Am J Cancer Res 2023;13(10):4989-5004
www.ajcr.us /ISSN:2156-6976/ajcr0153020

Brief Communication

Whole Exome Sequencing reveals clinically important pathogenic mutations in DNA repair genes across lung cancer patients

Lanlan Wang[1], Yali Ma[2], Wenjie Han[2], Qiumin Yang[2], Muhammad Jamil[3]

[1]Department of Medicine, The First People's Hospital of Shangqiu, Shangqiu 476100, Henan, China; [2]Department of Oncology, Shangqiu First People's Hospital, Shangqiu 476000, Henan, China; [3]PARC Arid Zone Research Center, Dera Ismail Khan 29050, Pakistan

Received August 19, 2023; Accepted September 24, 2023; Epub October 15, 2023; Published October 30, 2023

Abstract: Lung cancer remains a substantial health challenge, with distinct genetic factors influencing disease susceptibility and progression. This study aimed to decipher the landscape of DNA repair gene mutations in Pakistani lung cancer patients using Whole Exome Sequencing (WES) and to investigate their potential functional implications through downstream analyses. WES analysis of genomic DNA from 15 lung cancer patients identified clinically important pathogenic mutations in 6 DNA repair genes, including, BReast CAncer gene 1 (BRCA1), BReast CAncer gene 2 (BRCA2), Excision Repair Cross Complementing rodent repair deficiency, complementation group 6 (ERCC6), Checkpoint Kinase 1 (CHEK1), mutY DNA glycosylase (MUTYH), and RAD51D (RAD51 Paralog D). Kaplan-Meier (KM) analysis showed that pathogenic mutations in BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes were the prognostic biomarkers of worse OS in lung cancer patients. To explore the functional impact of these mutations, we performed Reverse transcription-quantitative polymerase chain reaction (RT-qPCR) and Immunohistochemistry (IHC) analyses. Our results revealed a down-regulation in the expression of the mutated genes, indicating a potential link between the identified mutations and reduced gene activity. This down-regulation could contribute to compromised DNA repair efficiency, thereby fostering genomic instability in lung cancer cells. Furthermore, targeted bisulfite sequencing analysis was employed to assess the DNA methylation status of the mutated genes. Strikingly, hypermethylation in the promoters of BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D was observed across lung cancer samples harboring pathogenic mutations, suggesting the involvement of epigenetic mechanism underlying the altered gene expression. In conclusion, this study provides insights into the genetic landscape of DNA repair gene mutations in Pakistani lung cancer patients. The observed pathogenic mutations in BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D, coupled with their down-regulation and hypermethylation, suggest a potential convergence of genetic and epigenetic factors driving genomic instability in lung cancer cells. These findings contribute to our understanding of lung cancer susceptibility and highlight potential avenues for targeted therapeutic interventions in Pakistani lung cancer patients.

Keywords: Lung cancer, Whole Exome Sequencing, mutations, hypermethylation

Introduction

Lung cancer remains one of the most prevalent and aggressive malignancies globally, exerting a significant toll on both public health and healthcare systems [1, 2]. Its diverse clinical manifestations, coupled with the intricate interplay of genetic and environmental factors, have fueled an urgent need to unravel the genetic underpinnings of lung cancer [3, 4]. One promising avenue in this pursuit is Whole Exome Sequencing (WES), a powerful genomics technique that offers unprecedented resolution in dissecting the intricate landscape of genetic alterations underlying various cancers, including lung cancer [5, 6].

WES has revolutionized the field of genomics by enabling the comprehensive interrogation of coding regions within the human genome [7, 8]. Unlike whole genome sequencing, which sequences the entirety of an individual's DNA,

Table 1. An overview of lung cancer patient's characteristics in the present study

Sr. no	Characteristics	Sample count (n)
1	Sex	
	Male	15
	Female	0
2	Age	
	> 60	2
	< 60	13
3	Smoking	
	Non-smoker	14
	Smoker	1

WES selectively captures the protein-coding regions, known as exons [9]. This approach strikes a balance between depth of coverage and cost-effectiveness, making it an ideal tool for identifying key genetic variations underlying diseases, such as lung cancer [9, 10].

The genetic basis of lung cancer is multifaceted, encompassing an intricate interplay between inherited genetic predisposition and somatic mutations acquired during a patient's lifetime [11-14]. These somatic mutations can accrue in crucial genes involved in diverse cellular processes, including DNA repair mechanisms [15-17]. DNA repair pathways, responsible for maintaining the integrity of the genome, are particularly pertinent in lung cancer, where genomic instability and accumulation of mutations are prevalent hallmarks [18-20]. Consequently, genetic alterations in DNA repair genes can critically impact disease initiation, progression, and response to treatment.

The genetic diversity present within different populations plays a pivotal role in shaping disease susceptibility, progression, and treatment outcomes [21, 22]. Given the unique genetic makeup of Pakistani individuals, it is imperative to decipher the mutational landscape of lung cancer in this ethnic group. Past studies have demonstrated that certain genetic variants may be enriched in specific populations due to founder effects, population bottlenecks, or distinct environmental exposures [23, 24].

The current study addresses this gap by presenting a comprehensive WES analysis of lung cancer samples obtained from Pakistani patients. The utilization of WES allows for the detection of mutations with high precision, offering a nuanced view of the genetic alterations driving tumorigenesis. The current research also sought to investigate the analysis of expression and promoter methylation of the mutated genes. The study delves into the genetic alterations within six DNA repair genes, including BReast CAncer gene 1 (BRCA1), BReast CAncer gene 2 (BRCA2), Excision Repair Cross Complementing rodent repair deficiency, complementation group 6 (ERCC6), Checkpoint Kinase 1 (CHEK1), mutY DNA glycosylase (MUTYH), and RAD51D (RAD51 Paralog D) shedding light on potential drivers of lung cancer within this unique patient population.

Methodology

Sample collection and preparation

Sample selection: This research received ethical approval from the research and ethics committee of the Pakistan Agriculture Research Center (PARC), Pakistan. Tumor tissue specimens were selected from a group of 15 patients diagnosed with lung cancer and admitted to District Head Quarter (DHQ), Teaching Hospital, Dera Ismail Khan, Khyber Pakhtunkhwa (KPK) between 2019 and 2023. Thorough clinical and pathological details were gathered for every patient, encompassing factors like tumor stage, histological subtype, and previous treatment experiences. The study adhered to the Helsinki guidelines [25], and prior to acquiring the samples, informed consent was obtained from the patients. Clinical information of the subjects is given in **Table 1**.

Tissue procurement: Surgical resection of the lung cancer patients was performed to obtain representative tumor tissue. Tissue samples were immediately snap-frozen in liquid nitrogen to preserve genetic material and stored at -80°C until nucleic acid extraction.

Nucleic acid extraction: Genomic DNA was extracted from tumor tissue using the organic method [26] while RNA was extracted using the TRIzole method [27]. The quantity and quality of extracted DNA and RNA were assessed using NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and gel electrophoresis technique [28].

Library preparation and Whole Exome Sequencing

Library construction: High-quality genomic DNA samples having an A260/A280 ratio within the range of 1.8 to 2.0 were used to prepare DNA libraries for sequencing. The DNA was fragmented using the mechanical method, followed by end-repair and adapter ligation steps.

Exome enrichment: DNA libraries were subjected to exome enrichment using biotinylated capture probes designed to target exonic regions of the human genome [29]. This step selectively captures protein-coding sequences, enhancing sequencing depth and coverage.

PCR amplification: Captured DNA fragments were amplified using PCR to increase the amount of DNA available for sequencing while maintaining relative representation of different genomic regions.

Sequencing process

Sequencing platform: Prepared libraries were sequenced using a high-throughput next-generation sequencing platform (Illumina, NovaSeq).

Paired-end sequencing: Paired-end sequencing was performed to obtain sequence data from both ends of the DNA fragments. This strategy provides additional confidence in read alignment and variant detection.

Read length and depth: Sequencing runs were configured to generate paired-end reads with sufficient read length (150 base pairs) to ensure accurate alignment. The depth of sequencing was determined to achieve high coverage (98%) across exonic regions.

Data generation: The sequencing run produced raw data in FASTQ format, containing the sequences of each read along with corresponding quality scores.

Bioinformatics analysis

Data preprocessing: Raw sequencing data underwent preprocessing steps, including adapter trimming and quality filtering. Cutadapt software was used in this study for this purpose [30].

Read alignment: Processed reads were aligned to the reference human genome (GRCh38) using alignment algorithms like Burrows-Wheeler Aligner (BWA) [31].

Variant calling: Variant calling software Genome Analysis Toolkit (GATK) was used to identify (Single Nucleotide Variants) SNVs and small indels [32]. The variant calling process involved base quality recalibration and local realignment around indels.

Variant annotation: Detected variants were annotated using the ClinVar database to provide clinical significance of the observed variants.

Mutational frequencies analysis

The Genome Aggregation Database (GnomAD) [33] is a comprehensive resource that aggregates and harmonizes genetic variant data from diverse human populations. It compiles genomic information from thousands of individuals, making it a valuable tool for researchers and clinicians alike. GnomAD offers a vast collection of variant frequencies, annotations, and functional predictions, aiding in the identification of rare and common genetic variations. In this study, the GnomeAD database was used to analyze the frequencies of observed mutations in the Asian population.

Kaplan-Meier analysis

The Kaplan-Meier (KM) analysis delving into mutated and wild-type patient cohorts offers valuable insights into survival disparities [34]. By meticulously tracking time-to-event outcomes, this method illuminates how genetic mutations impact patient prognosis. As the analysis progresses, survival curves for each group materialize, revealing potential distinctions in survival rates. This statistical method was applied to examine the differences in overall survival (OS) outcomes between lung cancer patients with and without genetic mutations. A $P < 0.05$ was used as the cutoff criterion for KM analysis.

Reverse transcription-quantitative polymerase chain reaction (RT-qPCR)

The RNA extracted was transcribed into complementary DNA (cDNA) using the Prime-Script

RT reagent kit (TaKaRa, Dalian, China). RT-qPCR analysis was performed on an ABI 7500 Real-Time PCR System (Applied Biosystems, USA) using the SYBR Premix Ex Taq TM II kit (TaKaRa). The expression levels were normalized to GAPDH. All experiments were independently conducted in triplicate. The 2^(-ΔΔCt) method was employed to assess the relative expression of each mutated gene [35]. A student t-test was applied to find differences in expression levels between two groups.

Receiver operating curve (ROC) generation

Based on the RT-qPCR expression data, ROC curves of mutated gene expression were generated using the SRPLOT web source (https://bioinformatics.com.cn/srplot).

Library preparation for targeted bisulfite sequencing analysis

In brief, total DNA (1 μg) was fragmented into approximately 200-300 bp fragments using a Covarias sonication system (Covarias, Woburn, MA, USA). Following purification, the DNA fragments underwent repair and phosphorylation of blunt ends using a mixture of T4 DNA polymerase, Klenow Fragment, and T4 polynucleotide kinase. The repaired fragments were then 3' adenylated using Klenow Fragment (3'-5' exo-) and ligated with adapters containing 5'-methylcytosine instead of 5'-cytosine and index sequences using T4 DNA Ligase. The constructed libraries were quantified using a Qubit fluorometer with the Quant-iT dsDNA HS Assay Kit (Invitrogen, Carlsbad, CA, USA) and sent to Beijing Genomic Institute (BGI), China for targeted bisulfite sequencing. Following sequencing, the methylation data was normalized into beta values.

Immunohistochemical (IHC) evaluation

For this study, IHC staining was performed on selective lung cancer samples having mutated and wild-type copies of the gene. Tissues were fixed in formalin and embedded in paraffin wax. Subsequently, sections of the tissues were obtained. The tissue sections underwent a sequential treatment with alcohol, starting with xylene, followed by a series of decreasing alcohol concentrations (100%, 95%, 90%, 80%, and 70%). Tissue antigen retrieval was achieved through boiling with sodium citrate buffer, while endogenous peroxidase inhibitors were added to inhibit peroxidase activity. To prevent non-specific binding, the tissue sections were blocked in 5% goat serum for 1 hour. Following that, separate drops of anti-BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D antibodies (dilution ratio: 1:300; MA1-137, MA1-137, 24291-1-AP), MA5-15145, PA5-27855, and PA5-27195) were applied onto the sections and incubated overnight at 4°C. Subsequently, secondary antibody (anti-rabbit) was added and incubated. Afterward, diaminobenzidine was introduced for color development, followed by hematoxylin re-staining. Finally, the staining results were assessed by blocking and photographing the sections.

cBioPortal analysis

The cBioPortal database stands as a pivotal platform in the realm of mutational analysis, facilitating an in-depth exploration of genomic alterations across The Cancer Genome Atlas (TCGA) samples of different cancer types [36]. With its user-friendly interface and comprehensive data integration, researchers and clinicians can gain invaluable insights into the intricate landscape of somatic mutations, copy number variations, and other genetic aberrations implicated in tumorigenesis. In the present study, we used this database to analyze clinically significant mutations across TCGA lung cancer samples.

Enrichment analysis

The MetaScape database stands as a vital resource in the realm of biological data analysis [37]. With its comprehensive collection of functional annotation and gene set enrichment tools, MetaScape empowers researchers to unravel intricate biological insights. By integrating diverse omics data and leveraging cutting-edge algorithms, it unveils complex relationships among genes, proteins, and pathways. In this study, we used this valuable resource for Gene Ontology (GO) and KEGG (Kyoto Encyclopedia of Genes and Genomes) analyses of the mutated genes. A $P < 0.05$ was used as the cutoff criterion for the functional enrichment analysis.

Drug prediction analysis

In our study, we harnessed the drug prediction feature of the DrugBank (http://www.drugbank.

ca) database [38], a comprehensive resource that provides valuable information on drug-target interactions and drug-related data. Leveraging this feature, we aimed to identify potential drugs that could target the hub genes identified in our study. By exploring the vast database of drug-target interactions, we sought to uncover drugs that may have regulatory effects on the expression of mutated genes.

Results

Sequencing and genomic variants

WES analysis produced a median of 98 million data reads per individual (with a range of 8 million to 173 million reads). This led to a median coverage of 96.6% (ranging from 92.3% to 98%) and 90% (ranging from 31.7% to 95.3%) for bases sequenced at least 10 and 50 times, respectively.

We extensively performed manual analyses on variants found within a set of 400 DNA repair genes. Variants were classified as either benign and likely benign, variant of uncertain significance (VUS), or pathogenic and likely pathogenic according to American College of Medical Genetics and Genomics (ACMG) recommendations [39].

In 400 DNA repair genes that were chosen for analyses, we identified 6 mutated genes including BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D in Pakistani lung cancer patients. A total of 42 non-silent variants were found in BRCA1 (n = 14), BRCA2 (n = 10), ERCC6 (n = 6), BRCA2 (n = 5), MUTYH (n = 5), and RAD51D (n = 2) genes (**Table 2** and **Figure 1**). Among 14 observed BRCA1 mutations, 12 mutations were benign and 2 were pathogenic (**Table 2**). Among 10 mutations observed in the BRCA2 gene, 7 mutations were benign, 1 was likely benign and 2 were pathogenic (**Table 2**). In the case of ERCC6, CHEK1, MUTYH, and RAD51D, in total 4, 4, 4, and 1 mutations were benign while 2, 1, 1, and 1 mutations were pathogenic, respectively (**Table 2**).

Clinically important genomic variants in BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes

Pathogenic variants are of great clinical significance as they play a crucial role in disease development [40]. In the current study among DNA repair-associated mutated genes, two clinically important pathogenic variants were observed in each BRCA1 (p.Arg1699Trp and p.Asp1692His), BRCA2 (p.Leu24Ter and p.Trp31Ter), and ERCC6 (p.Arg1221Ter and p.Trp31Ter) gene (**Table 2**). In the case of CHEK1 (p.Gln346X), MUTYH (p.Gln426Ter), and RAD51D (p.Trp314Ter), one clinically important pathogenic variant were observed in each of these genes (**Table 2**).

Frequencies of pathogenic mutations across Asian lung cancer patients in the GnomAD database

To assess the frequencies of pathogenic variants across lung cancer patients in Asian population, we analyzed the GnomAD database. Results showed that detected pathogenic variants in the present study, including BRCA1 (p.Arg1699Trp and p.Asp1692His), BRCA2 (p.Leu24Ter and p.Trp31Ter), ERCC6 (p.Arg1221Ter and p.Trp31Ter), CHEK1 (p.Gln346X), and MUTYH (p.Gln426Ter) had 0.000 mutational frequencies in the Asian lung cancer patients across GnomAD database. This finding underscores the distinct genetic makeup of the Asian population in relation to these specific pathogenic mutations in the context of lung cancer. Consequently, these results contribute significantly to our understanding of the genetic landscape of lung cancer within the Asian demographic and emphasize the importance of considering population-specific genetic factors in the prognosis and treatment of this disease.

Survival analysis of lung patients having pathogenic mutations in BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes

In the current research, we used KM curve analysis to assess the OS of the lung cancer patients harboring pathogenic mutations in BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes. In view of analysis results, BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D pathogenic mutation were the prognostic biomarkers of worse OS in lung cancer patients, as lung cancer patients with pathogenic mutated copies of these genes showed poor OS as compared to the other lung cancer patients with wild-type copies (**Figure 2**). This finding showed the potential role of these

Table 2. Count and types of mutations observed in BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes across lung cancer patients

Sr. no	Gene	NM:c.DNA	Protein	Nature	No. patients
1	BRCA1	NM_007294.4:c.5095C>T	p.Arg1699Trp	Pathogenic	3
2		NM_007294.4:c.5074G>C	p.Asp1692His	Pathogenic	4
3		NM_007294.4:c.5579A>C	p.His1860Pro	Benign	4
4		NM_007294.4:c.5576C>G	p.Pro1859Arg	Benign	11
5		NM_007294.4:c.5558A>C	p.Tyr1853Ser	Benign	4
6		NM_007294.4:c.5531T>G	p.Leu1844Arg	Benign	7
7		NM_007294.4:c.5411T>A	p.Val1804Asp	Benign	2
8		NM_007294.4:c.5402G>A	p.Gly1801Asp	Benign	1
9		NM_007294.4:c.5347A>T	p.Met1783Leu	Benign	13
10		NM_007294.4:c.5198A>G	p.Asp1733Gly	Benign	7
11		NM_007294.4:c.5158A>G	p.Thr1720Ala	Benign	9
12		NM_007294.4:c.5044G>A	p.Glu1682Lys	Benign	11
13		NM_007294.4:c.5024C>T	p.Thr1675Ile	Benign	4
14		NM_007294.4:c.4991T>C	p.Leu1664Pro	Benign	1
15	BRCA2	NM_000059.4:c.71T>G	p.Leu24Ter	Pathogenic	4
16		NM_000059.4:c.92G>A	p.Trp31Ter	Pathogenic	4
17		NM_000059.4:c.53G>A	p.Arg18His	Benign	11
18		NM_000059.4:c.128A>G	p.Asn43Ser	Benign	13
19		NM_000059.4:c.167A>C	p.Asn56Thr	Benign	1
20		NM_000059.4:c.223G>C	p.Ala75Pro	Benign	1
21		NM_000059.4:c.322A>C	p.Asn108His	Benign	5
22		NM_000059.4:c.502C>A	p.Pro168Thr	Benign	5
23		NM_000059.4:c.865A>C	p.Asn289His	Benign	2
24		NM_000059.4:c.868G>C	p.Val290Leu	Likely Benign	1
25	ERCC6	NM_000124.4:c.3661C>T	p.Arg1221Ter	Pathogenic	5
26		NM_000124.4:c.3445G>T	p.Glu1149Ter	Pathogenic	3
27		NM_000124.4:c.4322C>T	p.Thr1441Ile	Benign	1
28		NM_000124.4:c.4315G>C	p.Ala1439Pro	Benign	3
29		NM_000124.4:c.3965G>T	p.Gly1322Val	Benign	3
30		NM_000124.4:c.3689G>C	p.Arg1230Pro	Benign	3
31	CHEK1	NM_001274:c.1036C>T	p.Gln346X	Pathogenic	4
32		NM_001114122.3:c.1411A>G	p.Ile471Val	Benign	1
333		NM_001114122.3:c.1260A>G	p.Arg420=	Benign	1
34		NM_001114122.3:c.93T>C	p.Thr31=	Benign	8
35		NM_001114122.3:c.105C>T	p.Val35=	Benign	9
36	MUTYH	NM_001048174.2:c.1276C>T	p.Gln426Ter	Pathogenic	4
37		NM_001048174.2:c.1517G>A	p.Arg506Gln	Benign	3
38		NM_001048174.2:c.1460C>T	p.Ser487Phe	Benign	3
39		NM_001048174.2:c.930G>C	p.Gln310His	Benign	1
40		NM_001048174.2:c.929A>G	p.Gln310Arg	Benign	1
41	RAD51D	NM_002878.4:c.941G>A	p.Trp314Ter	Pathogenic	3
42		NM_002878.4:c.494G>A	p.Arg165Gln	Benign	9

BReast CAncer gene 1 = BRCA1, BReast CAncer gene 2 = BRCA2, Excision Repair Cross Complementing rodent repair deficiency, complementation group 6 = ERCC6, Checkpoint Kinase 1 = CHEK1, mutY DNA glycosylase = MUTYH, RAD51D = RAD51 Paralog D.

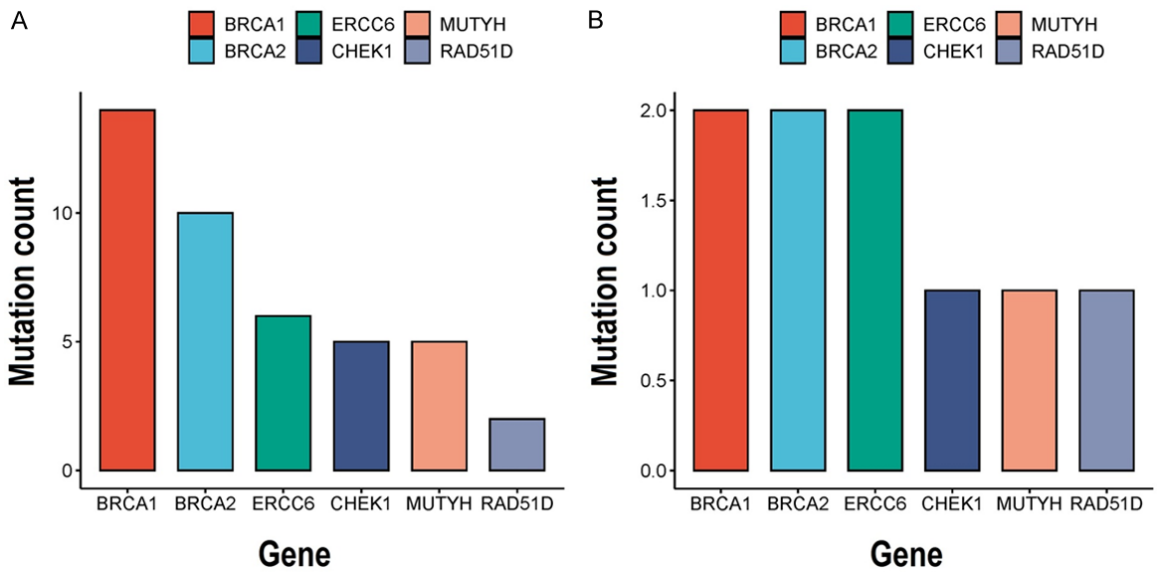


Figure 1. Total count of overall detected mutations and pathogenic mutations in six DNA repair genes (BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D) across lung cancer patients via WES. (A) An overall count of detected mutations in BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes across lung cancer patients, and (B) A count of detected pathogenic mutations in BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes across lung cancer patients. Whole Exome Sequencing = WES, BReast CAncer gene 1 = BRCA1, BReast CAncer gene 2 = BRCA2, Excision Repair Cross Complementing rodent repair deficiency, complementation group 6 = ERCC6, Checkpoint Kinase 1 = CHEK1, mutY DNA glycosylase = MUTYH, RAD51D = RAD51 Paralog D.

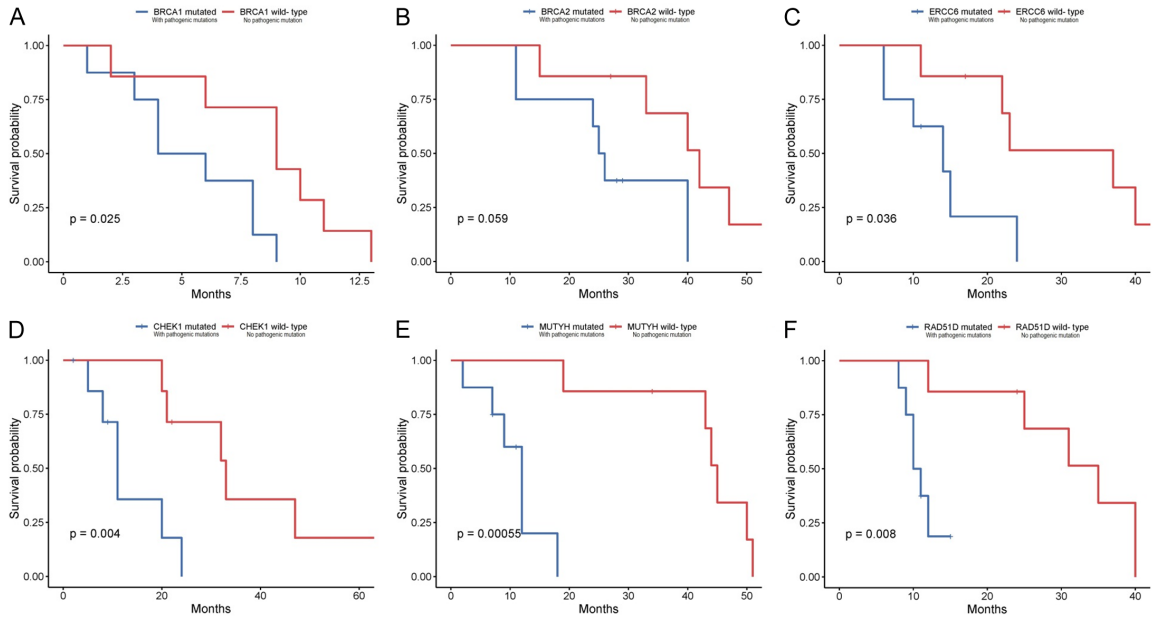


Figure 2. Kaplan-Meier survival analysis comparing two lung cancer sample groups: one harboring mutated copies of the BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes, and the other possessing the wild-type copies of these genes. BReast CAncer gene 1 = BRCA1, BReast CAncer gene 2 = BRCA2, Excision Repair Cross Complementing rodent repair deficiency, complementation group 6 = ERCC6, Checkpoint Kinase 1 = CHEK1, mutY DNA glycosylase = MUTYH, RAD51D = RAD51 Paralog D.

pathogenic mutations as prognostic biomarkers, shedding light on their significance in predicting the clinical outcomes of lung cancer patients.

Reverse transcription-quantitative polymerase chain reaction (RT-qPCR) analysis of BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes expression

Expression analysis of BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes was performed through RT-qPCR among two groups of the sample, i.e. first group consisted of lung cancer patients with wild-type genes (devoid of pathogenic mutations) and the second group of lung cancer samples which harbor pathogenic mutations in those genes. Results of the analysis revealed that BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes showed significant down-regulation in the lung cancer sample group that harbor pathogenic mutations as compared to the wild-type group of lung cancer patients which do not harbor pathogenic mutations in these genes (**Figure 3A**).

Moreover, significant sensitivity and specificity were observed in the ROC curves for BRCA1 (AUC: 0.714, *p*-value < 0.05), BRCA2 (AUC: 1.0, *p*-value < 0.05), ERCC6 (AUC: 0.671, *p*-value < 0.05), CHEK1 (AUC: 1.0, *p*-value < 0.05), MUTYH (AUC: 0.910, *p*-value < 0.05), and RAD51D (AUC: 0.671, *p*-value < 0.05) based on their expression levels, as illustrated in **Figure 3B**.

Promoter methylation analysis of BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D via bisulfite sequencing

The bisulfite-seq technique was employed in this study to analyze the promoter methylation patterns of BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes. This analysis was carried out in two distinct groups of lung cancer patients: the first group consisted of patients with wild-type genes (devoid of pathogenic mutations), while the second group included patients with lung cancer samples harboring pathogenic mutations in these genes. The results of the analysis revealed that the lung cancer patient group having pathogenic mutations exhibited lower beta values in the promoters of BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes compared to the lung cancer patient group having wild-type copies of these genes (**Figure 4**). These findings underscore the phenomenon of hypermethylation in these genes within the lung cancer patient group having pathogenic mutations (**Figure 4**).

Immunohistochemistry-based proteomic expression analysis of BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes

In our study, we conducted a proteomic expression analysis of BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D, within specific lung cancer tissue samples, which were mutated for the respective gene using the IHC technique. The outcome of this analysis unveiled a distinct pattern: the protein expression levels of these genes were noticeably reduced (staining: low) in the lung cancer sample having pathogenic mutated copies of these genes (**Figure 5**) when compared to the wild-type sample having wild-type copies of these genes (staining: high). This finding strongly suggests a correlation between the presence of mutations and the down-regulation of protein expression for these genes in lung cancer tissues.

Analysis of clinically important variants in BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes across the cancer genome atlas lung cancer samples

In our study, a comprehensive analysis of pathogenic variants within the BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes was conducted using the cBioPortal database across TCGA lung cancer samples. Intriguingly, the findings illuminated a significant disparity, as among all the observed variants in TCGA lung cancer samples, the pathogenic variants detected among Pakistani lung cancer patients in BRCA1 (p.Arg1699Trp and p.Asp1692His), BRCA2 (p.Leu24Ter and p.Trp31Ter), ERCC6 (p.Arg1221Ter and p.Trp31Ter), CHEK1 (p.Gln-346X), and MUTYH (p.Gln426Ter) were conspicuously absent within the TCGA lung cancer dataset (Supplementary Figure 1). This compelling revelation underscores the distinctiveness of these variants within the Pakistani population, highlighting the potential genetic heterogeneity that contributes to the unique landscape of lung cancer susceptibility in this specific demographic.

Enrichment analysis of BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes

Next, we performed GO and KEGG enrichment analyses. Among GO terms, BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes

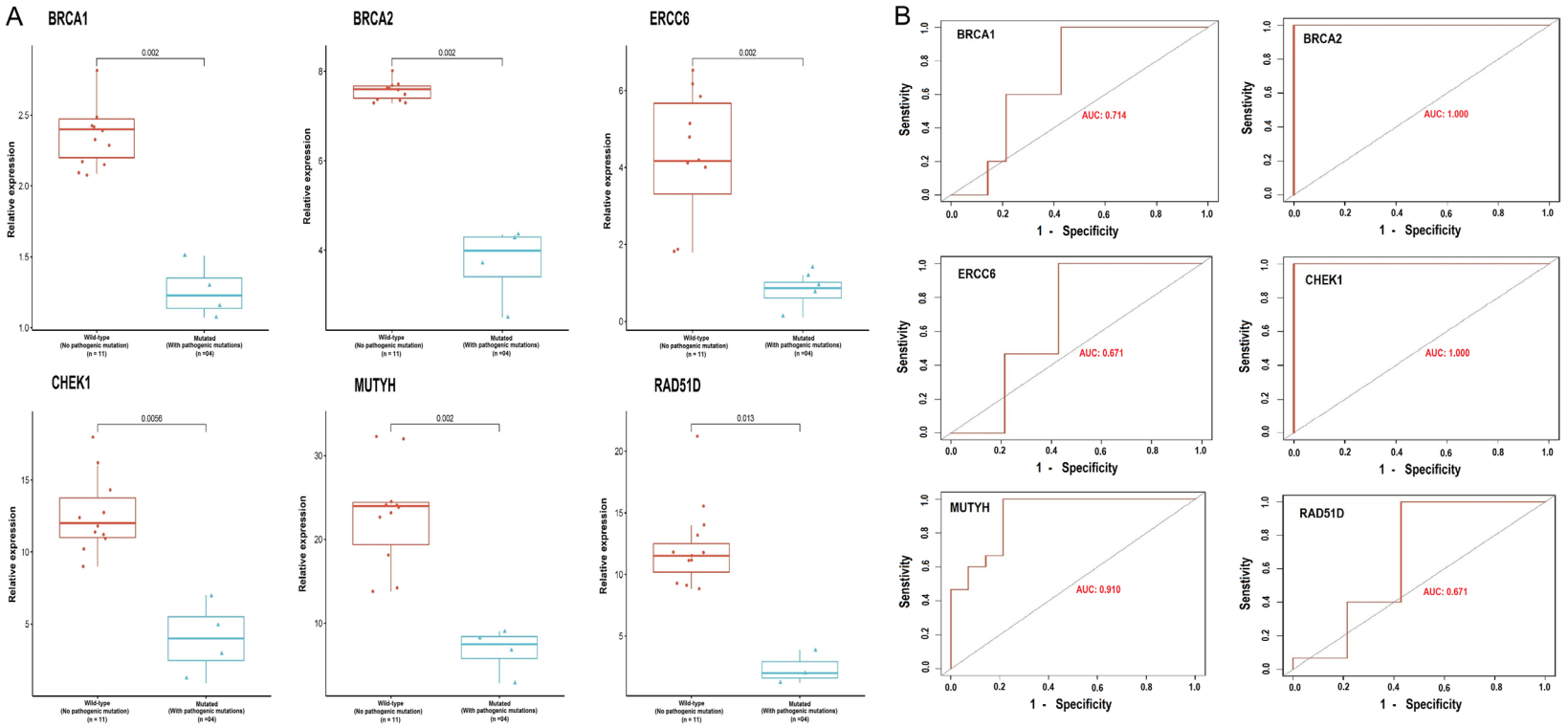


Figure 3. Relative expression and ROC curve analysis of BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes between two lung cancer sample groups: one harboring pathogenic mutated copies of the BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes, and the other possessing the wild-type copies of these genes. (A) Relative expression analysis of BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes via RT-qPCR, and (B) RT-qPCR expression-based ROC curves of BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D. A $P < 0.05$ was used as the cut-off criterion. Reverse transcription-quantitative polymerase chain reaction = RT-qPCR, AUC = Area Under the Cover, BReast CAncer gene 1 = BRCA1, BReast CAncer gene 2 = BRCA2, Excision Repair Cross Complementing rodent repair deficiency, complementation group 6 = ERCC6, Checkpoint Kinase 1 = CHEK1, mutY DNA glycosylase = MUTYH, RAD51D = RAD51 Paralog D.

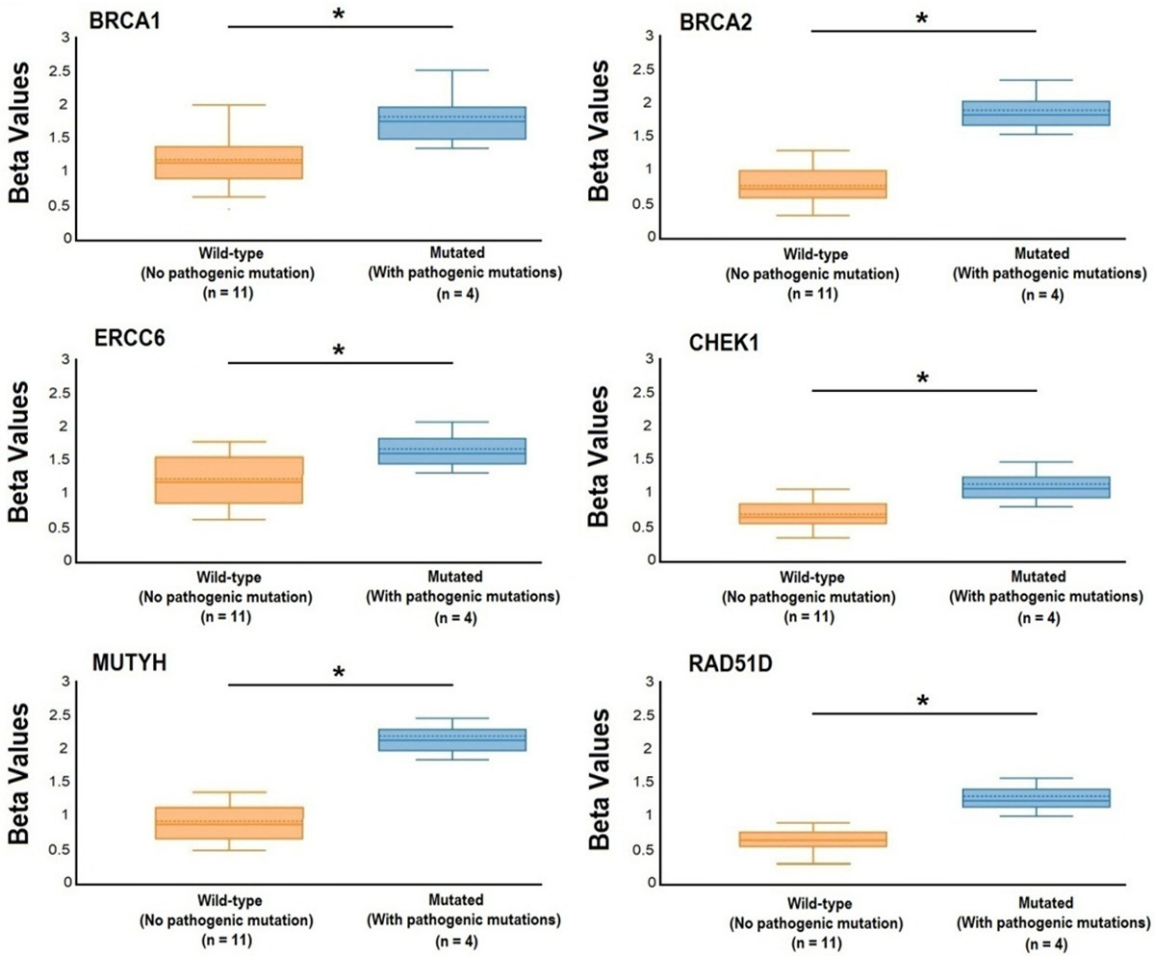


Figure 4. Targeted bisulfite sequencing-based methylation level exploration of BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes between two lung cancer sample groups: one harboring pathogenic mutated copies of the BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes, and the other possessing the wild-type copies of these genes. A $P < 0.05$ was used as the cut-off criterion. BReast CAncer gene 1 = BRCA1, BReast CAncer gene 2 = BRCA2, Excision Repair Cross Complementing rodent repair deficiency, complementation group 6 = ERCC6, Checkpoint Kinase 1 = CHEK1, mutY DNA glycosylase = MUTYH, RAD51D = RAD51 Paralog D.

were enriched in "consensed nuclear chromosome, lateral element, chromosome, telomeric region, condensed chromosonme, and synaptonemal structure" etc., CC terms (Supplementary Figure 2A), "gamma-tubulin binding, Tubulin binding, damaged DNA binding, and DNA-dependent ATPAse activity" etc., MF terms (Supplementary Figure 2B), "DNA double strand repair via homologous recombination, recombinational repair, and double strand break repair" etc., BP terms (Supplementary Figure 2C), and "homologous recombination, fanconi anemia pathway, base excision repair, and nucleotide excision repair" etc., KEGG terms (Supplementary Figure 2D).

Drug prediction analysis of BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes

In this comprehensive study, we leveraged the DrugBank database to systematically investigate potential therapeutic options for enhancing the expression of mutated down-regulated genes, including BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D. Our meticulous analysis unveiled a diverse array of promising drug candidates, including Acetaminophen, Estradiol, Cytarabine, and Bortezomib with the capacity to up-regulate the expression of these target genes (**Table 3**). These findings hold significant implications for advancing our under-

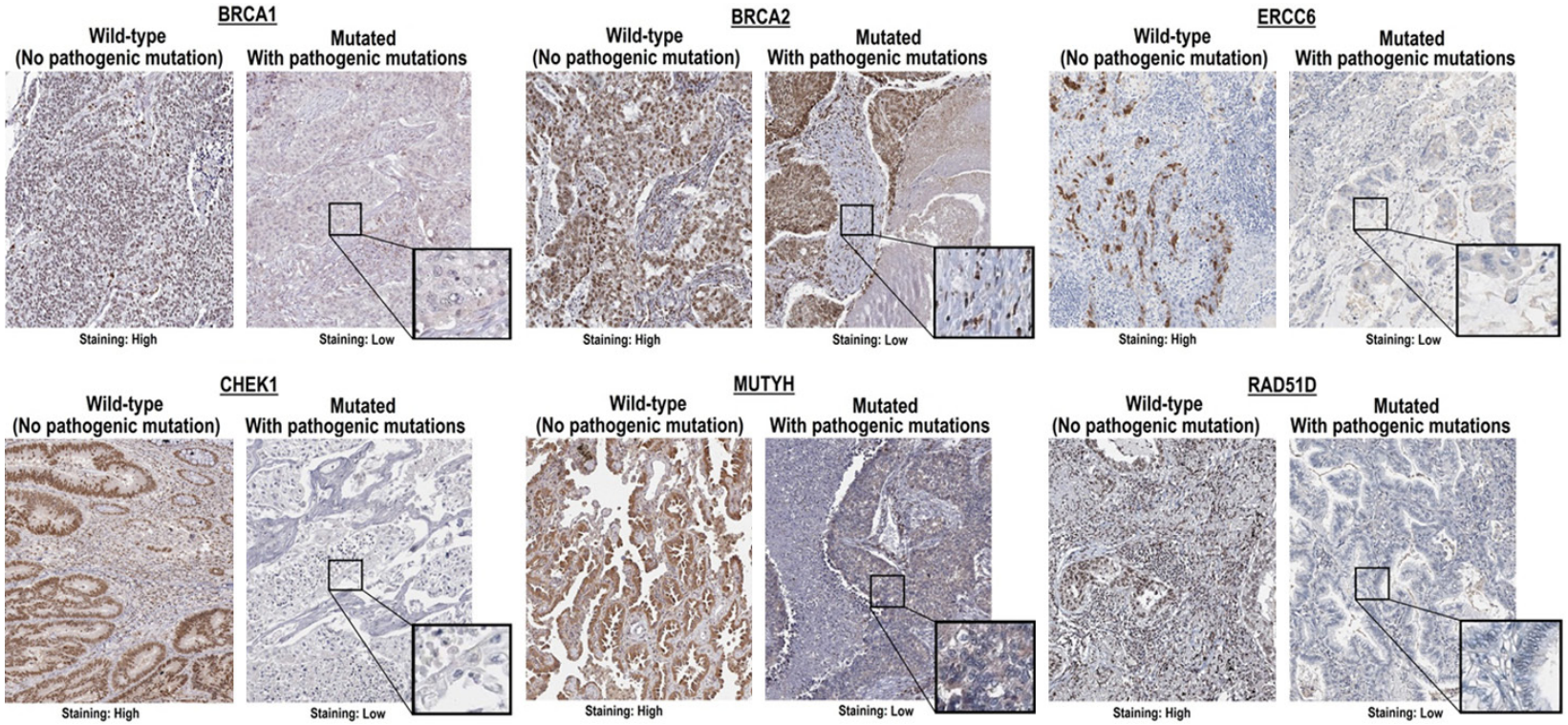


Figure 5. IHC-based proteomic expression analysis of BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes between two lung cancer samples: one harboring pathogenic mutated copies of the BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes, and the other possessing the wild-type copies of these genes. IHC = Immunohistochemistry, BReast CAncer gene 1 = BRCA1, BReast CAncer gene 2 = BRCA2, Excision Repair Cross Complementing rodent repair deficiency, complementation group 6 = ERCC6, Checkpoint Kinase 1 = CHEK1, mutY DNA glycosylase = MUTYH, RAD51D = RAD51 Paralog D.

Table 3. DrugBank-based BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D associated drugs

Sr. No	Hub gene	Drug name	Effect	Reference	Group
1	BRCA1	Acetaminophen	Increase expression of BRCA1 mRNA	A20418	Approved
		Estradiol		A21155	
2	BRCA2	Acetaminophen	Increase expression of BRCA2 mRNA	A20418	Approved
		Estradiol		A21155	
3	ERCC6	Acetaminophen	Increase expression of ERCC6 mRNA	A20418	Approved
		Cytarabine		A20508	
4	CHEK1	Acetaminophen	Increase expression of CHEK1 mRNA	A20418	Approved
		Cytarabine		A20508	
5	MUTYH	Bortezomib	Increase expression of MUTYH mRNA	A21437	Approved
		Acetaminophen		A20418	
6	RAD51D	Acetaminophen	Increase expression of RAD51D mRNA	A20418	Approved
		Estradiol		A21155	

BReast CAncer gene 1 = BRCA1, BReast CAncer gene 2 = BRCA2, Excision Repair Cross Complementing rodent repair deficiency, complementation group 6 = ERCC6, Checkpoint Kinase 1 = CHEK1, mutY DNA glycosylase = MUTYH, RAD51D = RAD51 Paralog D.

standing of potential avenues for modulating DNA repair mechanisms and may pave the way for novel therapeutic strategies in the realm of genetic stability and cancer prevention.

Discussion

In this study, we analyzed genetic variants in a set of 15 lung cancer tissue samples from the Pakistani population. Employing the WES technique, our focus was directed toward pinpointing the prevalent mutations within a comprehensive repertoire of 400 DNA repair genes. The extensive reservoir of genomic insights generated through this study holds the potential to revolutionize our existing comprehension of lung cancer, subsequently propelling the advancement of tailored therapeutic approaches for lung cancer patients.

Outcomes of this study have illuminated significant clinically important Pakistani population-specific pathogenic mutations within 6 key DNA repair genes, including BRCA1 (p.Arg1699Trp and p.Asp1692His), BRCA2 (p.Leu24Ter and p.Trp31Ter), and ERCC6 (p.Arg1221Ter and p.Trp31Ter) gene (**Table 2**). In case of CHEK1 (p.Gln346X), MUTYH (p.Gln426Ter), and RAD51D (p.Trp314Ter). These findings resonate profoundly with the intricate landscape of DNA repair and maintenance mechanisms, shedding light on potential links between these genetic aberrations and lung cancer development.

BRCA1 and BRCA2 genes are mainly involved in maintaining genomic stability [41]. Mutations in those genes are widely recognized for their pivotal roles in hereditary breast and ovarian cancers, emerging evidence suggests their potential involvement as causative factors in lung cancer as well [42, 43]. Although pathogenic mutations in those genes are more common in breast and ovarian tissues, their significance in lung cancer pathogenesis has gained attention due to their roles in DNA repair mechanisms and maintenance of genomic integrity [44, 45]. Dysfunctional BRCA1 and BRCA2 genes could compromise the cell's ability to repair DNA damage, potentially leading to an accumulation of mutations in lung cells and increasing the risk of tumorigenesis [46]. Recent studies have reported the presence of BRCA1 and BRCA2 mutations in lung cancer patients, particularly in individuals with a strong familial history of these mutations [47, 48].

ERCC6, an essential component of the transcription-coupled nucleotide excision repair pathway, plays a pivotal role in rectifying DNA damage [49]. A study by Ma et al. highlighted ERCC6 mutations in lung cancer patients, implying its potential contribution to genomic instability and tumorigenesis [50]. CHEK1, a key regulator of cell cycle checkpoints and DNA damage responses, is another gene that gained massive attention recently [51]. Mutations

in CHEK1 might lead to unchecked and increased cell division [52]. A study conducted by Alorjani et al. reported frequent CHEK1 mutations in a subset of lung cancer cases, indicating their possible involvement in disease onset [53]. MUTYH, responsible for DNA base excision repair, has garnered interest due to its connection with oxidative DNA damage [54]. Mutations in MUTYH could lead to the accumulation of DNA lesions, possibly contributing to lung cancer progression [55]. Recent studies demonstrated the presence of MUTYH mutations in lung adenocarcinomas, suggesting a role in the mutational landscape of the disease [56, 57]. RAD51D, an essential component of homologous recombination repair, has implications for genome stability [58]. RAD51D mutations could impair DNA repair processes and contribute to genomic alterations in lung cells [59]. A study by Grundy et al. identified RAD51D mutations in lung squamous cell carcinomas, hinting at their potential involvement in the disease's molecular mechanism [60].

The presence of mutated genes has been correlated with poorer OS outcomes in cancer patients [61, 62]. In our study, detected pathogenic mutations in BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D have demonstrated a significant association with reduced OS in mutated lung cancer patients relative to the wild-type patients. This underscores the clinical relevance of mutated genes as potential prognostic indicators and therapeutic targets.

Moreover, targeted bisulfite-seq, RT-qPCR, and IHC analyses revealed that BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes were significantly hypermethylated and down-regulated in mutated lung cancer samples as compared to the wild-type lung cancer samples. This down-regulation could compromise DNA repair pathways, heightening mutation accumulation and genomic instability in cancer cells. These findings align with the concept that gene expression alterations contribute to oncogenesis [63, 64].

Moreover, the analysis of pathways indicated substantial engagement of mutated BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes in the modulation of various important pathways, encompassing "homologous recombination", "fanconi anemia pathway", "base excision repair", and "nucleotide excision repair" among lung cancer patients.

Conclusion

In summary, using the WES technique, we have identified clinically important pathogenic mutations in six DNA repair genes, including BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D across Pakistani lung cancer patients for the first time. Further downstream analysis showed that mutations in those genes were also prognostic biomarkers of worse OS in Pakistani lung cancer patients.

Disclosure of conflict of interest

None.

Address correspondence to: Muhammad Jamil, PARC Arid Zone Research Center, Dera Ismail Khan 29050, Pakistan. E-mail: jamilmatrah@parc.gov.pk; Qiumin Yang, Department of Oncology, Shangqiu First People's Hospital, Shangqiu 476000, Henan, China. E-mail: sqsyxfzb@126.com

References

[1] Zhou Y and Chen T. Human microbiota: a crucial gatekeeper in lung cancer initiation, progression, and treatment. Med Microecol 2022; 13: 100055.

[2] Usman M, Okla MK, Asif HM, AbdElgayed G, Muccee F, Ghazanfar S, Ahmad M, Iqbal MJ, Sahar AM, Khaliq G, Shoaib R, Zaheer H and Hameed Y. A pan-cancer analysis of GINS complex subunit 4 to identify its potential role as a biomarker in multiple human cancers. Am J Cancer Res 2022; 12: 986-1008.

[3] Sinjab A, Rahal Z and Kadara H. Cell-by-cell: unlocking lung cancer pathogenesis. Cancers (Basel) 2022; 14: 3424.

[4] Sial N, Saeed S, Ahmad M, Hameed Y, Rehman A, Abbas M, Asif R, Ahmed H, Hussain MS, Rehman JU, Atif M and Khan MR. Multi-omics analysis identified TMED2 as a shared potential biomarker in six subtypes of human cancer. Int J Gen Med 2021; 14: 7025-7042.

[5] Dietz S, Schirmer U, Mercé C, Von Bubnoff N, Dahl E, Meister M, Muley T, Thomas M and Sültmann H. Low input whole-exome sequencing to determine the representation of the tumor exome in circulating DNA of non-small cell lung cancer patients. PLoS One 2016; 11: e0161012.

[6] Zheng CX, Gu ZH, Han B, Zhang RX, Pan CM, Xiang Y, Rong XJ, Chen X, Li QY and Wan HY. Whole-exome sequencing to identify novel somatic mutations in squamous cell lung cancers. Int J Oncol 2013; 43: 755-764.
[7] Stranneheim H and Wedell A. Exome and genome sequencing: a revolution for the discovery and diagnosis of monogenic disorders. J Intern Med 2016; 279: 3-15.
[8] Usman M, Hameed Y, Ahmad M, Iqbal MJ, Maryam A, Mazhar A, Naz S, Tanveer R, Saeed H, Bint-E-Fatima, Ashraf A, Hadi A, Hameed Z, Tariq E and Aslam AS. SHMT2 is associated with tumor purity, CD8+ T immune cells infiltration, and a novel therapeutic target in four different human cancers. Curr Mol Med 2023; 23: 161-176.
[9] Warr A, Robert C, Hume D, Archibald A, Deeb N and Watson M. Exome sequencing: current and future perspectives. G3 (Bethesda) 2015; 5: 1543-1550.
[10] Sial N, Rehman JU, Saeed S, Ahmad M, Hameed Y, Atif M, Rehman A, Asif R, Ahmed H, Hussain MS, Khan MR, Ambreen A and Ambreen A. Integrative analysis reveals methylenetetrahydrofolate dehydrogenase 1-like as an independent shared diagnostic and prognostic biomarker in five different human cancers. Biosci Rep 2022; 42: BSR20211783.
[11] Shigematsu H and Gazdar AF. Somatic mutations of epidermal growth factor receptor signaling pathway in lung cancers. Int J Cancer 2006; 118: 257-262.
[12] Hameed Y and Ejaz S. TP53 lacks tetramerization and N-terminal domains due to novel inactivating mutations detected in leukemia patients. J Cancer Res Ther 2021; 17: 931-937.
[13] Khan M and Hameed Y. Discovery of novel six genes-based cervical cancer-associated biomarkers that are capable to break the heterogeneity barrier and applicable at the global level. J Cancer Res Ther 2022; 9000: 104.
[14] Usman M and Hameed Y. GNB1, a novel diagnostic and prognostic potential biomarker of head and neck and liver hepatocellular carcinoma. J Cancer Res Ther 2022; 9000: 148.
[15] Rosell R and Karachaliou N. Large-scale screening for somatic mutations in lung cancer. Lancet 2016; 387: 1354-1356.
[16] Shinmura K, Yamaguchi S, Saitoh T, Kohno T and Yokota J. Somatic mutations and single nucleotide polymorphisms of base excision repair genes involved in the repair of 8-hydroxyguanine in damaged DNA. Cancer Lett 2001; 166: 65-69.
[17] Gabriel AAG, Atkins JR, Penha RCC, Smith-Byrne K, Gaborieau V, Voegele C, Abedi-Ardekani B, Milojevic M, Olaso R, Meyer V, Boland A, Deleuze JF, Zaridze D, Mukeriya A, Swiatkowska B, Janout V, Schejbalová M, Mates D, Stojšić J, Ognjanovic M; ILCCO consortium, Witte JS, Rashkin SR, Kachuri L, Hung RJ, Kar S, Brennan P, Sertier AS, Ferrari A, Viari A, Johansson M, Amos CI, Foll M and McKay JD. Genetic analysis of lung cancer and the germline impact on somatic mutation burden. J Natl Cancer Inst 2022; 114: 1159-1166.
[18] Burma S, Chen BP and Chen DJ. Role of nonhomologous end joining (NHEJ) in maintaining genomic integrity. DNA Repair (Amst) 2006; 5: 1042-1048.
[19] Martin SA, Hewish M, Lord CJ and Ashworth A. Genomic instability and the selection of treatments for cancer. J Pathol 2010; 220: 281-289.
[20] Salmaninejad A, Ilkhani K, Marzban H, Navashenaq JG, Rahimirad S, Radnia F, Yousefi M, Bahmanpour Z, Azhdari S and Sahebkar A. Genomic instability in cancer: molecular mechanisms and therapeutic potentials. Curr Pharm Des 2021; 27: 3161-3169.
[21] Campbell MC and Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annu Rev Genomics Hum Genet 2008; 9: 403-433.
[22] Burrell RA, McGranahan N, Bartek J and Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. Nature 2013; 501: 338-345.
[23] Pauls SU, Nowak C, Bálint M and Pfenninger M. The impact of global climate change on genetic diversity within populations and species. Mol Ecol 2013; 22: 925-946.
[24] Pharoah PD, Dunning AM, Ponder BA and Easton DF. Association studies for finding cancer-susceptibility genetic variants. Nat Rev Cancer 2004; 4: 850-860.
[25] World Medical Association. World Medical Association declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013; 310: 2191-2194.
[26] Gupta N. DNA extraction and polymerase chain reaction. J Cytol 2019; 36: 116-117.
[27] Rio DC, Ares M Jr, Hannon GJ and Nilsen TW. Purification of RNA using TRIzol (TRI reagent). Cold Spring Harb Protoc 2010; 2010: pdb.prot5439.
[28] Lee PY, Costumbrado J, Hsu CY and Kim YH. Agarose gel electrophoresis for the separation of DNA fragments. J Vis Exp 2012; 3923.
[29] Gaudin M and Desnues C. Hybrid capture-based next generation sequencing and its application to human infectious diseases. Front Microbiol 2018; 9: 2924.
[30] Chen C, Khaleel SS, Huang H and Wu CH. Software for pre-processing Illumina next-generation sequencing short read sequences. Source Code Biol Med 2014; 9: 8.

[31] Li H and Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics 2009; 25: 1754-1760.
[32] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M and DePristo MA. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010; 20: 1297-1303.
[33] Gudmundsson S, Singer-Berk M, Watts NA, Phu W, Goodrich JK, Solomonson M; Genome Aggregation Database Consortium, Rehm HL, MacArthur DG and O'Donnell-Luria A. Variant interpretation using population databases: lessons from gnomAD. Hum Mutat 2022; 43: 1012-1030.
[34] Goel MK, Khanna P and Kishore J. Understanding survival analysis: Kaplan-Meier estimate. Int J Ayurveda Res 2010; 1: 274-278.
[35] Livak KJ and Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 2001; 25: 402-408.
[36] Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C and Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012; 2: 401-404.
[37] Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C and Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 2019; 10: 1523.
[38] Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, Gautam B and Hassanali M. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res 2008; 36: D901-D906.
[39] Miller DT, Lee K, Chung WK, Gordon AS, Herman GE, Klein TE, Stewart DR, Amendola LM, Adelman K, Bale SJ, Gollob MH, Harrison SM, Hershberger RE, McKelvey K, Richards CS, Vlangos CN, Watson MS and Martin CL; ACMG Secondary Findings Working Group. ACMG SF v3.0 list for reporting of secondary findings in clinical exome and genome sequencing: a policy statement of the American College of Medical Genetics and Genomics (ACMG). Genet Med 2021; 23: 1381-1390.
[40] Marian AJ. Clinical interpretation and management of genetic variants. JACC Basic Transl Sci 2020; 5: 1029-1042.
[41] Gorodetska I, Kozeretska I and Dubrovska A. BRCA genes: the role in genome stability, cancer stemness and therapy resistance. J Cancer 2019; 10: 2109-2127.
[42] Mehrgou A and Akouchekian M. The importance of BRCA1 and BRCA2 genes mutations in breast cancer development. Med J Islam Repub Iran 2016; 30: 369.
[43] Angeli D, Salvi S and Tedaldi G. Genetic predisposition to breast and ovarian cancers: how many and which genes to test? Int J Mol Sci 2020; 21: 1128.
[44] Larsen MJ, Thomassen M, Gerdes AM and Kruse TA. Hereditary breast cancer: clinical, pathological and molecular characteristics. Breast Cancer (Auckl) 2014; 8: 145-155.
[45] Tomasova K, Cumova A, Seborova K, Horak J, Koucka K, Vodickova L, Vaclavikova R and Vodicka P. DNA repair and ovarian carcinogenesis: impact on risk, prognosis and therapy outcome. Cancers (Basel) 2020; 12: 1713.
[46] Krishnan R, Patel PS and Hakem R. BRCA1 and metastasis: outcome of defective DNA repair. Cancers (Basel) 2021; 14: 108.
[47] Lee YC, Lee YC, Li CY, Lee YL and Chen BL. BRCA1 and BRCA2 gene mutations and lung cancer sisk: a meta-analysis. Medicina (Kaunas) 2020; 56: 212.
[48] Yan B, Xie B, Huang M, Guo J, Sun J, Chen J, Tao Y and Xiao D. Mutations and expressions of breast cancer 1/2 in lung cancer. Thorac Cancer 2023; 14: 1753-1763.
[49] Lin Z, Zhang X, Tuo J, Guo Y, Green B, Chan CC, Tan W, Huang Y, Ling W, Kadlubar FF, Lin D and Ning B. A variant of the Cockayne syndrome B gene ERCC6 confers risk of lung cancer. Hum Mutat 2008; 29: 113-122.
[50] Ma H, Hu Z, Wang H, Jin G, Wang Y, Sun W, Chen D, Tian T, Jin L, Wei Q, Lu D, Huang W and Shen H. ERCC6/CSB gene polymorphisms and lung cancer risk. Cancer Lett 2009; 273: 172-176.
[51] Patil M, Pabla N and Dong Z. Checkpoint kinase 1 in DNA damage response and cell cycle regulation. Cell Mol Life Sci 2013; 70: 4009-4021.
[52] Wu M, Pang JS, Sun Q, Huang Y, Hou JY, Chen G, Zeng JJ and Feng ZB. The clinical significance of CHEK1 in breast cancer: a high-throughput data analysis and immunohistochemical study. Int J Clin Exp Pathol 2019; 12: 1-20.
[53] Alorjani M, Aburub M, Al-Trad B, Hamad MA, AbuAlarja M, Bashir SA, Al-Batayneh K and Zoubi MA. The prevalence of CHEK1 and CHEK2 mutations in prostate cancer: a retrospective cohort study. Med Arch 2023; 77: 8-12.
[54] Kairupan C and Scott RJ. Base excision repair and the role of MUTYH. Hered Cancer Clin Pract 2007; 5: 199-209.
[55] Robinson PS, Thomas LE, Abascal F, Jung H, Harvey LMR, West HD, Olafsson S, Lee BCH,

Coorens THH, Lee-Six H, Butlin L, Lander N, Truscott R, Sanders MA, Lensing SV, Buczacki SJA, Ten Hoopen R, Coleman N, Brunton-Sim R, Rushbrook S, Saeb-Parsy K, Lalloo F, Campbell PJ, Martincorena I, Sampson JR and Stratton MR. Inherited MUTYH mutations cause elevated somatic mutation rates and distinctive mutational signatures in normal human cells. Nat Commun 2022; 13: 3949.

[56] Curia MC, Catalano T and Aceto GM. MUTYH: not just polyposis. World J Clin Oncol 2020; 11: 428-449.

[57] Reckamp KL, Behrendt CE, Slavin TP, Gray SW, Castillo DK, Koczywas M, Cristea MC, Babski KM, Stearns D, Marcum CA, Rodriguez YP, Hass AJ, Vecchio MM, Mora P, Cervantes AE, Sand SR, Mejia RM, Tsou TC, Salgia R and Weitzel JN. Germline mutations and age at onset of lung adenocarcinoma. Cancer 2021; 127: 2801-2806.

[58] Reh WA, Nairn RS, Lowery MP and Vasquez KM. The homologous recombination protein RAD51D protects the genome from large deletions. Nucleic Acids Res 2017; 45: 1835-1847.

[59] Baldock RA, Pressimone CA, Baird JM, Khodakov A, Luong TT, Grundy MK, Smith CM, Karpenshif Y, Bratton-Palmer DS, Prakash R, Jasin M, Garcin EB, Gon S, Modesti M and Bernstein KA. RAD51D splice variants and cancer-associated mutations reveal XRCC2 interaction to be critical for homologous recombination. DNA Repair (Amst) 2019; 76: 99-107.

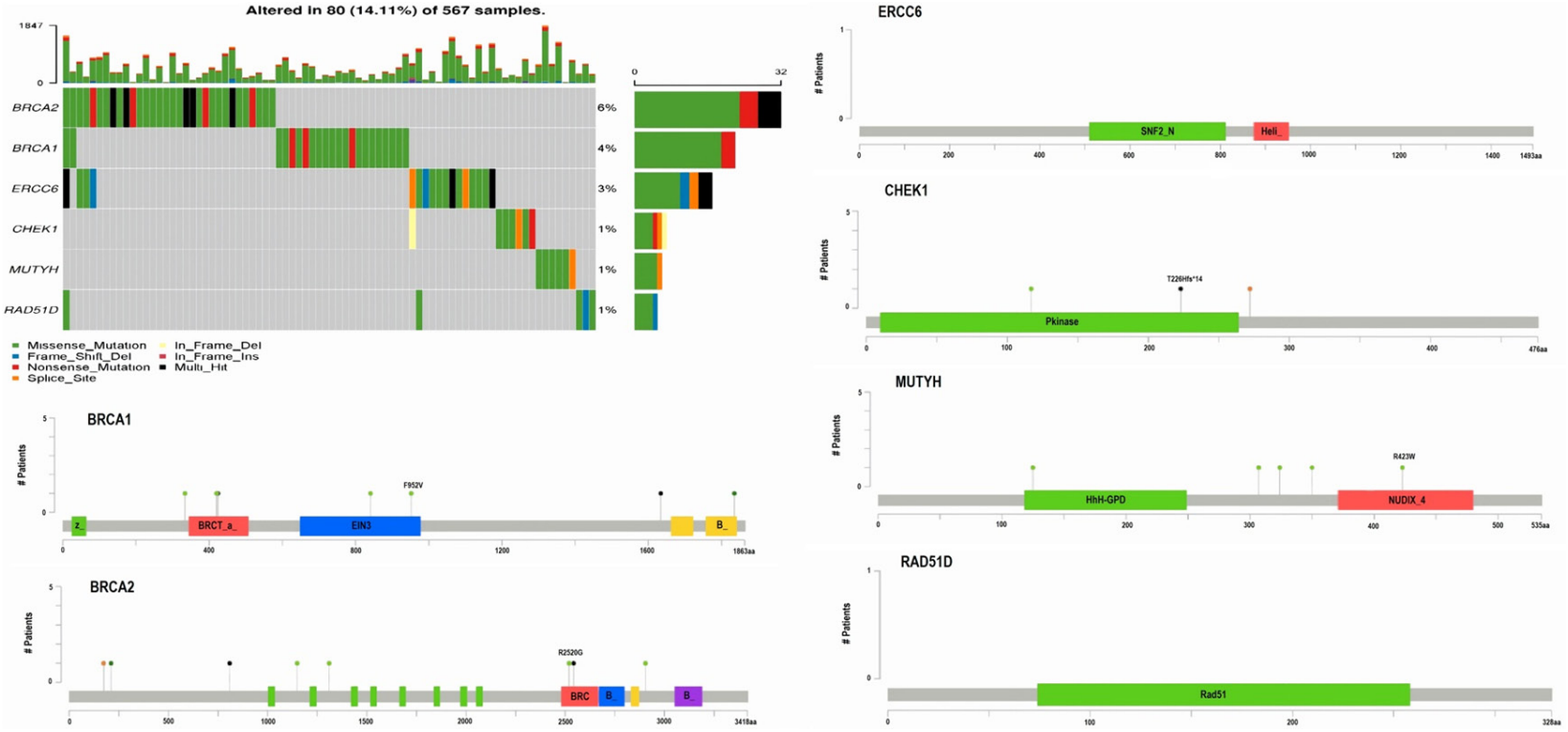
[60] Grundy MK, Buckanovich RJ and Bernstein KA. Regulation and pharmacological targeting of RAD51 in cancer. NAR Cancer 2020; 2: zcaa024.

[61] Mangone FR, Bobrovnitchaia IG, Salaorni S, Manuli E and Nagai MA. PIK3CA exon 20 mutations are associated with poor prognosis in breast cancer patients. Clinics (Sao Paulo) 2012; 67: 1285-1290.

[62] Tan E, Whiting J, Xie H, Imanirad I, Carballido E, Felder S, Frakes J, Mo Q, Walko C, Permuth JB, Sommerer K, Kim R, Anaya DA, Fleming JB and Sahin IH. BRAF mutations are associated with poor survival outcomes in advanced-stage mismatch repair-deficient/microsatellite high colorectal cancer. Oncologist 2022; 27: 191-197.

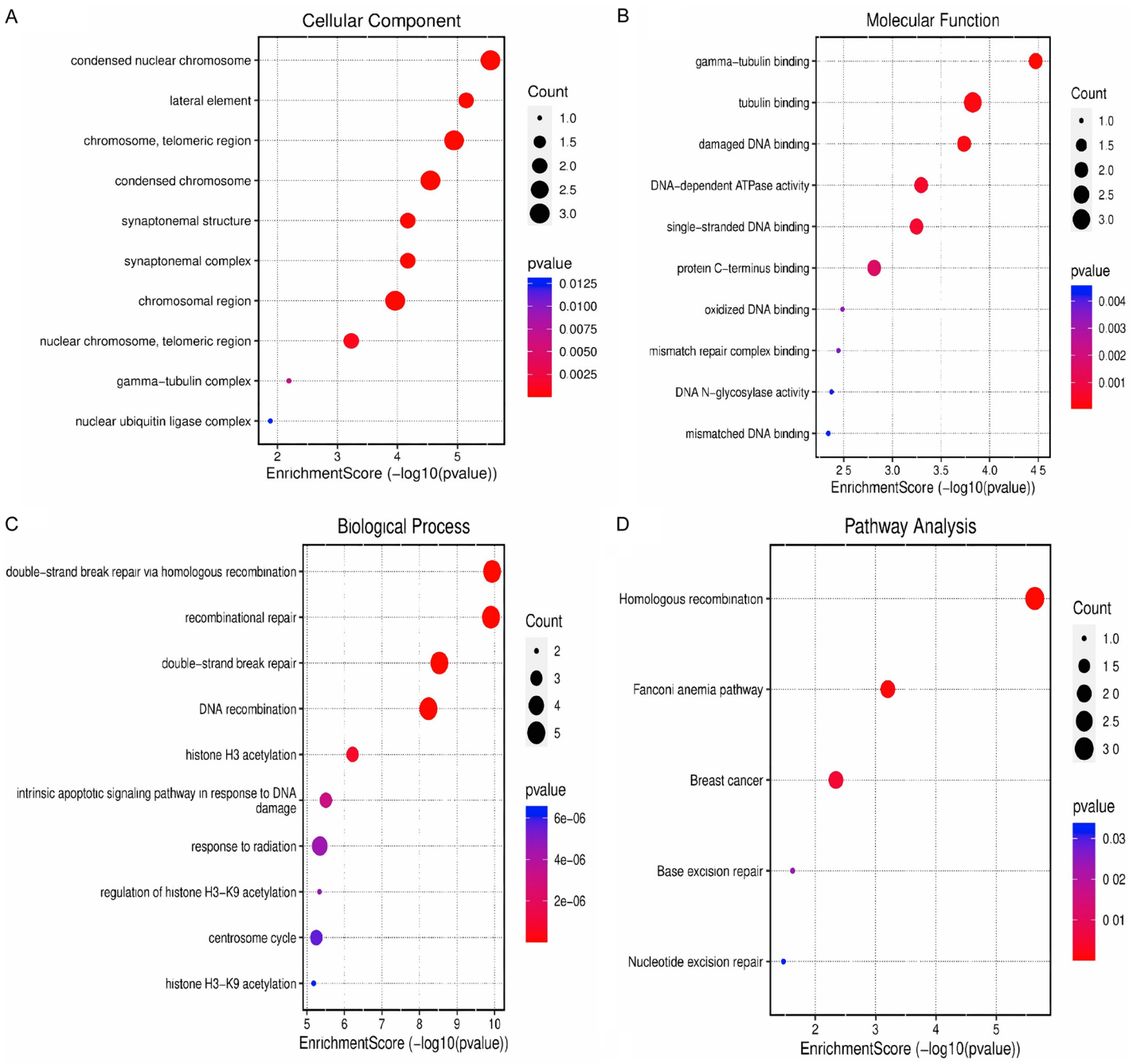
[63] Chakravarthi BV, Nepal S and Varambally S. Genomic and epigenomic alterations in cancer. Am J Pathol 2016; 186: 1724-1735.

[64] Herceg Z and Hainaut P. Genetic and epigenetic alterations as biomarkers for cancer detection, diagnosis and prognosis. Mol Oncol 2007; 1: 26-41.



Supplementary Figure 1. Mutational analysis results of BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes across TCGA lung cancer samples via cBioPortal platform.

WES analysis of lung cancer patients



Supplementary Figure 2. GO and KEGG analyses of BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes via Metascape. (A) BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes-related CC terms, (B) BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes-related MF terms, (C) BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes-related BP terms, and (D) BRCA1, BRCA2, ERCC6, CHEK1, MUTYH, and RAD51D genes-related KEGG terms. A $P < 0.05$ was used as the cut-off criterion.