of this kind of control ensures that the replication is accurate, that the mutations are not accumulated, and it also provides the first line of defense against malignant transformation.

The opposite picture is presented by the defective DRGs, no matter if the defects are caused by germline mutations, somatic alterations, or epigenetic silencing. The repair machinery is remodeled as a result of the defects. The failure to repair DNA will eventually lead to the accumulation of mutations, chromosomal instability, and the activation of oncogenic pathways. For instance, BRCA1/2 mutations disrupt homologous recombination and as a result the individual becomes susceptible to breast and

ovarian cancers. On the other hand, in the process of mismatch repair, the loss of MLH1 or MSH2 function leads to microsatellite instability which is a factor in colorectal carcinogenesis. Therefore, it can be said that the DNA repair genes that normally serve as genome integral guards turn into cancer development, progression, and therapeutic resistance drivers when defective.

3. METHODOLOGY

i) Compilation of gene list from different databases

To analyse DNA repair genes systematically, comprehensive literature and databases were reviewed so that an authoritative list of human DNA repair genes could be compiled. Collected DNA repair genes from multiple authoritative lists which includes UNIPROT, NCBI Gene database, and some human DNA repair genes were included from the list of wood laboratory which is maintained by MD Anderson cancer centre. Woods laboratory database is one of the most comprehensive and regularly updated resources in which there are annotations of all the genes involved in major DNA repair pathways. By using this multi-source approach, we got high level coverage and cross validation.

In initial compilation there were 1252 genes that had potential connections with DNA repair mechanisms. To refine this list we applied stringent criteria so that core DNA repair genes and DNA repair associated genes were distinguished. Core DNA repair genes were defined as those encoding proteins with direct enzymatic roles in DNA repair processes or essential regulatory functions in repair pathway coordination. Those genes which help in DNA repair mechanisms by playing indirect roles like chromatin remodeling, cell cycle checkpoint control or DNA damage signalling. Through systemic literature review and functional annotation analysis the initial set of genes were filtered and 269 genes were identified which are bonafide dna repair genes according to our criteria. Each gene was categorized according to its involvement in dna repair pathways including Base excision repair, nucleotide excision repair, mismatch repair, homologous recombination, non-homologous end joining, direct reversal, translesion synthesis, and Fanconi anemia pathways. Detailed documentation was maintained which linked each gene to relevant literature and supported its classification and pathway assessment.

ii) GEPIA2 Analysis Platform

GEPIA2(gene expression profiling interactive analysis 2) was used across different cancer types for each dna repair gene for survival and expression analysis. GEPIA2 is a comprehensive web server which provides access to RNA sequencing data of the cancer genome atlas (TCGA) and GTEx (genotype tissue expression). GEPIA2 represents a significant advancement over

previous analysis platforms, offering enhanced functionality for large-scale expression profiling and interactive analysis. The platform's integration of TCGA tumor samples with GTEx normal tissue samples provided essential control data for determining the statistical significance of expression alterations.

iii) Gene expression analysis

GEPIA2 platform enabled systematic analysis of 269 DNA repair genes across 33 cancer types listed in TCGA. Individual gene analysis was performed to identify patterns of expression dysregulation. Expression analysis involved comparison between gene expression levels in tumor and normal tissue samples. Box plot analysis provided visual representation for expression distributions of every gene, which made identification of expression changes in both magnitude and consistency possible. We classified DNA repair genes into significantly upregulated or downregulated based on whether or not they meet the statistical criteria (fold-change | log2FC|& gt;1 and p<0.05 to identify significantly altered genes). Then among these statistically upregulated/downregulated genes those showing significant expression dysregulation in (≥17) cancer types among the 33 cancer types listed in TCGA, were identified. During the analysis, DNA repair gene expression levels were measured as log(TPM+1) on the Y- axis. The key parameters had following cutoff values (log2FC cutoff = 1) and (p-value cutoff= 0.05) while 0.4 jitter size was applied. The study matched both TCGA normal data and GTEx data.

iv) Survival analysis

Survival analysis was a critical component in comprehensive assessment for clinical significance of DNA repair genes. We employed log-rank tests with Kaplan-Meier survival analysis to evaluate prognostic

impact of DNA repair gene expression levels. The overall survival data of TCGA patients was incorporated in the analysis, which provides adequate follow up time for meaningful survival assessment.

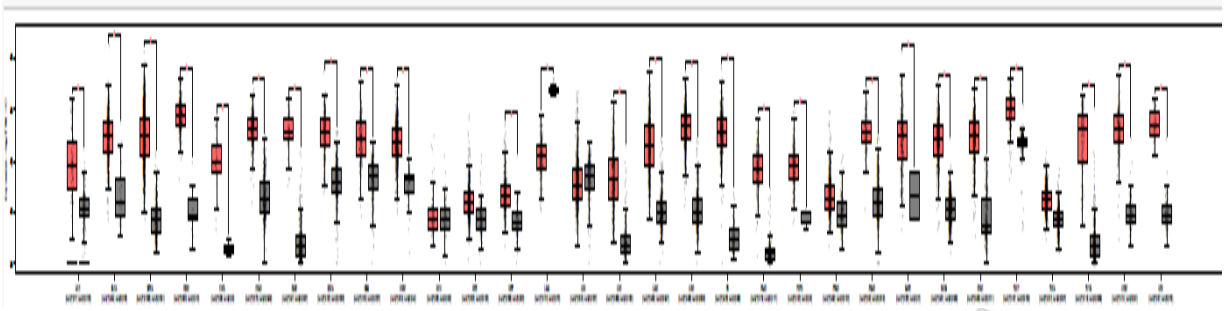
Survival analysis was performed for each of the 269 DNA repair genes across all 33 cancer types, combined as well as individually. By generating Hazard ratio (HR) and p-values, relation between expression levels and patient survival was quantified. Genes with (HR ≥ 1, log-rank p < 0.05) were associated with higher expression and reduced survival and could be said to show poor prognosis while those with (HR ≤ 1, log-rank p < 0.05) correlated to higher expression and improved survival hence, showed favorable prognosis. This systematic approach enabled identification of DNA repair genes associated with consistent prognostic significance across multiple cancer types.

4. RESULTS

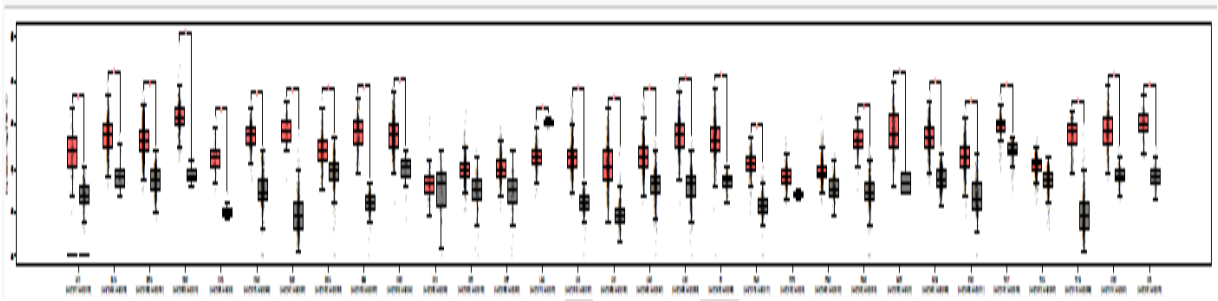
i) Pan-cancer expression analysis shows that most DNA repair genes (DRGs) are Upregulated in most cancers.

The expression analysis, done by generating pan-cancer box plots of individual DRGs using GEPIA2, showed that most of the DRGs are upregulated in cancers, with 157 of them showing upregulation, while only mere 25 showing downregulation. Of the upregulated DRGs, 24 DRGs showed upregulation across multiple cancers which were of statistical significance, while only 1 gene showed statistical downregulation for many cancers. Some of the representative box plots are given below. It may be noted that of the 33 cancers studied, box plots for two cancers, namely, Mesothelioma (MESO) and Uveal Melanoma (UVM) could not be found across all DRGs.

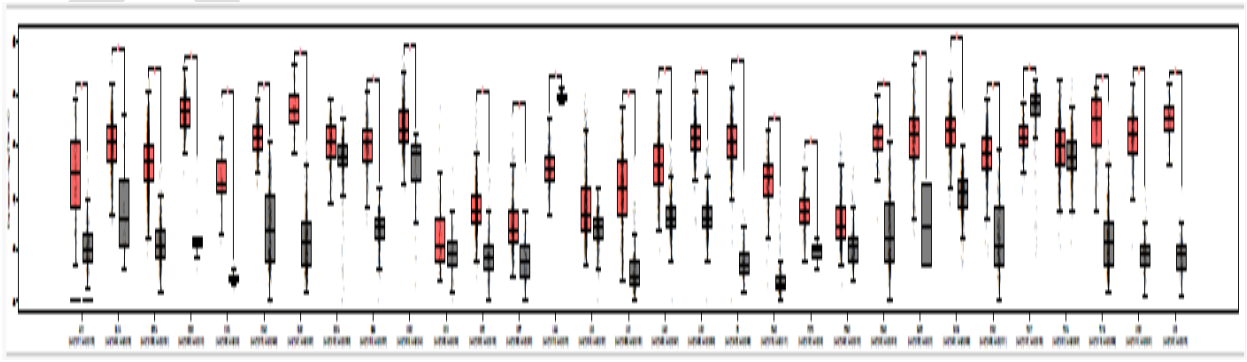
1. **Fig1: Box plots showing Pan-cancer expression for the gene UBE2T. The gene showed upregulation that was of statistical significance across 25 different types of cancers. (p-value = 0.05)**



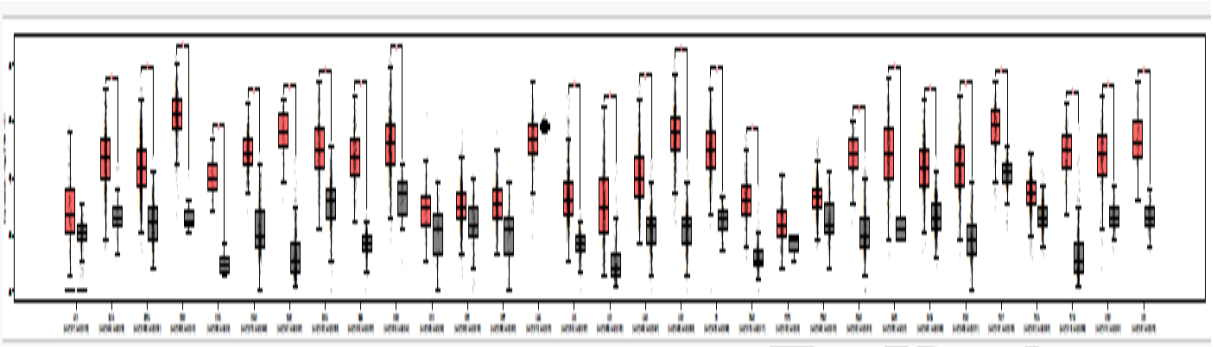
2. **Fig 2: Box plots showing Pan-cancer expression for the gene RNASEH2A. The gene showed upregulation that was of statistical significance across 24 different types of cancers. (p value = 0.05)**



3. **Fig 3: Box plots showing Pan-cancer expression for the gene PTTG1. The gene showed upregulation that was of statistical significance across 24 different types of cancers. (p value = 0.05)**



4. **Fig 4: Box plots showing Pan-cancer expression for the gene MCM2. The gene showed upregulation that was of statistical significance across 23 different types of cancers. (p value = 0.05)**



5. **Fig 5: Box plots showing Pan-cancer expression for the gene NUDT1. The gene showed upregulation that was of statistical significance across 23 different types of cancers. (p value = 0.05)**

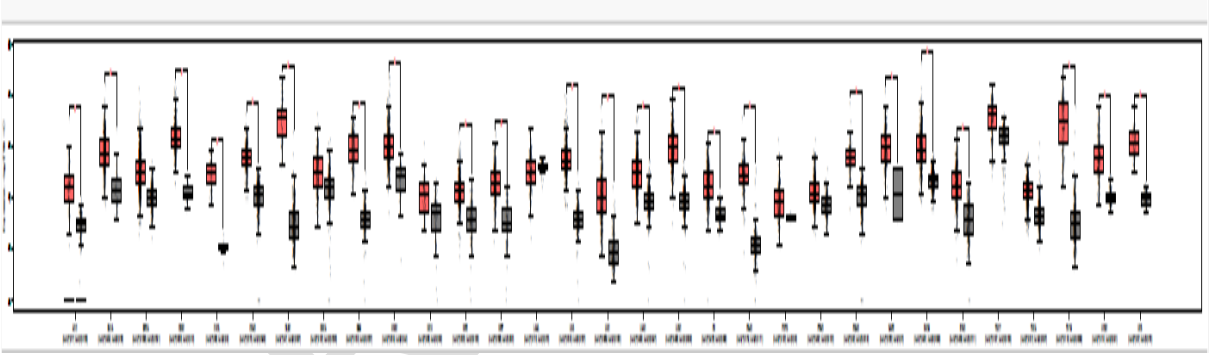


Table-1: Pan-Cancer Expression Analysis of DNA Repair Genes

S.No	Gene Symbol	Statistically Upregulated in cancers (out of 33 cancers)	Gene Symbol	Statistically Down regulated in cancers (out of 33 cancers)
1	UBE2T	25	NEIL1	19
2	RNASEH2A	24		
3	PTTG1	24		

4	MCM2	23		
5	UHRF1	23		
6	NUDT1	23		
7	HELLS	22		
8	RHNO1	22		
9	EXO1	22		
10	FANCI	22		
11	RMI2	22		
12	CHEK1	22		
13	FANCA	22		
14	RECQL4	22		
15	RAD51	21		
16	BLM	21		
17	FEN1	21		
18	GAPDH	21		
19	TIMELESS	19		
20	RAD54L	19		
21	RPA3	19		
22	MSH2	18		
23	BRCA1	17		
24	RFC2	17		
25	FANCD2	17		

These observations show that majority of the DRGs may have an oncogenic property.

ii) Pan-cancer survival analysis

The pan-cancer survival analysis of DRGs was done by using GEPIA2, wherein the overall effect of each gene was studied across 33 different cancers taken together. The plots showed that, of 276 DRGs, 121 showed low survival due to high expression and 60 showed low survival due to low expression. The remaining genes showed no significant correlation, owing to either an insignificant log rank p value (i.e., greater than 0.05), or an overlap between the plots of high and low expression. The pan-cancer plots of some representative DRGs are given below:

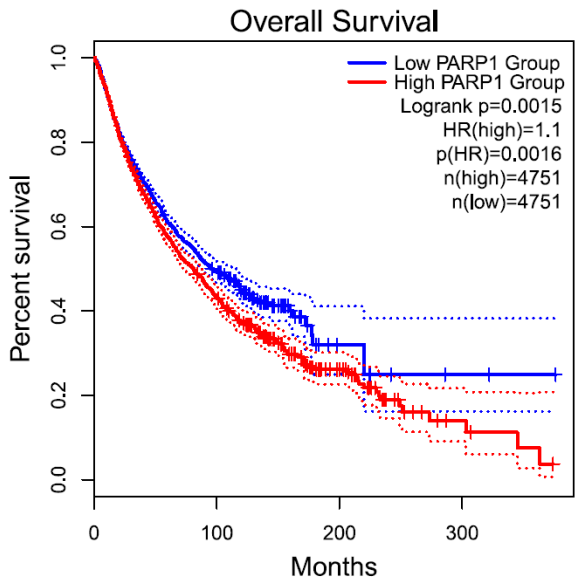
a) Low survival due to high expression

Table-2A:

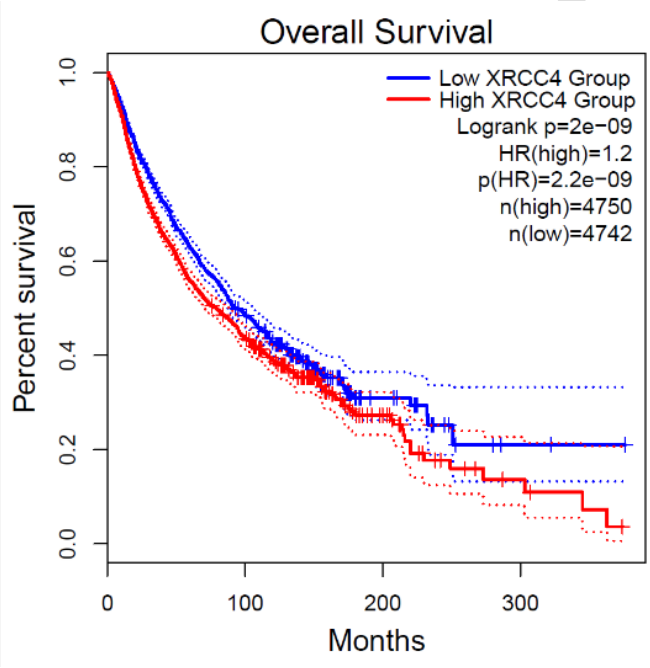
List of DNA Repair Genes showing low survival due to high expression across different cancers.

Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol
NSMCE2	FANCD2	HLTF	CHEK1	DDX5
HELLS	NABP1	DMC1	CHEK2	TIMELESS
HOTAIRM1	PELI1	SMC1A	FANCA	RAD18
PPP1R15A	FANCE	POLD3	FANCG	TP53
PARP9	POLM	SLX1A	POLQ	TP73
ZNF384	UBE2T	GEN1	RAD1	CREB1
PHF5A	POLE3	FAAP24	RAD51D	HSPA1A
RHNO1	FANCI	FAAP100	RECQL4	HDAC2
SPRTN	RTEL1	POLA1	RNASEH2A	FEN1
MSH5	APEX2	RPA4	MYB	GBA1

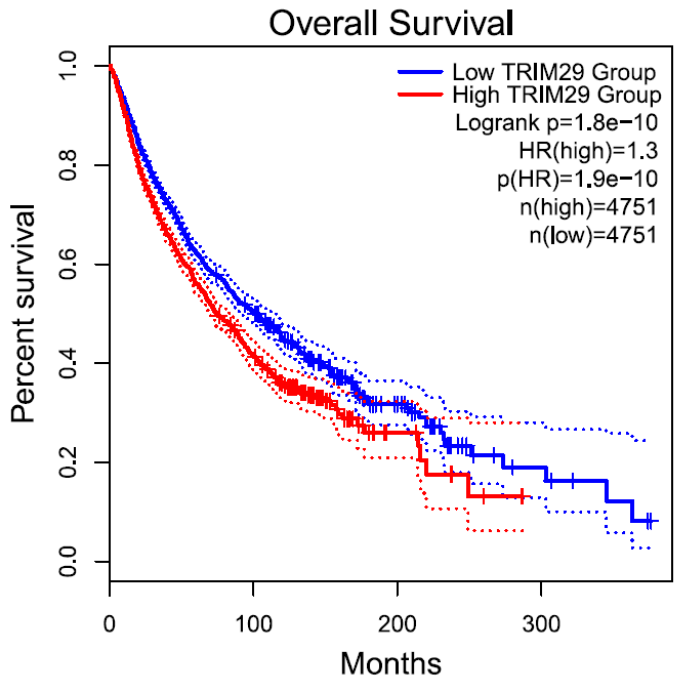
BRCA1	MAD2L2	RAG1	XRCC4	POLE
BRCA2	MCM8	TDP1	MAPK8	NCOA3
ERCC2	ZMYND8	NCOR2	MEIOB	RAD54L
GTF2H1	DDX1	FLNA	NEIL3	RAD23A
LIG 1	WRNIP1	TRIM29	TOPBP1	UBE2A
MRE11A	DCLRE1C	RPS6KB1	HDAC10	RAD54B
MSH6	UHRF1	PIM1	PMS2P3	BLM
POLD1	BRIP1	RECQL	RAD50	PRKCD
POLD2	RMI2	CFL1	SMC6	ATR
RFC2	ALKBH3	CHD1L	BARD1	GTF2H3
EXO1	PALB2	TDG	RAD9A	NUDT1
SAMHD1	PRKDC	UBE2B	RAD51	USF1
UNG	FANCC	COPS5	ATM	SRSF2
PTTG1	RIF1	ERCC6L	SLX4	GTF2H5
GAPDH				



1. **Fig 6: Pan-cancer survival plot showing the effect of high and low expression of the gene PARP1 on survival. High expression of the gene leads to significantly low percent survival (Logrank p = 0.0015), becoming as low as nearly 0.1% survival at 300 months.**



2. **Fig 7: Pan-cancer survival plot showing the effect of high and low expression of the gene XRCC4 on survival. High expression of the gene leads to significantly low percent survival (Logrank p = $2*10^{-9}$), becoming as low as nearly 0.1% survival at 300 months.**



3. **Fig 8: Pan-cancer survival plot showing the effect of high and low expression of the gene TRIM29 on survival. High expression of the gene leads to significantly low percent survival ($1.8*10^{-10}$), becoming as low as nearly 0.14% survival at 300 months.**

4.

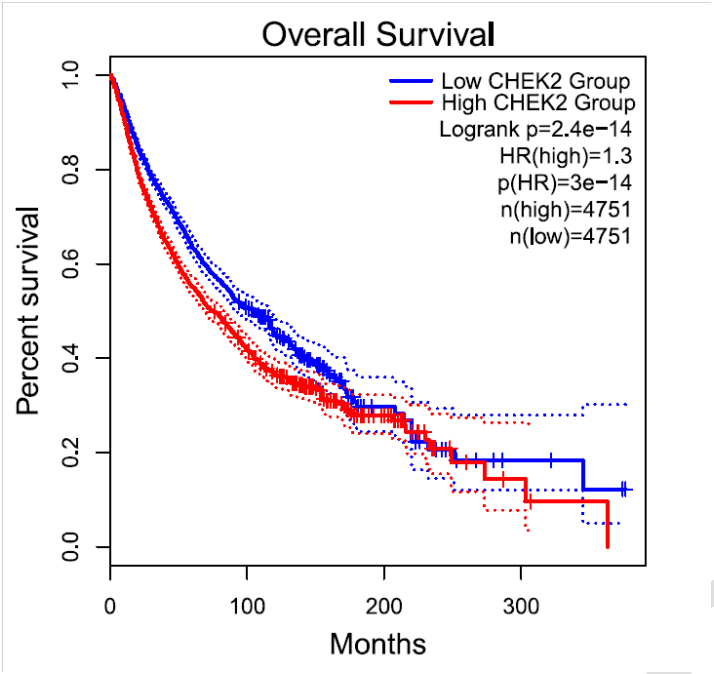
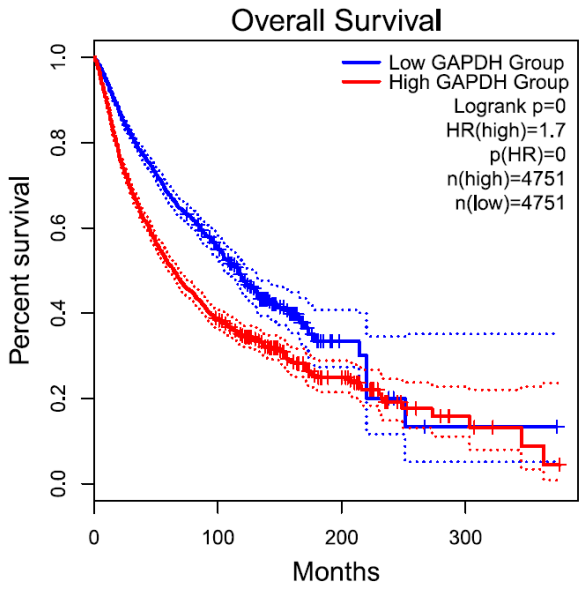


Fig 9: Pan-cancer survival plot showing the effect of high and low expression of the gene CHEK2 on survival. High expression of the gene leads to significantly low percent survival ($2.4*10^{-14}$), becoming as low as nearly 0.1% survival at 300 months.



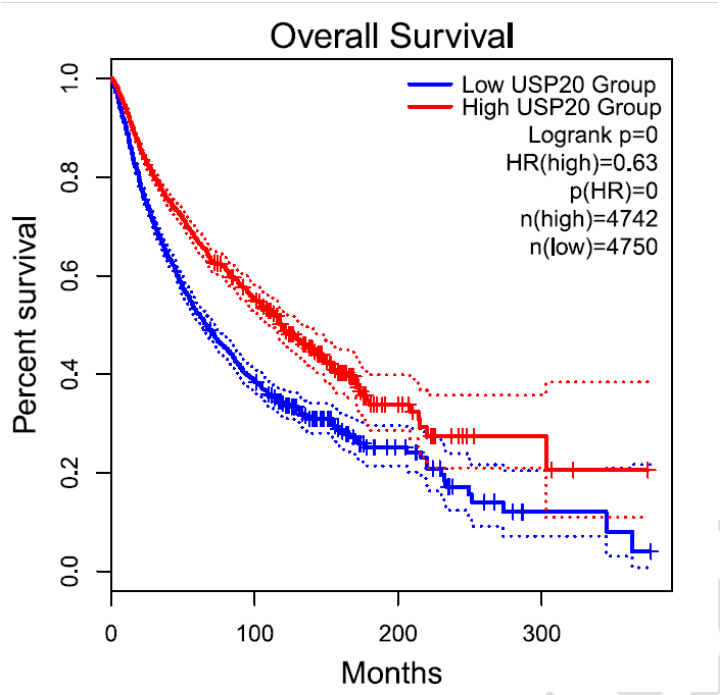
5. **Fig10: Pan-cancer survival plot showing the effect of high and low expression of the gene GAPDH on survival. High expression of the gene leads to significantly low percent survival (Logrank p = 0), becoming as low as nearly 0.2% survival at 300 months.**

Low survival due to low expression

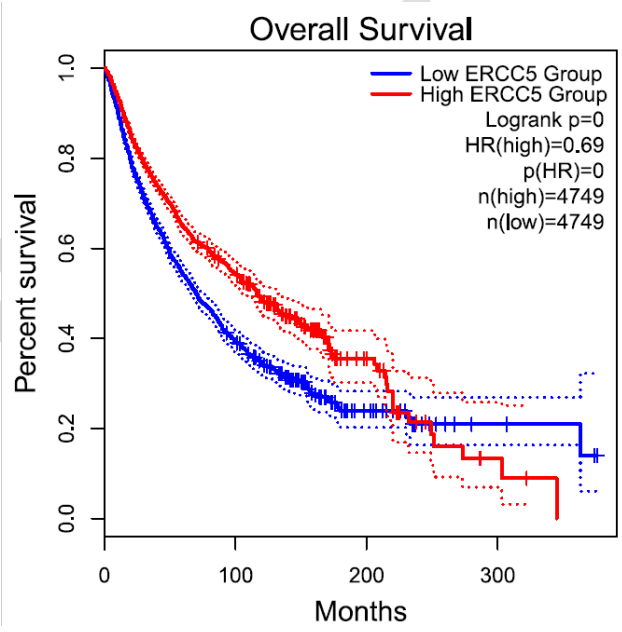
Table-2B:

List of DNA Repair Genes showing low survival due to low expression across different cancers.

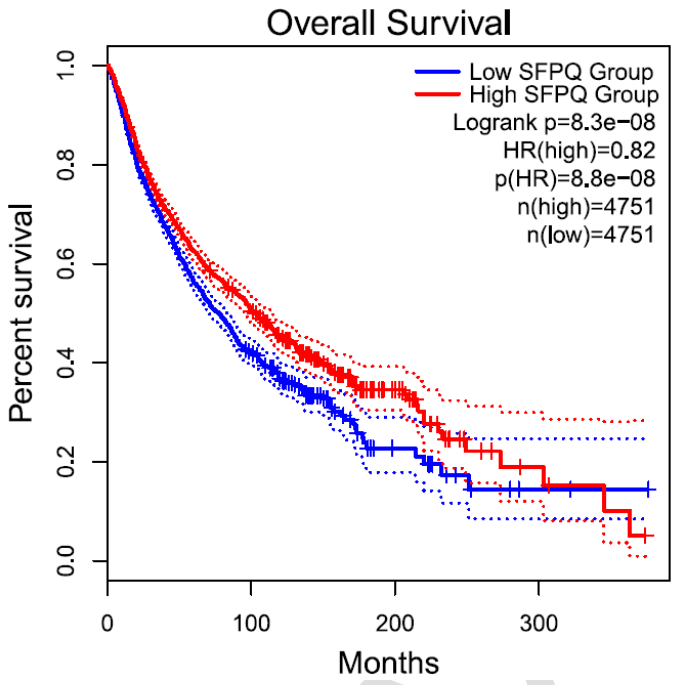
Gene Symbol	**Gene Symbol**	**Gene Symbol**	**Gene Symbol**
USP20	TREX1	FANCF	XRCC6
PHF2	PDS5B	HIC1	XRCC5
PMVK	SMARCAL1	RPA3	RPA2
GTF2H2C	POLK	MLH3	MSH3
ERCC5	POLL	PARP1	KDM6A
FUS	INO80	POLB	PNKP
MCM2	POLI	RAD17	REV3L
MLH1	TREX2	RAD51C	AKT3
MPG	MGMT	RECQL5	NEIL2
SFPQ	TP53BP1	ATF4	APTX
XRCC1	SPIDR	DDB1	RNASEH2B
FAN1	HELQ	ZSWIM7	SIRT1
NEIL1	SWI5	REV1	MEAF6
KAT5	SWSAP1	ENDOV	XAB2
FBXW7	ERCC6		



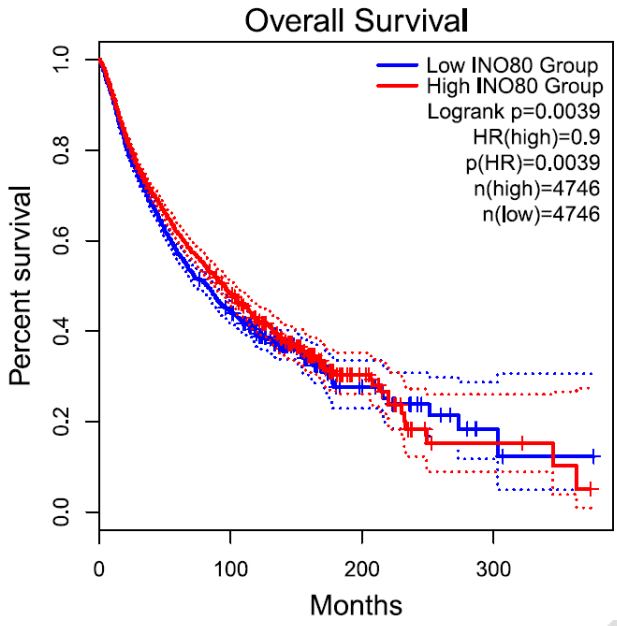
1. **Fig11: Pan-cancer survival plot showing the effect of high and low expression of the gene USP20 on survival. Low expression of the gene leads to significantly low percent survival (Logrank p = 0), becoming as low as nearly 0.12% survival at 300 months.**



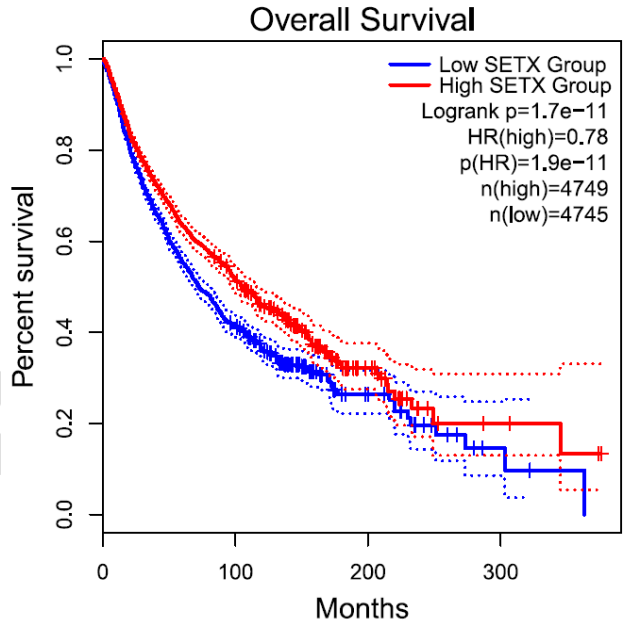
2. **Fig 12: Pan-cancer survival plot showing the effect of high and low expression of the gene ERCC5 on survival. Low expression of the gene leads to significantly low percent survival (Logrank p = 0), becoming as low as nearly 0.2% survival at 300 months.**



3. **Fig 13: Pan-cancer survival plot showing the effect of high and low expression of the gene SFPQ on survival. Low expression of the gene leads to significantly low percent survival (Logrank p = $8.3*10^{-8}$), becoming as low as nearly 0.18% survival at 300 months.**



4. **Fig 14: Pan-cancer survival plot showing the effect of high and low expression of the gene INO80 on survival. Low expression of the gene leads to significantly low percent survival (Logrank p = 0.0039), becoming as low as nearly 0.11% survival at 300 months.**



5. **Fig 15: Pan-cancer survival plot showing the effect of high and low expression of the gene SETX on survival. Low expression of the gene leads to significantly low percent survival (Logrank p = $1.7*10^{-11}$), becoming as low as nearly 0.09% survival at 300 months.**

The overall pan-cancer survival analysis of the DRGs show that most of the genes show low survival due to high expression, indicating the possibility of them having oncogenic properties.

iii) Survival Analysis of DRGs for individual cancers contradict the observations of the Pan-cancer survival and expression analyses

Survival analysis of each DRG for each cancer was done through plots retrieved using GEPIA2. Based on the results obtained during the pan-cancer survival analysis, it was expected that these genes would exhibit significant levels of up/downregulation in a majority of the cancers (at least 17), that would predict low survival rates. However, the results contradict his assumption, as this correlation was only observable case of the gene FANCI, which showed low survival due to high expression only in 14 cancer types. Thus, it may be understood that it is difficult to isolate a gene or even a gene panel that could be used for survival analysis and thus clinical prognoses for multiple cancers, as this would require an understanding of the role of each DRGs in the various cancer pathways, which was not done in the present study.

Additionally, the number of genes showing significant up/down-regulation were analysed for individual cancer types, as given in the table and graphs below:

Table-3: Summary of the number of DRGs showing significant up/down-regulation for individual cancers, indicating low percent survival.

Cancer Type	HIGH*	LOW*
ACC	115	
BLCA	16	12
BRCA	17	4
CESC	13	20
CHOL	4	6

COAD	9	8
DLBC		3
ESCA	3	3
GBM	6	1
HNSC	10	10
KICH	56	4
KIRC	28	85
KIRP	41	9
LAML	38	1
LGG	94	7
LIHC	32	3
LUAD	63	3
LUSC	2	5
MESO	63	11
OV	7	9
PAAD	29	11
PCPG	10	2
PRAD	14	1
SARC	51	7
SKCM	33	14
STAD	9	5
TGCT	3	

THCA	14	4
THYM	7	46
UCEC	2	6
UCS		6
UVM	28	11

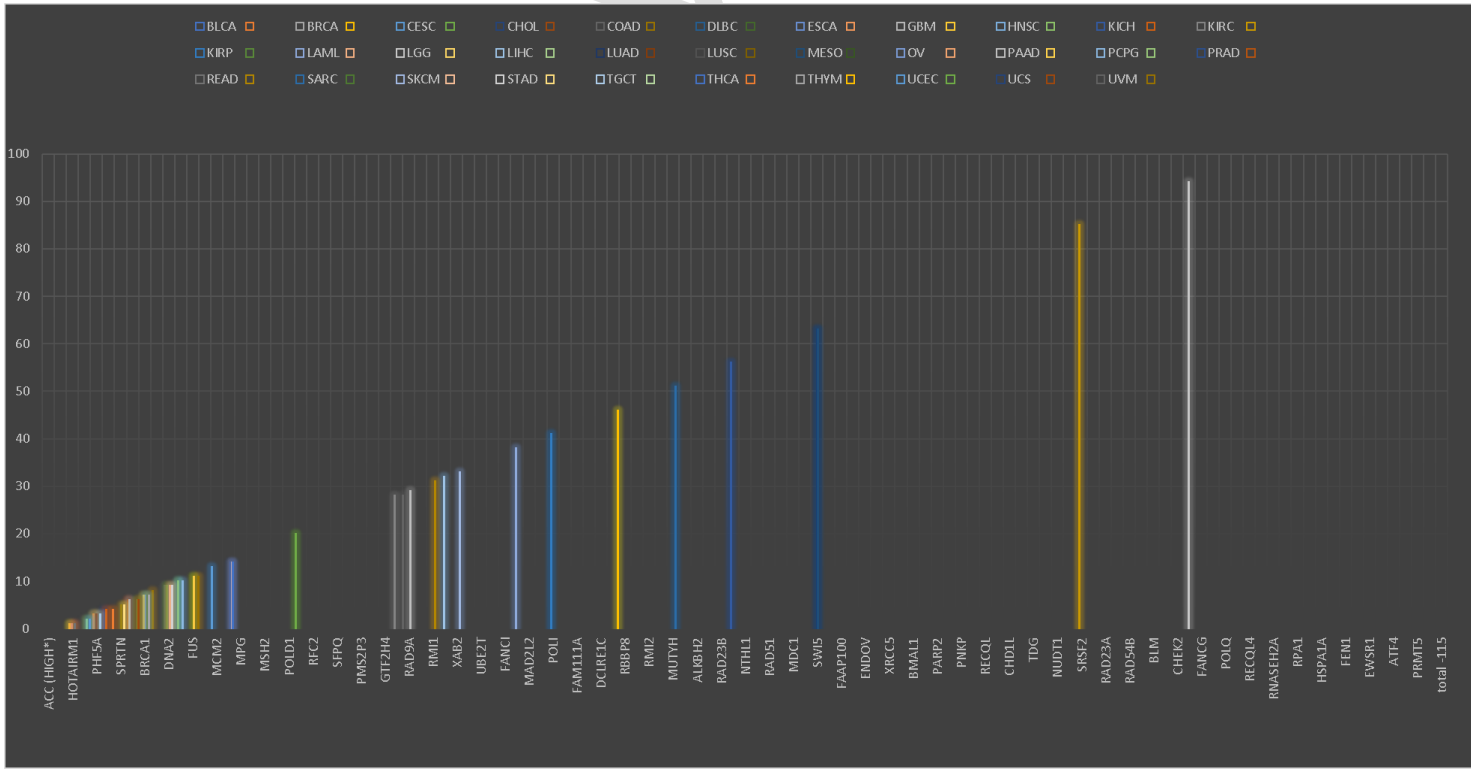


Fig 16: Plot showing the number of DRGs showing significant levels of expression in different cancer types.

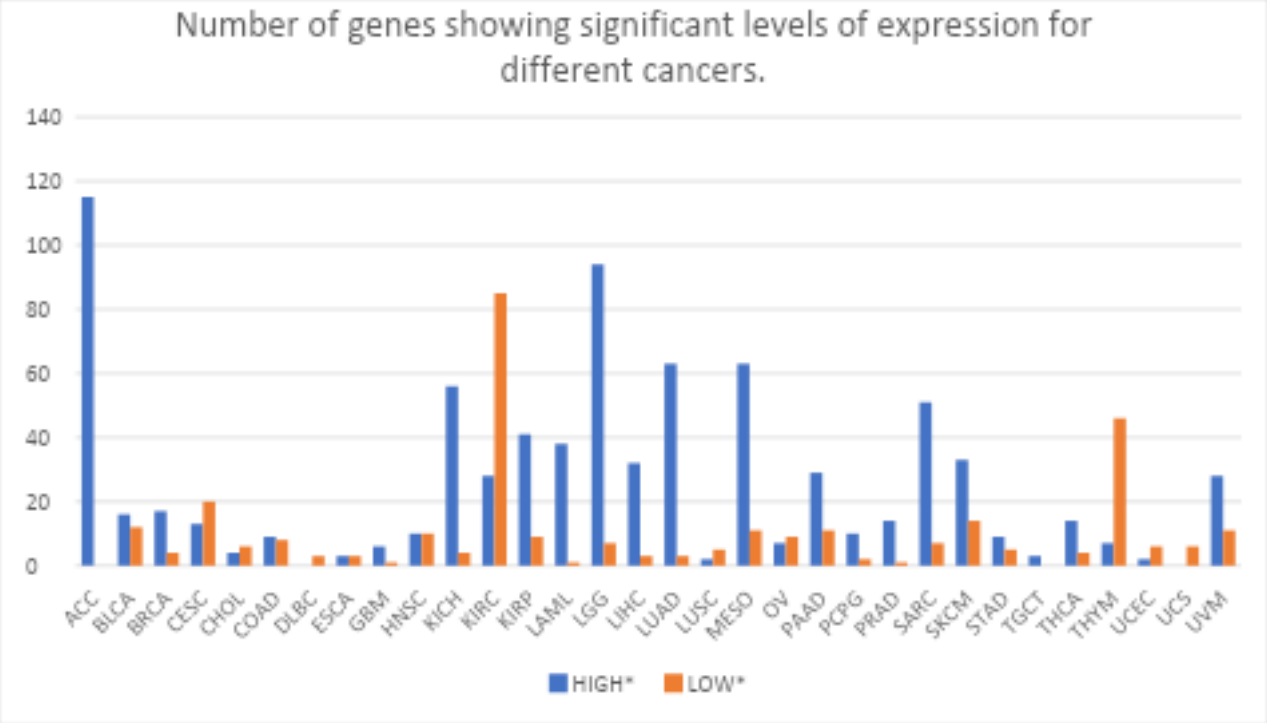


Fig 17: Plot showing number of DRGs leading to low survival in individual cancer types (due to high/low expression)

These plots show that a gene panel can be created for individual cancers, which may be utilised for making prognoses for specific cancers.

5. DISCUSSION

This study aimed to do the Pan cancer analysis of DNA Repair genes to understand their potential role in cancer progression. To achieve this, A comprehensive list of human DNA repair genes was compiled through a systematic literature review and various databases, such as UniProt, NCBI Gene, and the Wood Laboratory database (MD Anderson Cancer Centre). Out of an initial collection of 1,252 genes, rigorous selection criteria pinpointed 269 genuine DNA repair genes, categorized based on primary repair pathways including base excision, nucleotide excision, mismatch repair, homologous recombination, and others. Subsequently, The GEPIA2 platform was utilized for analyses of gene expression and survival,

combining datasets from TCGA and GTEx. Levels of expression for each DNA repair gene were analysed in tumour and normal tissues across 33 different cancer types and genes that exhibited notable dysregulation were identified by applying the thresholds of $|\log FC| > 1$ and $p < 0.05$.

Afterward, Kaplan–Meier survival analysis along with log-rank tests as conducted to evaluate prognostic significance. Hazard ratios (HR) and p-values were determined to associate gene expression with patient survival overall — Genes exhibiting ($HR \geq 1$, log-rank $p < 0.05$) were linked to elevated expression and decreased survival, indicating a poor prognosis, whereas those with ($HR \leq 1$, log-rank $p < 0.05$) were associated with higher expression and better survival, thus reflecting a favourable prognosis. This integrated bioinformatics method facilitated the cross-validation of expression patterns and underscored important DNA repair genes with possible pan-cancer prognostic relevance.

The pan-cancer expression study of DNA Repair Genes (DRGs) showed that most of these genes are elevated in expression across different cancer types. Among the 276 DRGs examined with GEPIA2, 157 were identified as upregulated, whereas merely 25 exhibited downregulation. Additionally, 24 DRGs showed statistically significant upregulation in various cancer types, indicating that abnormal activation of DNA repair pathways is a prevalent characteristic of tumorigenesis. This extensive upregulation probably indicates the elevated DNA damage load and genomic instability typical of cancer cells, requiring enhanced DNA repair functions for tumour persistence and advancement. Nonetheless, for two cancers—Mesothelioma (MESO) and Uveal Melanoma (UVM)—box plot data for specific genes were missing, which might somewhat restrict the thoroughness of the expression analysis.

The pan-cancer survival analysis also showed that elevated levels of various DRGs are linked to worse overall survival. In particular, 121 DRGs exhibited poor survival rates with high expression, whereas 60 displayed low survival with low expression. This result implies that the increased expression of various DNA repair genes may provide cancerous advantages, potentially by improving tumor cells' capacity to withstand genotoxic stress and avoid apoptosis. This behaviour is consistent with the idea of "DNA repair addiction," where cancer cells depend on overly active repair processes to maintain their elevated growth and mutation levels.

Interestingly, when the survival analysis was broadened to specific cancer types, the trend seen in the pan-cancer analysis was not consistently mirrored. A single gene, FANCI, demonstrated a strong correlation with low survival in a significant number of cancers (14 types). This mismatch underscores the complexity and context-sensitivity of DNA repair gene roles in cancer biology. The same DRG can

have either tumour-promoting or tumour-suppressive effects based on the cellular context, mutation characteristics, or the main DNA repair pathway functioning in that specific cancer type. Thus, although pan-cancer analyses provide important overarching insights, they can obscure tissue-specific molecular mechanisms and gene-pathway interactions.

6. REFERENCES

1. Kiwerska, Katarzyna, Szyfter, Krzysztof. DNA repair in cancer initiation, progression, and therapy—a double-edged sword.
2. Alhmoud, F. Jehad, Woolley, F. John, Moustafa, Al Al-Eddin, Malki, Imad Mohammed. DNA Damage/Repair Management in Cancers. 2023 Apr 23
3. Sancar, Aziz. DNA repair in humans.
4. Ronen, Amiram, Glickman, W. Barry. Human DNA Repair Genes.
5. Wood, D. Richard, Mitchell, Michael, Sgouros, John, Lindahl, Tomas. Human DNA repair genes. 2001 Feb 16
6. Wood, D. Richard, Mitchell, Michael, Lindahl, Tomas. Human DNA repair genes, 2005. 2005 Sep 4
7. http://gepia2.cancer-pku.cn/#index
8. https://www.ncbi.nlm.nih.gov
9. https://www.uniprot.org/